

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic
Theory in its Relation to Statistics and Mathematics*

CONTENTS

TIMOTHY BESLEY AND TORSTEN PERSSON: State Capacity, Conflict, and Development ...	1
MATTHEW GENTZKOW AND JESSE M. SHAPIRO: What Drives Media Slant? Evidence From U.S. Daily Newspapers	35
BRUNO BIAIS, THOMAS MARIOTTI, JEAN-CHARLES ROCHE, AND STÉPHANE VILLENEUVE: Large Risks, Limited Liability, and Dynamic Moral Hazard.....	73
DONALD W. K. ANDREWS AND GUSTAVO SOARES: Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection	119
GARY CHAMBERLAIN: Binary Response Models for Panel Data: Identification and Information	159
JOSEPH P. ROMANO AND AZEEM M. SHAIKH: Inference for the Identified Set in Partially Identified Econometric Models.....	169
MARK ARMSTRONG AND JOHN VICKERS: A Model of Delegated Project Choice.....	213
RENÉ CALDENTEY AND ENNIO STACCHETTI: Insider Trading With a Random Deadline ...	245
DAVID RAHMAN AND ICHIRO OBARA: Mediated Partnerships	285
ALESSANDRO CITANNA AND PAOLO SICONOLFI: Recursive Equilibrium in Stochastic Overlapping-Generations Economies	309
NOTES AND COMMENTS:	
JOHN E. STOVALL: Multiple Temptations	349
PEDRO CARNEIRO, JAMES J. HECKMAN, AND EDWARD VYTLACIL: Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin	377
BRENDAN K. BEARE: Copulas and Temporal Dependence	395
ANNOUNCEMENTS	
FORTHCOMING PAPERS	411
REPORT OF THE SECRETARY	413
REPORT OF THE TREASURER	415
REPORT OF THE EDITORS 2008–2009	425
ECONOMETRICA REFEREES 2008–2009	433
REPORT OF THE EDITORS OF THE MONOGRAPH SERIES	437
SUBMISSION OF MANUSCRIPTS TO THE ECONOMETRIC SOCIETY MONOGRAPH SERIES	447
	451

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic
Theory in its Relation to Statistics and Mathematics*

Founded December 29, 1930

Website: www.econometricsociety.org

EDITOR

STEPHEN MORRIS, Dept. of Economics, Princeton University, Fisher Hall, Prospect Avenue, Princeton, NJ 08544-1021, U.S.A.; morris@econometricsociety.org

MANAGING EDITOR

GERI MATTSON, 2002 Holly Neck Road, Baltimore, MD 21221, U.S.A.; mattsonpublishingservices@comcast.net

CO-EDITORS

DARON ACEMOGLU, Dept. of Economics, MIT, E52-380B, 50 Memorial Drive, Cambridge, MA 02142-1347, U.S.A.; daron@econometricsociety.org

WOLFGANG PESENDORFER, Dept. of Economics, Princeton University, Fisher Hall, Prospect Avenue, Princeton, NJ 08544-1021, U.S.A.; wpesendorfer@econometricsociety.org

JEAN-MARC ROBIN, Maison des Sciences Économiques, Université Paris 1 Panthéon-Sorbonne, 106/112 bd de l'Hôpital, 75647 Paris Cedex 13, France and University College London, U.K.; jnrobin@econometricsociety.org

LARRY SAMUELSON, Dept. of Economics, Yale University, 20 Hillhouse Avenue, New Haven, CT 06520-8281, U.S.A.; samuelson@econometricsociety.org

JAMES H. STOCK, Dept. of Economics, Harvard University, Littauer M-24, 1830 Cambridge Street, Cambridge, MA 02138, U.S.A.; jstock@econometricsociety.org

HARALD UHLIG, Dept. of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, U.S.A.; uhlig@econometricsociety.org

ASSOCIATE EDITORS

YACINE AÏT-SAHALIA, Princeton University

OLIVER LINTON, London School of Economics

JOSEPH G. ALTONJI, Yale University

BART LIPMAN, Boston University

JAMES ANDREONI, University of California,
San Diego

THIERRY MAGNAC, Toulouse School of Economics
(GREMAQ and IDEI)

JUSHAN BAI, Columbia University

GEORGE J. MAILATH, University of Pennsylvania

MARCO BATTAGLINI, Princeton University

DAVID MARTIMORT, IDEI-GREMAQ,

PIERPAOLO BATTIGALLI, Università Bocconi

Université des Sciences Sociales de Toulouse

DIRK BERGEMANN, Yale University

STEVEN A. MATTHEWS, University of

XIAOHONG CHEN, Yale University

Pennsylvania

VICTOR CHERNOZHUKOV, Massachusetts
Institute of Technology

ROSA L. MATZKIN, University of California,
Los Angeles

J. DARRELL DUFFIE, Stanford University

LEE OHANIAN, University of California,

JEFFREY ELY, Northwestern University

Los Angeles

HALUK ERGIN, Washington University
in St. Louis

WOJCIECH OLSZEWSKI, Northwestern University

MIKHAIL GOLOSOV, Yale University

NICOLA PERSICO, New York University

FARUK GUL, Princeton University

BENJAMIN POLAK, Yale University

JINYONG HAHN, University of California,
Los Angeles

PHILIP J. RENY, University of Chicago

PHILIP A. HAILE, Yale University

SUSANNE M. SCHENNACH, University of Chicago

MICHAEL JANSSON, University of California,
Berkeley

UZI SEGAL, Boston College

PHILIPPE JEHIEL, Paris School of Economics
and University College London

NEIL SHEPHERD, University of Oxford

PER KRUSELL, Princeton University and
Stockholm University

MARCIANO SINISCALCHI, Northwestern University

FELIX KUBLER, University of Zurich

JEROEN M. SWINKELS, Northwestern University

EDITORIAL ASSISTANT: MARY BETH BELLANDO, Dept. of Economics, Princeton University, Fisher Hall, Princeton, NJ 08544-1021, U.S.A.; econometrica@econometricsociety.org

Information on MANUSCRIPT SUBMISSION is provided in the last two pages.

Information on MEMBERSHIP, SUBSCRIPTIONS, AND CLAIMS is provided in the inside back cover.

STATE CAPACITY, CONFLICT, AND DEVELOPMENT

BY TIMOTHY BESLEY AND TORSTEN PERSSON¹

The absence of state capacities to raise revenue and to support markets is a key factor in explaining the persistence of weak states. This paper reports on an ongoing project to investigate the incentive to invest in such capacities. The paper sets out a simple analytical structure in which state capacities are modeled as forward looking investments by government. The approach highlights some determinants of state building including the risk of external or internal conflict, the degree of political instability, and dependence on natural resources. Throughout, we link these state capacity investments to patterns of development and growth.

KEYWORDS: State building, civil war, weak states.

A STRIKING FEATURE of economic development is an apparent symbiotic evolution of strong states and strong market economies. However, traditional analyses of economic development tend to focus on the expansion of the market economy with less attention paid to the expansion of the state. Just as private physical and human capital accumulation is a key engine of private sector growth, the buildup of public capital is also an engine of state expansion. It is arguable that a good part of investing in state effectiveness comes from improving the state's ability to implement a range of policies, something which we refer to as *state capacity*. Nowadays, this concept is commonplace in other branches of social science. Coined by historical sociologists, such as Charles Tilly, state capacity originally referred to the power of the state to raise revenue. Here we broaden it to capture the wider range of competencies that the state acquires in the development process, which includes the power to enforce contracts and support markets through regulation or otherwise.

The issue of state capacity is also common currency in the applied development community, where it is intimately associated with the concept of *weak* or *fragile* states. Weak states tend to be hopelessly poor—unable to maintain basic economic functions and raise the revenue required to deliver basic services to their citizens. They are also often plagued by civil disorder or outright conflict. This propensity toward conflict and weak government institutions tends to be clustered with low income levels and stagnation.

This paper puts forward a simple model of investments in state capacity. It provides a unifying framework for thinking about a range of issues that, so far, have been discussed as disparate phenomena: the risk of external or internal conflict, the degree of political instability, and economic dependence on natural resources. It provides answers, albeit in a stylized way, to a range

¹This paper was the basis for Persson's Presidential Address to the Econometric Society in 2008. We are grateful to a co-editor and four referees, as well as a number of participants in regional meetings for comments. We thank David Seim and Prakarsh Singh for research assistance, and CIFAR, ESRC, and the Swedish Research Council for financial support.

of questions: What are the main economic and political determinants of the state's capacity to raise revenue and support markets? How do risks of violent conflict affect the incentives to invest in state building? Does it matter whether conflicts are external or internal to the state? What may be the mechanisms whereby weak states are associated with lower income levels and growth rates than strong states? What relations should we expect between resource rents, civil wars, and economic development? These questions are now occupying the attention of many scholars who try to understand patterns of development across time and place.

Section 1 of the paper presents a basic model in which building state capacity to raise taxes (fiscal capacity) and support markets (legal capacity) are modeled as investments under uncertainty. Our model yields a series of benchmark results, detailing how investments in state capacity depend on a number of structural factors. It shows why we might expect the two forms of state capacity to be complements and hence to develop together, and illustrates why a lower risk of external conflict, a higher degree of resource dependence, as well as lower political stability, weaken the incentive for state building. This basic framework serves as a building block and is put to work in the subsequent two sections.

Section 2 models political stability endogenously, with the rate of turnover being affected by internal conflicts initiated by an opposition group of insurgents. Here, we model internal—as opposed to external—violent conflict, by allowing incumbent and opposition groups to invest in violence. Having characterized the circumstances when the economy ends up in peace, government repression, or civil war, we revisit the analysis of investments in state capacity. The results illustrate how high resource dependence may jointly trigger a high propensity toward conflict, low income, and low investments in legal and fiscal capacity.

In Section 3, we examine how building fiscal capacity can improve other aspects of policy making. Here, we extend the basic framework by allowing for quasi-rents in production. In this model version, political instability can keep the economy in an investment trap, where low investments in fiscal capacity perpetuate inefficient regulatory policies to redistribute income through rent creation/protection rather than through taxation. This in turn leads to factor market distortions, lower investments in market support, and low income/growth. The results suggest another channel that links together weak state capacity and low income, which again works through weak incentives to build the state.

The association of weak states (manifested in low state capacity) with poor economic performance is a theme that runs across all three sections. A unified model of the incentive to invest in state capacity is at the heart of each section and lays bare a common set of factors that shape low levels of state capacity, which have not been joined together in previous approaches. In each section, the theoretical results are summarized in a few key propositions. We discuss

the implications of the theory and comment on its relationship to the existing literature, as well as mentioning some relevant empirical work. A short concluding section takes stock of the findings and suggests topics for further research.

1. THE ORIGINS OF STATE CAPACITY

This section develops the core model for analyzing the incentive to invest in state capacity, based on Besley and Persson (2009a). As we mentioned in the introduction, economists have paid little attention to state capacity investments. For example, researchers in public finance, political economics, or development rarely assume that a government, which finds a certain tax rate for a certain tax base optimal and incentive-compatible, is constrained by fiscal infrastructure. Similarly, economic theory rarely assumes that the state is constrained by a lack of legal infrastructure when it comes to enforcing private contracts or, more generally, supporting private markets.

This contrasts with the approach taken by political and economic historians who view the state's capacity to raise revenue as an important phenomenon in itself. They link to a thirst for military success and regard it as a key factor behind the successful development of nation states (see, e.g., Tilly (1985), Levi (1988), or Brewer (1989)). In line with the core thesis, the tax systems in countries such as the United States, the United Kingdom, and Sweden, have indeed been reformed and expanded in connection with actual or latent external conflicts. Political scientists such as Migdal (1988) have emphasized that one of the major problems of developing countries is that their states are often too weak and lack the capacity to raise revenue and to govern effectively. State capacities and weak states are also major concepts in the development policy community.²

The starting point taken outside of economics has some attraction given the practical experience of economic development. Presupposing sufficient capacities to tax and support markets does not sit well with the experience of many states, either in history or in the developing world of today. Moreover, international data suggest that the ability to raise revenue from advanced tax systems is strongly positively related to the ability to support markets, as well as to the level of economic development.

Figure 1 illustrates these patterns in the data. It shows the positive correlations in contemporary data between the tax share of gross domestic product (GDP; vertical axis), an index of property-rights protections (horizontal axis), and income (gray dots denote above and black dots denote below median income in 1980). There is no good reason to believe that these correlations can be interpreted causally. Indeed, our core model will emphasize the joint determination of these variables, where institutions, historical shocks, and initial

²See, for example, Rice and Patrick (2008) for a discussion and definition of weak states.

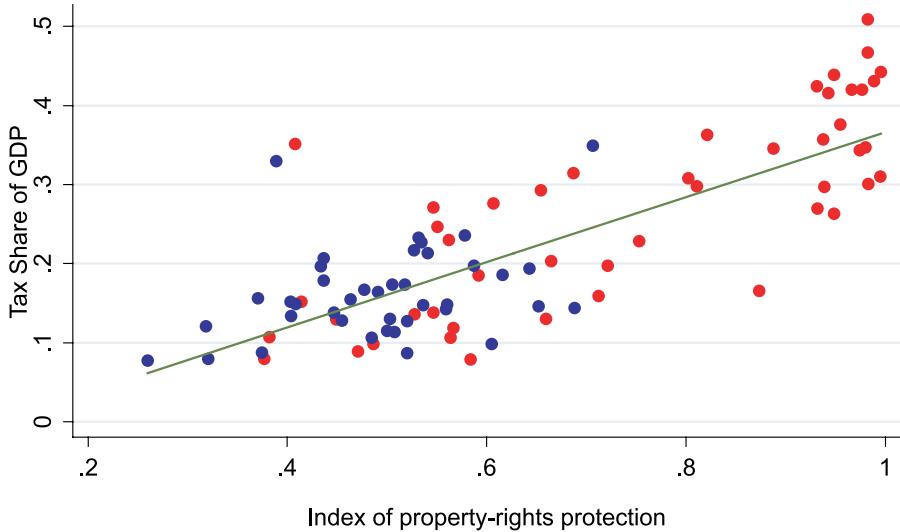


FIGURE 1.—Patterns of fiscal and legal capacity. ● denotes above median income in 1980; ● denotes below median income in 1980; — denotes fitted values.

conditions are common omitted factors that jointly drive taxation, property-rights protection, and income.

The model of Besley and Persson (2009a) separates decisions about investments that enhance the feasible set of policies from decisions about the policies themselves.³ Thus, taxes and market-supporting policies are constrained by the state's fiscal and legal capacity. Expansion of these capacities is viewed as forward-looking investments under uncertainty. A central result that emerges from the framework, under specific assumptions, is an important complementarity between fiscal and legal capacity. This implies that the two forms of state capacity are likely to be positively correlated with each other and with income as Figure 1 suggests.

1.1. Basic Model Setup

The model is stripped down to give a simple and transparent account of the important factors. Total population size is normalized to 1. There are two groups, each of which comprises half the population in every time period. For the purposes of this paper, two alternative timing structures give essentially the same results. In one, time is infinite and one generation is alive in each

³Recent related papers include Acemoglu (2005), where governments can increase their future tax revenues by spending on public goods, and Acemoglu, Ticchi, and Vindigni (2009), who studied the buildup of government bureaucracies. Earlier, fiscal capacity was studied by Cukierman, Edwards, and Tabellini (1992); legal capacity investment was studied by Svensson (1998).

period, making investment decisions based on a warm-glow bequest motive. In the other, which we adhere to here, there are just two time periods, $s = 1, 2$, and the world ends after period 2. Although artificial, this two-period approach allows us to make the main points of economic interest.

At the beginning of period 2, the group that held power at the end of period 1 is the incumbent government, denoted by I_1 . The other group is the opposition, denoted by O_1 . Power can be peacefully transferred to the opposition, which happens with exogenous probability given by parameter γ . This can be thought of as the reduced form of some underlying political process, which we do not model. As a result, whoever wins becomes the new incumbent, I_2 , and whoever loses becomes the new opposition, O_2 . At the end of period s , the current incumbent, I_s , sets a tax on the income of each group member, denoted by t^{I_s} , where $J_s \in \{I_s, O_s\}$. It also chooses a level of legal support for each group, p^{I_s} , and spends on general public goods, G_s . At the end of period 1, incumbent I_1 also makes investments in the next period's state capacity (see below). In addition to tax income, the government earns natural resource rents R_s . These are stochastic and drawn from a two-point distribution $\{R_L, R_H\}$ where $R_s = R_H$ with probability ρ in each period. None of the resource rents accrue directly to the private sector.⁴

The precise timing of these events is spelled out below.

Individual Incomes and Utility

In period s , individuals consume and produce, with members of group J_s earning a market income

$$w^{J_s} = w(p^{J_s}),$$

where $w(\cdot)$ is an increasing concave function. The policy variables p^{J_s} can be interpreted in a number of ways. In a broad sense, we view them as a reduced form for market-supporting policies that raise private incomes of group J . This might include the provision of productive physical infrastructure such as roads, ports, and bridges. The distinctive feature of policy is that the way such capacity is deployed, reflected by p^{J_s} , is distinct from the *capacity* to use the policy, a feature that we introduce below. Following Besley and Persson (2009a), we refer to p^{J_s} throughout as if they are policies that affect legal enforcement and raise incomes by facilitating gains from trade in capital markets.⁵

⁴We could add private natural resources as accruing additively to private incomes without any effect on the incentives that we model in this paper. In this case, R_s can be thought of as the share of rents that accrue to the public sector.

⁵Besley and Persson (2009a) developed a microfounded model with less than perfect enforcement (by the state) of collateral in (private) credit-market contracts. The policy p^{J_s} in this context is interpreted as policies that allow greater use of collateral to support trade in credit markets.

One feature of our formulation is worth emphasizing as it is somewhat non-standard. Having created legal capacity π_s , we allow this level of market support to be enjoyed costlessly by *both* groups. However, whether these benefits are extended is a policy decision by government, that is, $0 \leq p^{J_s} \leq \pi_s$. The government can therefore choose to protect the property rights of the two groups to different degrees, given its legal capacity (see below). Creating legal capacity can thus be (conceptually) distinct from regulating access to it.

Individual utility in period s is linear and given by

$$(1) \quad \alpha_s G_s + c^{J_s} = \alpha_s G_s + (1 - t^{J_s})w(p^{J_s}),$$

where c^{J_s} is private consumption and G_s is the level of public goods, with parameter α_s reflecting the value of public goods. We assume that α_s has a two-point distribution $\{\alpha_L, \alpha_H\}$, with $\alpha_H > 2 > \alpha_L$, and we use ϕ to denote the probability that $\alpha_s = \alpha_H$. A specific interpretation is that G_s denotes spending on external defense, while α_s and ϕ capture the severity and risk of external conflict. The equality in (1) arises because we assume that individuals do not save between periods 1 and 2.

Constraints on Government

Policies are constrained by state capacity. The levels of fiscal capacity τ_s , and legal capacity π_s are inherited from the previous period. The incumbent group in period 1 chooses these levels for period 2 given the political institutions in place.

In concrete terms, τ represents fiscal infrastructure, such as a set of competent tax auditors or the institutions necessary to tax income at source or to impose a value-added tax: we can think about τ as decreasing the share of market income $(1 - \tau)$ an individual can earn in the informal sector. Fiscal capacity does not depreciate, but can be augmented by I_1 through nonnegative investments which cost $F(\tau_2 - \tau_1)$, where $F(\cdot)$ is an increasing convex function with $F(0) = F_\tau(0) = 0$. A higher τ_s allows the incumbent I_s to charge higher tax rates, such that $t^{J_s} \leq \tau_s$. To allow for redistribution in a simple way, we allow negative tax rates.

In concrete terms, π represents legal infrastructure investments such as building court systems, educating and employing judges, and registering property or credit. Like fiscal capacity, legal capacity does not depreciate, but can be augmented with nonnegative investments at cost $L(\pi_2 - \pi_1)$, where $L(\cdot)$ is an increasing convex function with $L(0) = L_\pi(0) = 0$. As we mentioned in the last section, a higher π_s allows government I_s to better support private markets with $0 \leq p^{J_s} \leq \pi_s$.

The government budget constraint in period s can be written as

$$(2) \quad 0 \leq \sum_{J_s \in \{I_s, O_s\}} \frac{t^{J_s} w^{J_s}}{2} - G_s + R_s - \begin{cases} L(\pi_2 - \pi_1) - F(\tau_2 - \tau_1), & \text{if } s = 1, \\ 0, & \text{if } s = 2. \end{cases}$$

Given that the opposition takes over with probability γ , this parameter becomes a crude measure of political instability.⁶

Timing

Each period has the following timing:

Stage 1. The initial conditions are $\{\tau_s, \pi_s\}$ and the identity of last period's incumbent is I_{s-1} .

Stage 2. The values of public goods α_s and natural resource rents R_s are realized.

Stage 3. Group I_{s-1} remains in office with probability $1 - \gamma$.

Stage 4. The new incumbent I_s determines a vector of tax rates, legal support, and spending on public goods: $\{(t^{I_s}, p^{I_s}), G_s\}$. The period-1 incumbent also chooses state capacities π_2 for the next period τ_2 .

Stage 5. Payoffs for period s are realized and consumption takes place.

1.2. *Equilibrium Policy*

We begin with the policy choices at Stage 4 of period s . Linearity allows us to study these separately from the choices of state capacity for period 2. With the assumed policy weights, we can write the objective of incumbent I_s as

$$(3) \quad V^{I_s} = w(p^{I_s})(1 - t^{I_s}) + \alpha_s \left[\frac{t^{I_s}w(p^{I_s}) + t^{O_s}w(p^{O_s})}{2} + z_s \right],$$

where we have replaced G_s via the government budget constraint (2) and where residual revenue z_s is defined by

$$z_s = R_s - \begin{cases} L(\pi_2 - \pi_1) - F(\tau_2 - \tau_1), & \text{if } s = 1, \\ 0, & \text{if } s = 2. \end{cases}$$

This objective is maximized subject to $G_s \geq 0$, $t^{I_s} \leq \tau_s$, and $p^{I_s} \leq \pi_s$.

Taxation and Spending on Public Goods

The simple form of (3) makes it easy to derive equilibrium fiscal policy. Whenever $\alpha_s = \alpha_H > 2$, it is optimal for I_s to tax its own group maximally, $t^{I_s} = \tau_s$, and use the revenue to expand G_s . Because I_s puts zero weight on the opposition group, it also sets $t^{O_s} = \tau_s$. If $\alpha_s = \alpha_L < 2$, it becomes optimal

⁶Besley and Persson (2009a) assumed that in its decisions the government internalizes the preferences of the opposition group according to a weight $\theta \in [0, \frac{1}{2}]$ that captures, in a simple and reduced-form way, the inclusiveness of political institutions through checks and balances or electoral systems. Here, we simplify the analysis by assuming that any government acts purely selfishly by maximizing the expected utility of its own group (i.e., we assume $\theta = 0$).

to switch to a redistributive policy, where the opposition is still taxed fully, $t^{O_s} = \tau_s$, but no public goods are provided and

$$-t^{I_s}w(p^{I_s}) = \tau_s w(p^{O_s}) + 2z_s.$$

Thus, whether we have high or low demand for common-interest public goods is crucial. For high α , the incumbent taxes both groups at full capacity and spends all available revenue (less investment costs if $s = 1$) on public goods. When public goods are not very valuable, no public goods are provided and all available revenue is transferred to the incumbent group (through a negative tax rate).⁷ We refer to $\alpha_s = \alpha_H$ as the *common-interest* state and refer to $\alpha_s = \alpha_L$ as the *redistributive* state.

The *realized* value of government funds in period s , which is obtained by differentiating V^{I_s} with regard to z_s , is state dependent and is given by:

$$\lambda_s = \text{Max}[\alpha_s, 2].$$

Legal Protection

It is straightforward to see that (3) is increasing in the legal protection afforded to each group. Thus, it becomes optimal to exploit any existing legal capacity fully and set

$$p^{O_s} = p^{I_s} = \pi_s.$$

Intuitively, the incumbent group can only gain from improving property rights to both groups, either directly via a higher wage or indirectly via a higher tax base. Simple as it is, this production efficiency result is in the spirit of Diamond and Mirrlees (1971). The result does not mean that property rights are well protected everywhere, however, since this hinges on the value of π_s reflecting past investment decisions.

The key point—which can be broadly applied—is that whatever the state’s capacity to improve productivity, it will be shared universally on an open access basis, but as we show in Section 3, when rents are present, the state’s capacity to tax those rents becomes important. In the present setting, however, the result holds regardless of the level of fiscal capacity.

Even though the setup is a bit different, the results on policy are similar to those in Besley and Persson (2009a). Collecting all results allows the following statement.

PROPOSITION 1: *In all states, $p^{J_s} = \pi_s$ for $J_s \in \{I_s, O_s\}$ and $t^{O_s} = \tau_s$. In common-interest states, $G_s = \tau w(\pi_s) + z_s$ and $t^{I_s} = \tau_s$, while in redistributive states, $G_s = 0$ and $-t^{I_s} = \tau_s + 2(z_s/(w(\pi_s)))$.*

⁷Besley and Persson (2009a) emphasized that public goods will generally be underprovided relative to a utilitarian optimum. However, given the two potential values of α_s , this underprovision result is absent here.

1.3. Equilibrium State Capacity

Preliminaries

Using the equilibrium policies in Proposition 1, we can write the expected future payoff to the incumbent at Stage 4 of period 1, taking as given the state capacity for period 2:

$$(4) \quad E[V^{I_1}(\pi_2, \tau_2)] = w(\pi_2)(1 - \tau_2) + E(\lambda_2)[\tau_2 w(\pi_2) + E(z_2)].$$

The expression $E(\lambda_2) = \phi\alpha_H + (1 - \phi)(1 - \gamma)2$ is the *expected* value of government funds in period 2 viewed from the perspective of period 1 and is a key magnitude that determines investment incentives. It depends on three underlying parameters. With probability ϕ , the value of public goods (risk of external conflict) is high, α_H , the future is a common-interest state, and all revenue is used to supply private goods. With probability $(1 - \phi)$, the future is a redistributive state and the incumbent captures a marginal return of 2 with probability $(1 - \gamma)$, namely when it stays in power.

State Capacity Choices

The choice by incumbent group I_1 of state capacity for period 2 maximizes

$$(5) \quad E[V^{I_1}(\pi_2, \tau_2)] - \lambda_1[L(\pi_2 - \pi_1) + F(\tau_2 - \tau_1)],$$

subject to $\pi_2 \geq \pi_1$ and $\tau_2 \geq \tau_1$. Thus the choice of I_1 trades off the period-2 expected benefits against the period-1 costs of investment, given the realized value of public funds. When doing so, it takes into account the uncertainties about the future values of public goods and resource rents, as well as the prospects of government turnover.

Carrying out the maximization and using (4), we can write the first-order (complementary-slackness) conditions as

$$(6) \quad w_p(\pi_2)\{1 + \tau_2[E(\lambda_2) - 1]\} \leq \lambda_1 L_\pi(\pi_2 - \pi_1)$$

and

$$(7) \quad w(\pi_2)[E(\lambda_2) - 1] \leq \lambda_1 F_\tau(\tau_2 - \tau_1),$$

where (6) concerns legal capacity and (7) concerns fiscal capacity.

Conditions (6) and (7) reproduce, in somewhat different notation, the gist of the results in Besley and Persson (2009a). Since $L_\pi(0) = F_\tau(0) = 0$, it is easy to see that if $E(\lambda_2) > 1$, there is always positive investment in both kinds of state capacity. Moreover, in this case, fiscal and legal capacity are complements. To simplify the discussion, we focus on this case here.⁸ It will prevail as long as the probability, ϕ , of the common interest state is large enough or political instability, γ , is low enough: sufficient conditions are either $\phi > \frac{1}{2}$ or $\gamma < \frac{1}{2}$.

⁸Besley and Persson (2009a) discussed some implications of this not being the case.

Determinants of State Capacity

When $E(\lambda_2) > 1$, the left-hand side of (6) is increasing in τ_2 , while the left-hand side of (7) is increasing in π_2 . The resulting complementarity is interesting in its own right. However, it also simplifies the analysis since it implies that the payoff function (5) is supermodular. This means that we can use standard results on monotone comparative statics (see, e.g., Milgrom and Shannon (1994)). Thus, any factor that increases (decreases) the expected value of government funds $E(\lambda_2)$, for given λ_1 , will increase (decrease) investment in *both* legal and fiscal capacity. The same is true for any factor that weakly decreases (increases) the right-hand side of the two expressions for given $E(\lambda_2)$.

Using (6) and (7) together with the definition of $E(\lambda_2)$, we establish the following result:

PROPOSITION 2: *Investments in both legal and fiscal capacity increase with*

- (i) *wages (for given π),*
- (ii) *the share of national income not generated by natural resources,*
- (iii) *the expected value of public goods (risk of external conflict),*
- (iv) *the level of political stability,*
- (v) *lower costs in either type of investment (for given π or τ).*

The proof of this and subsequent results is given in the Appendix.

1.4. Implications

The first part of Proposition 2 is consistent with Figure 1, where we saw that taxation and property-rights protection are both positively correlated with income across countries. We return to the relation between legal capacity and income (growth) later on in this section.

Second, Proposition 2 suggests that investment in state capacity is declining in the share of resource rents in GDP— $R_s/Y_s = R_s/(w(\pi_s) + R_s)$ —for given Y_s . This is because we have assumed that only produced output is taxed and that legal capacity is only useful for produced output.

The third part of Proposition 2 is in line with Tilly's (1985) claim that war is important for building fiscal capacity, but extends it to legal capacity. While external defense is a natural example, the result applies to any national common-interest program, such as a universal welfare state or health program. If the demand for such public goods or services is expected to be high, any group that is in power has a greater incentive to invest in fiscal capacity to finance future common-interest spending. In the second half of the 18th century, continued state capacity building by the dominant British elite culminated in the launch of an income tax during the Napoleonic wars, when the British government could raise taxes equal to a remarkable 36% of GDP (Mathias and O'Brien (1976)).

Part (iv) of Proposition 2 holds because the incumbent group faces a smaller risk of the opposition using a larger fiscal capacity to redistribute against the incumbent. Thus, we should observe higher political stability to induce more

developed economic institutions.⁹ We know of no systematic evidence on this prediction, but a historical case in point is England after the Glorious Revolution. During a parliament dominated by the Whigs for more than 40 years, tax income rose to 20% of GDP, and institutions for charging excise and indirect taxes were put in place (see, e.g., Stasavage (2007) and O'Brien (2005)).

One interpretation of the fifth part of the proposition is a theoretical rationale for legal origins, the subject of many studies following La Porta, Lopez de Silanes, Shleifer, and Vishny (1998). If some form of legal origin, such as the common-law tradition, makes it cheaper to facilitate private contracting, then we would expect this to promote investments in the legal system. Less trivially, we would also expect the same legal origin to promote investments in the tax system, because of the complementarity of legal and fiscal capacity.

Correlations in International Data

Besley and Persson (2009a) explored the cross-sectional correlations in international data, motivated by results like Proposition 2, which identifies a number of *common* determinants of legal and fiscal capacity. First, they took the historical incidence of war as a proxy for past demand for common public goods and used data from the Correlates of War data set to measure the share of all years between 1816—or independence, if later—and 1975 that a country was involved in external military conflict. Second, they considered indicators of legal origin from La Porta et al. (1998) as proxies for the cost of legal infrastructure. To gauge current legal and fiscal capacity, they considered four different indicators of each form of state capacity, including measures of contract enforcement, protection of property rights, and various aspects of tax structure.

Besley and Persson (2009a) showed that a higher share of external conflict years in the past is always associated with higher measures of legal capacity as well as fiscal capacity in the present. While English legal origin is uncorrelated with legal capacity (except when it comes to contract enforcement), German and Scandinavian legal origins do display a robust positive correlation, not only with legal capacity but also with fiscal capacity. Key determinants identified by our theory thus appear to have stable correlations with the state's capacity to support markets as well as to raise revenue.¹⁰

⁹In their richer model, Besley and Persson (2009a) found that this effect should be stronger in countries with less inclusive political institutions. They also found that more inclusive political institutions by themselves generally promote investments in state capacity.

¹⁰In line with their more extensive model, Besley and Persson (2009a) also measured inclusive political institutions in the past by the incidence of democracy and parliamentary democracy. They found that current state capacity of both types is generally correlated with these measures of politically inclusive institutions.

Growth

Beyond these direct implications, the model makes a prediction about economic growth between periods 1 and 2. Using Proposition 1, this is given by

$$(8) \quad \frac{Y_2 - Y_1}{Y_1} = \frac{w(\pi_2) - w(\pi_1) + R_2 - R_1}{w(\pi_1) + R_1}.$$

If we ignore the exogenous resource rents, higher growth is generated solely by having higher legal capacity and hence better support for private markets. This would show up in the data as higher total factor productivity.

Legal capacity may be closely related to financial development (in the microfounded model of Besley and Persson (2009a), e.g., private credit to GDP is proportional to π). Financial development due to better institutions can thus cause growth, but the relationship can easily go the other way: According to the second part of Proposition 2, higher income generally raises incentives to invest in legal capacity, leading to financial development.

The complementarity between fiscal and legal capacity has interesting implications for the relationship between taxation and growth. If greater legal capacity is driven by the determinants suggested by Proposition 2, we would expect it go hand-in-hand with greater fiscal capacity. Variation in these determinants would tend to induce a *positive* correlation between taxes and growth. Even in the case where $E(\lambda_2) < 1$ (when investment in fiscal capacity is zero), legal capacity and national income are still positively correlated even though taxation and growth are uncorrelated.¹¹

These observations relate to recent empirical findings in the macroeconomics of development. Many researchers have found a positive correlation between measures of financial development, or property-rights protection, and economic growth (e.g., King and Levine (1993), Hall and Jones (1999), and many subsequent papers), although the first part of Proposition 2 warns us that such correlations may not reflect a causal effect of financial markets, but reverse causation. But many researchers who expected to find a negative relation between taxes and growth have found nothing (see, e.g., the overview in Benabou (1997)). Simple though it is, our model suggests a possible reason for these findings.

Our approach focuses on state capacity and hence ignores the standard engine of growth through private capital accumulation. When one extends the model to include private investment, building fiscal capacity does have a more “standard” disincentive effect on growth, because higher τ_2 raises expected taxes and lowers expected net private returns. However, building legal capacity has an additional positive effect on growth, because it can raise the gross

¹¹However, Besley and Persson (2009a) showed that changes in income distribution drive fiscal and legal capacity in opposite directions, inducing a *negative* correlation between taxes and growth.

return to investing, which stimulates private accumulation. With complementarity between fiscal and legal capacity, both kinds of state capacity may still expand with overall income.

2. CONFLICT AND STATE CAPACITY

This section extends our approach to include the possibility of violent internal conflict. We modify the model by allowing for the possibility that public and private resources are used by incumbent and opposition to maintain or gain control of the state. As a by-product of this, we endogenize political instability. In our model, conflict might arise in the redistributive state (when $\alpha_s = \alpha_L$), since this entails a greater advantage of becoming a residual claimant on public resources, including natural resource rents.

Our analysis is motivated by the observation that political instability and high risk of conflict are clustered in the data with weak states and low levels of development. Our approach based on investments in state capacity will show how all of these have common underlying roots. Moreover, the factors identified by Proposition 2 as affecting investment in state capacity play a key role in this clustering.

There now exists a large literature on conflict in the third world (see, e.g., Sambanis (2002) and Blattman and Miguel (2009) for broad reviews). Counting all countries and years since 1950, the incidence of civil war is about 6%, with a yearly peak of more than 12% (in 1991 and 1992), according to the Correlates of War data set. The cumulated death toll in civil conflicts since the Second World War exceeds 15 million (Lacina and Gledtisch (2005)). A robust empirical fact is that poor countries are disproportionately more likely to be involved in civil war. There are two leading interpretations of this correlation in the literature: Fearon and Laitin (2003) see conflict in poor countries as reflecting limited capacity to put down rebellions by weak states, while Collier and Hoeffler (2004) see it as reflecting lower opportunity costs of fighting.

The civil war literature typically treats incomes and state capacity as exogenous,¹² but the dynamic implications of conflict are likely to be important. Although our approach is simple and stylized, it offers a first step toward a dynamic approach that emphasizes the state capacity channel.¹³ The analysis will also speak to the link between natural resources, conflict, and development.¹⁴ In our model, large resource rents raise the risk of civil war and diminish the

¹²Miguel, Satyanath, and Sergenti (2004) took a step toward treating incomes as endogenous. They used weather shocks to instrument for growth in African countries from the 1980s and onward, and found that lower growth raises the probability of civil conflict.

¹³In a previous paper (Besley and Persson (2008a)) we argued that internal and external conflict may have opposite effects on the incentives to invest in fiscal capacity, but there we took the probability of civil war to be exogenous.

¹⁴See Ross (2004) for a survey of the research on natural resources and civil war.

incentive to invest in state capacity, thus creating a negative feedback loop to the level of development.

2.1. Conflict and Takeover

The key change in the model is to modify the way in which political power is transferred. As in Section 1, this may happen peacefully. However, we add the possibility that power changes hand through violent conflict. Our approach is very simple. Suppose that the incumbent can raise an army, the size of which (in per capita terms) is denoted by $\delta^{I_{s-1}} \in \{0, A^I\}$, where $0 < A^I < 1$ (recall that total population size is unity). This discrete-choice formulation, which is relaxed in Besley and Persson (2008b), is somewhat artificial, but makes the analysis simpler. There is no conscription, so soldiers must be compensated for their lost income. The army, which costs $w^{I_{s-1}} \delta^{I_{s-1}}$, is financed out of the public purse.

The opposition can also raise an army, denoted by $\delta^{O_{s-1}} \in \{0, A^O\}$, with $0 < A^O < 1$, which it uses to mount an insurgency to take over the government. When in opposition, we assume that each group has the capacity to tax its own citizens to finance a private militia. The decision on $\delta^{O_{s-1}}$ is made by the opposition group, but the resources have to be raised within the group.

The probability that group O_{s-1} wins power and becomes the new incumbent I_s is

$$\gamma(\delta^{O_{s-1}}, \delta^{I_{s-1}}) \in [0, 1].$$

This probability of turnover depends on the resources devoted to fighting. We assume that this is increasing in the first argument and decreasing in the second so that there are returns to each side from fighting. We make the following assumption on the underlying conflict technology:

ASSUMPTION 1: *The contest function satisfies*

$$\begin{aligned} & \frac{1 - \gamma(A^O, A^I)}{\gamma(A^O, 0) - \gamma(A^O, A^I)} \\ & \leq \frac{1 - \gamma(0, A^I)}{\gamma(0, 0) - \gamma(0, A^I)} \\ & < \min \left\{ 1 + \frac{\frac{A^O}{2A^I}}{\gamma(A^O, A^I) - \gamma(0, A^I)}, \frac{\frac{A^O}{2A^I}}{\gamma(A^O, 0) - \gamma(0, 0)} \right\}. \end{aligned}$$

This assumption rules out the possibility of an undefended insurgency. It will hold if the marginal return to fighting is low enough for the opposition and high enough for the incumbent.¹⁵

Given this technology for conflict, we make two substantive changes to the model described in Section 1.1. First, the government budget constraint has to be rewritten to reflect the financing of the state army. This is now¹⁶

$$(9) \quad 0 \leq \sum_{J_s \in \{I_s, O_s\}} \frac{t^{J_s} w^{J_s}}{2} - G_s + z_s - w^{I_{s-1}} \delta^{I_{s-1}}.$$

Second, Stage 3 in the timing is replaced by the following sequence:

Stage 3a. Group O_{s-1} chooses the level of any insurgency $\delta^{O_{s-1}}$.

Stage 3b. The incumbent government I_{s-1} chooses the size of its army $\delta^{I_{s-1}}$.

Stage 3c. Group I_{s-1} remains in office with probability $1 - \gamma(\delta^{O_{s-1}}, \delta^{I_{s-1}})$.

In this setting, we interpret civil war as $\delta^{O_{s-1}} = A^O$ and $\delta^{I_{s-1}} = A^I$, that is, both groups are investing in violence, while $\delta^{O_{s-1}} = 0$ and $\delta^{I_{s-1}} = A^I$ is interpreted as repression by government to stay in power.

2.2. Incidence of Civil War and Repression

Preliminaries

It is easy to show that the (new) incumbent's policy choices at Stage 4 of each period in Proposition 1 still apply. Making use of this, we can derive the government's objective function after the resolution of uncertainty over α_s and R_s at Stage 2, but prior to the choice of armies at Stage 3. For the incumbent at Stage 3b, the appropriate expression depends on the realized value of α_s and is given by

$$(10) \quad E[V^{I_{s-1}}(\pi_s, \tau_s) | \alpha_s = \alpha_H] = \alpha_H [\tau_s w(\pi_s) + z_s - w(\pi_s) \delta^{I_{s-1}}] \\ + w(\pi_s)(1 - \tau_s)$$

¹⁵The assumption is consistent with a variety of assumptions about the functional form of the "contest function." In the case of a linear model where

$$\gamma(\delta^{O_{s-1}}, \delta^{I_{s-1}}) = \gamma + \mu(A^O - A^I),$$

Assumption 1 is satisfied if

$$1 - \gamma + \mu A^I < 1/2.$$

¹⁶This formulation assumes that resource revenues are large enough to finance the incumbent's army or, alternatively, that the new incumbent pays for the army ex post, honoring any outstanding "war debts."

and

$$(11) \quad \begin{aligned} E[V^{I_{s-1}}(\pi_s, \tau_s) | \alpha_s = \alpha_L] \\ = w(\pi_s)(1 - \tau_s) + (1 - \gamma(\delta^{O_{s-1}}, \delta^{I_{s-1}})) \\ \times 2[\tau_s w(\pi_s) + z_s - w(\pi_s)\delta^{I_{s-1}}]. \end{aligned}$$

The opposition chooses its army $\delta^{O_{s-1}}$, at Stage 3a, to maximize the group's expected utility, which is given by

$$(12) \quad \begin{aligned} E[V^{O_{s-1}} | \alpha_s = \alpha_H] = \alpha_H[\tau_s w(\pi_s) + z_s - w(\pi_s)\delta^{I_{s-1}}] \\ + w(\pi_s)(1 - \tau_s - \delta^{O_{s-1}}) \end{aligned}$$

and

$$(13) \quad \begin{aligned} E[V^{O_{s-1}} | \alpha_s = \alpha_L] = \gamma(\delta^{O_{s-1}}, \delta^{I_{s-1}})2[\tau_s w(\pi_s) + z_s - w(\pi_s)\delta^{I_{s-1}}] \\ + w(\pi_s)(1 - \tau_s - \delta^{O_{s-1}}). \end{aligned}$$

The main difference between these expressions reflects the fact that the incumbent uses the government budget to finance its army, whereas the opposition uses its private resources.

We are now in a position to characterize the unique subgame perfect equilibrium of the game where the insurgents (opposition) move first. The equilibrium strategies are denoted by $\{\hat{\delta}^{O_{s-1}}, \hat{\delta}^{I_{s-1}}\}$.

Common-Interest States

We begin by stating a useful (if perhaps obvious) result in the case when demand for public goods is high:

PROPOSITION 3: *There is never conflict when $\alpha_s = \alpha_H : \hat{\delta}^{O_{s-1}} = \hat{\delta}^{I_{s-1}} = 0$.*

Intuitively, all spending in the common-interest state will be on common-interest goods, independently of who holds power, so there is nothing to fight over. Given our interpretation of α_H as (a high risk of) external conflict, it is interesting to note that very few—less than half a percent—of the country-years in the Correlates of War data set entail simultaneous external and internal conflict.

This result implies that the probability of political turnover in common-interest states is $\gamma(0, 0)$.

Redistributive States

When $\alpha_s = \alpha_L$, the situation is different. The payoffs (11) and (13) reveal a trade-off: decision makers must weigh the opportunity cost of higher armed forces against a higher probability of takeover and control over state resources.

Given Assumption 1, we get a straightforward characterization of conflict regimes by the size of public revenues and other parameters in terms of three main regimes. Define

$$Z(z_s; \pi_s, \tau_s) = \frac{\tau_s w(\pi_s) + z_s}{w(\pi_s)},$$

the ratio of total government revenue per capita to the real wage (nonresource share of GDP), as well as a lower and an upper bound for this variable:

$$\underline{Z} = \left[\frac{1 - \gamma(0, A^I)}{\gamma(0, 0) - \gamma(0, A^I)} \right] A^I$$

and

$$\overline{Z} = \frac{A^O}{[\gamma(A^O, A^I) - \gamma(0, A^I)]2} + A^I,$$

where $\overline{Z} > \underline{Z}$, by the second inequality in Assumption 1. We now have the following statement.

PROPOSITION 4: *Suppose that Assumption 1 holds and $\alpha_s = \alpha_L$ (a redistributive state). Then there are three possibilities:*

- (i) *If $Z(z_s; \pi_s, \tau_s) > \overline{Z}$, then there is civil conflict with $\widehat{\delta}^{O_{s-1}} = A^O$ and $\widehat{\delta}^{I_{s-1}} = A^I$.*
- (ii) *If $\underline{Z} \leq Z(z_s; \pi_s, \tau_s) \leq \overline{Z}$, then the state is repressive with $\widehat{\delta}^{O_{s-1}} = 0$ and $\widehat{\delta}^{I_{s-1}} = A^I$.*
- (iii) *If $Z(z_s; \pi_s, \tau_s) < \underline{Z}$, then there is peace with $\widehat{\delta}^{O_{s-1}} = 0$ and $\widehat{\delta}^{I_{s-1}} = 0$.*

If $Z(z_s; \pi_s, \tau_s)$ is very high, which corresponds to low wages (low π_s), high fiscal capacity, or high natural resource rents, then the outcome is conflict because it is cheap to fight and there is a large cake to redistribute for the winner. If $Z(z_s; \pi_s, \tau_s)$ is in an intermediate range, then the government represses the opposition to increase the probability that it stays in power. Finally, if $Z(z_s; \pi_s, \tau_s)$ is low enough, then there is peace.¹⁷ The main role of Assumption 1 is to rule out an undefended insurgency. While this is a theoretical possibility, such cases do not seem common in practice.

Proposition 4 gives a link between natural resource rents, real wages, and the likelihood of conflict. For given state capacities (π_s, τ_s) , variable Z_s varies

¹⁷The parameter restriction in Assumption 1 is the reason that the ordering is straightforward. In Besley and Persson (2008b), we also obtained an ordering result of this form (under weaker assumptions) in a related model where the choice of armies is continuous and institutions constrain the behavior of the incumbent and the opposition ex post. For some parameter restrictions, it is possible to have an outcome where the government does not defend against an insurgency (passive acceptance of terrorism).

stochastically with natural resource rents R_s and real wages w_s . By this route, we expect commodity prices to predict civil war. Besley and Persson (2008b) explored the empirical link between commodity prices and the incidence of civil conflict. Using trade volume data from the NBER-UN Trade data set and international price data for about 45 commodities from UNCTAD, they constructed country-specific commodity export and commodity import price indexes for about 125 countries since 1960.¹⁸ According to the open-economy model in Besley and Persson (2008b), higher export price index can be interpreted as a positive shock to natural resource rents, and a higher import price index can be interpreted as a negative shock to (real) income. In line with Proposition 4, they found a robust empirical link between these price indexes and the incidence of civil war.

Proposition 4 also suggests that government repression and civil war may reflect the same underlying determinants, namely resource rents and real wages. Indeed, the proposition suggests that the regimes of peace, repression, and civil war can be looked upon as ordered states. Interpreting government repression as infringements on human rights, Besley and Persson (2009b) pushed this argument further and estimated the likelihood of observing these states as an ordered probit.

2.3. Investment in State Capacity

The analysis in the previous subsection takes legal and fiscal capacity as given. We now explore the implications of conflict for the incentive to invest in state capacity.

When there is no risk of future civil war, the analysis in Section 1.3 applies with $\gamma(0, 0) = \gamma$. To highlight the new mechanisms added by the possibility of conflict, we assume that the period-1 incumbent knows for sure that the value of public goods in the future is low (i.e., that $\alpha_2 = \alpha_L$). Except for the issue of incumbency, the only remaining uncertainty—and the only determinant of the risk of conflict—then concerns the level of natural resource rents.

There are two new effects on state capacity investment beyond those found in the nonconflict model of Section 1. The first of these comes from observing that conflict changes the probability that the incumbent group will stay in power and hence affects political instability. To see this formally, we can use the result in Proposition 4 to write the equilibrium probability of turnover as

$$\Gamma(Z(R_2; \pi_2, \tau_2)) = \begin{cases} \gamma(A^O, A^I), & \text{if } Z(R_2; \pi_2, \tau_2) > \bar{Z}, \\ \gamma(0, A^I), & \text{if } Z(R_2; \pi_2, \tau_2) \in [\underline{Z}, \bar{Z}], \\ \gamma(0, 0), & \text{if } Z(R_2; \pi_2, \tau_2) < \underline{Z}. \end{cases}$$

¹⁸The price indexes for a given country have fixed weights, computed as the share of exports and imports of each commodity in the country's GDP in a given base year.

The constituent probabilities depend on the exogenous level of resource rents and the endogenous levels of state capacity. Note that the probability of turnover is not monotonic in natural resource rents: survival is largest in the middle range where the government represses the opposition. Whether outright conflict increases political instability is not clear *a priori*: this depends on whether the government is more or less likely to survive in the conflict regime compared to peace (i.e., $\gamma(A^O, A^I) \geq \gamma(0, 0)$). Given our observation in Proposition 2 that political stability affects investments in state capacity, this makes it unlikely that there is any general proposition linking conflict and state development working through this channel. Thus, to wash this effect out and home in on other considerations, we will make another assumption.

ASSUMPTION 2: $\gamma(A^O, A^I) \approx \gamma(0, 0)$.

One corollary of this assumption is that conflict is clearly Pareto inefficient with resources being spent without any material change (*ex ante*) in who holds power.

The second effect of adding conflict to the model comes from the fact that the incumbent government has to pay the real market wage to employ the soldiers in its army. Thus, incumbents may be more reluctant, all else equal, to raise incomes by investing in legal capacity (or any other institution raising the wage).

We consider two cases. In Case 1, a country cycles between peace and civil war, whereas in Case 2 it cycles between repression and civil war.

CASE 1— $Z(R_H; \pi_2, \tau_2) > \bar{Z} > \underline{Z} > Z(R_L; \pi_2, \tau_2)$: Suppose the prize from winning a conflict is high enough for both incumbent and opposition to arm when resource rents are high, whereas neither of them arms when resource rents are low. Implicitly, we thus assume that variations in investment in fiscal capacity τ_2 are never large enough to induce changes in the conflict regime.

Under these assumptions, and following the same approach as in Section 1, we can write the payoff of the period-1 incumbent controlling the state-capacity investment decisions as

$$\begin{aligned} E[V^{I_s}(\pi_2, \tau_2) | \alpha_s = \alpha_L] \\ = w(\pi_2)(1 - \tau_2) \\ + E(\lambda_2)[\tau_2 w(\pi_2) + E(z_2)] - \rho[1 - \gamma(A^O, A^I)]2w(\pi_2)A^I, \end{aligned}$$

where the expected value of future government funds is given by $E(\lambda_2) = [1 - ((1 - \rho)\gamma(0, 0) + \rho\gamma(A^O, A^I))]2$. As in Section 1, we focus on the case where $E(\lambda_2) > 1$ so that investments in both kinds of state capacity remain

complements.¹⁹ Compared to our earlier expression (4) in the baseline (no-conflict) model in Section 1, the objective function has a new and third term, which captures the cost of conflict. That this term is multiplied by ρ reflects the fact that conflict occurs only when resource rents are high.

The first-order conditions for investments in legal and fiscal capacity are

$$(14) \quad w_p(\pi_2) [\{1 + \tau_2[E(\lambda_2) - 1]\} - \rho[1 - \gamma(A^O, A^I)]2A^I], \\ \leq \lambda_1 L_\pi(\pi_2 - \pi_1),$$

$$(15) \quad w(\pi_2)[E(\lambda_2) - 1] \leq \lambda_1 F_\tau(\tau_2 - \tau_1).$$

When Assumption 2 holds, the probability of conflict, ρ , has a negligible effect on the expected value of public funds, $E(\lambda_2)$. Then the only first-order effect on investments of a higher probability of conflict comes from the second term on the left-hand side of (14). Evidently, a higher ρ reduces the marginal return to investing in legal capacity, since a higher share of the economy's labor is expected to be devoted to conflict. Taking the complementarity between fiscal and legal capacity into account, we now have another result.

PROPOSITION 5: *Suppose that the future state is always redistributive ($\alpha_2 = \alpha_L$), there is either conflict or peace depending on the level of natural resource rents, and that Assumption 2 holds. Then an exogenous increase in the probability of conflict, via a higher value of ρ , reduces the incentive to invest in both fiscal and legal capacity.*

Proposition 5 illustrates a particular channel through which the static inefficiency of conflict is compounded by a dynamic inefficiency via a lower incentive to invest in state capacity: investing in economic development makes it more expensive for the government to finance its troops should a conflict arise. This highlights a specific mechanism through which conflict risk perpetuates a weak state.

CASE 2— $Z(R_H; \pi_2, \tau_2) > \bar{Z} > Z(R_L; \pi_2, \tau_2) > \underline{Z}$: In this case, changes in resource rents cycle the economy between repression and civil war; the incumbent always finds it optimal to arm, while the opposition arms only when resource rents are high. In this instance, the probability of high resource rents, ρ , has a direct effect on the expected probability of turnover for the period-1 incumbent, even if Assumption 2 does not hold.

¹⁹Note, however, that an increase in ρ (now the probability of conflict since conflict occurs when natural resource rents are high) may increase or decrease the future expected value of public funds. Depending on the relative values of A^I and A^O , this can raise or cut the likelihood that state capacities are substitutes rather than complements.

Now, the expected payoff to the incumbent is

$$(16) \quad E[V^{I_s}(\pi_2, \tau_2) | \alpha_s = \alpha_L] = w(\pi_2)(1 - \tau_2) \\ + E(\lambda_2)[\tau_2 w(\pi_2) + E(z_2) - w(\pi_2)A^I],$$

where $E(\lambda_2) = \{1 - [(1 - \rho)\gamma(0, A^I) + \rho\gamma(A^O, A^I)]\}/2$. After some manipulation, the first-order conditions for investing in state capacity become

$$(17) \quad w_p(\pi_2)\{(1 - A^I) + (\tau_2 - A^I)[E(\lambda_2) - 1]\} \leq \lambda_1 L_\pi(\pi_2 - \pi_1),$$

$$(18) \quad w(\pi_2)[E(\lambda_2) - 1] \leq \lambda_1 F_\tau(\tau_2 - \tau_1).$$

Note that the condition for positive investments in legal capacity—namely a positive left-hand side of (17)—may now be stronger than $E(\lambda_2) > 1$. Clearly, $E(\lambda_2) > 1$ together with $\tau_2 > A^I$ is a sufficient condition. In fact, the term $(1 - A^I)w_p(\pi_2)$ always represents a drag on investment in state capacity similar to the effect identified in Case 1.

Assuming that this condition is met, we can contemplate the effect of a change in ρ on the incentive to invest. From the first-order conditions, we have the following proposition:

PROPOSITION 6: *Suppose that the future state is always redistributive ($\alpha_2 = \alpha_L$) and there is either conflict or repression, depending on the level of natural resource rents. Then an increase in the exogenous probability of conflict, via a higher value of ρ , reduces the marginal incentive to invest in both fiscal and legal capacity.*

The result in Proposition 6 follows by complementarity and by noting that a higher value of ρ decreases the left-hand side of both (17) and (18), the latter because $\frac{\partial \text{LHS}}{\partial \rho} = -2w_p(\pi_2)[\gamma(A^O, A^I) - \gamma(0, A^I)](\tau_2 - A^I) < 0$, where the sign follows from the foregoing sufficient condition for positive investment. Intuitively, the direct effect through the probability of survival always outweighs the effect through the expected value of fighting.

This result is analogous to part (iv) of Proposition 2, whereby higher political instability reduces investments in state capacity. However, the instability is now modeled as an equilibrium outcome, where conflict (relative to repression) makes it less likely that the incumbent survives. This result gives a further theoretical explanation as to why the prospect of conflict might perpetuate weak states in both raising taxes and supporting markets.

2.4. Implications

Propositions 5 and 6 highlight two key mechanisms through which the possibility of civil conflict may perpetuate weak states, with lower levels of income as a consequence. Our examples have focused on marginal incentives within a regime (corresponding to the maintained assumptions defining our two cases).

[Proposition 4](#) defines a threshold for wages relative to resource rents above which conflict ends. Because of this, a government may strive for a big enough investment in legal capacity to raise wages so as to generate peace. To the extent that this is important, we might expect incentives to go in the opposite direction of those driving the results in [Propositions 5](#) and [6](#). There may then be scope for a “big push” to raise wages and to break out of the conflict trap.

The results also suggest a note of caution for researchers who pursue empirical studies of the determinants of civil war. Our model shows why it may be hazardous to interpret the correlation between poverty and civil war as a causal effect from poverty to the incidence of conflict. Indeed, the results in this section imply that both of the two leading explanations of this correlation—low opportunity cost of fighting due to low wages and low state capacity in poor countries—may reflect common omitted factors rather than a causal mechanism. In particular, low state capacity in terms of raising tax revenue, as well as low wages (due to poor support of markets), may be simultaneously determined with a high probability of civil war by factors such as high resource rents.

Finally, the state-capacity channel developed here also provides a theoretical connection between conflict and low growth. This is apparent by returning to equation (8), which links low investment in π_2 to low growth.

3. STATE CAPACITY, DISTORTIONS, AND INCOME

We now explore the link between state capacity investments—particularly investment in fiscal capacity—and policy distortions which lower the level of income. We show how investments in fiscal capacity can underpin efficiency-enhancing changes in the *form* of redistribution, diminishing the use of other “regulatory” distortions which make the economy less productive. We provide an example in which a government with insufficient fiscal capacity chooses legal protection in an inefficient way. While this general point has been made before (for example, by [Acemoglu \(2006\)](#)), this takes state capacity as given. We show that the production inefficiencies may persist over time when state capacity is chosen endogenously because the economy may be caught in an investment trap. The apparatus developed in [Section 1](#) explains the factors that underpin this.

This analysis of the role of state capacity in encouraging efficient production provides a unique window on debates about the consequences of large government for the economy. As we noted in [Section 1.4](#), it is hard to find evidence in macroeconomic studies of aggregate data that high taxes affect the growth rate. Most microeconomic studies of individual data also tend to find fairly modest behavioral effects of taxes on investment behavior. The mechanism that we identify here whereby fiscal capacity increase production efficiency may constitute an important offsetting effect of increasing the power to tax. Our approach also provides an alternative to the standard macroeconomic view of government’s role in enhancing growth, as exemplified by [Barro \(1990\)](#)

and Barro and Sala-i-Martin (1992), who emphasize the role of tax-financed public capital accumulation, such as building ports and roads.

To make these points as simply as possible, we drop the extension to endogenous conflict in Section 2. Instead, we extend the basic framework of Section 1 in a different direction, adding an additional factor of production so that the model includes both labor and capital. Capital becomes a source of producer rents, and it is the seeking of these rents that can generate persistent production inefficiencies when the economy is caught in an investment trap for state capacity. This way, we illustrate another mechanism that may generate a link between low income and low state capacity.

3.1. A Simple Two-Factor Economy

We modify the production side of the economy to have two factors of production. Suppose now that $w(p^{I_s})$ is a form of capital, the productivity of which depends on property-rights protection for group J in period s . A share of each group, denoted by σ , are entrepreneurs and have access to a constant-returns Cobb–Douglas technology that combines capital and raw labor, l , to produce output. The capital share is denoted by η .²⁰ The remaining $1 - \sigma$ share of the population supplies a single unit of raw labor to an economy-wide labor market. The production technology on intensive form is $l^{I_s}(k^{J_s})^\eta$, where k^{J_s} is the capital-to-labor ratio $w(p^{I_s})/l^{I_s}$. Since aggregate labor supply is $l = (1 - \sigma)$, the aggregate capital–labor ratio

$$(19) \quad k(p^{I_s}, p^{O_s}) = \frac{\sigma[w(p^{I_s}) + w(p^{O_s})]}{2(1 - \sigma)}$$

is increasing in the property-rights protection of *each* group. An individual capital owner in group J_s sets optimal labor demand according to the condition $(1 - \eta)(k^{J_s})^\eta = \omega$, where ω is the economy-wide wage. Because the technology is common across groups, the equilibrium wage is given by the same condition, evaluated at $k(p^{I_s}, p^{O_s})$:

$$(1 - \eta)(k(p^{I_s}, p^{O_s}))^\eta = \omega(p^{I_s}, p^{O_s}).$$

Thus, the wage depends on property-rights protection for the two groups and is increasing in both of these policy variables, since

$$\frac{\partial \omega}{\partial p^{J_s}} = (1 - \eta)\eta(k(p^{I_s}, p^{O_s}))^{\eta-1} \frac{\sigma w_p(p^{J_s})}{2(1 - \sigma)} > 0.$$

²⁰Assuming a common share σ across groups simplifies the algebra. Relaxing this assumption makes it easier to prove the possibility of inefficient outcomes (see Propositions 3 and 4). An incumbent group, I , with a large share σ^I of capital owners is more willing to select inefficient policies to boost the group's rents than is a group with a small share.

Intuitively, more productive capital in any sector drives up the demand for labor, which raises the equilibrium wage.

Finally, we can define the income of a representative member of group J_s as

$$(20) \quad y^{J_s}(p^{I_s}, p^{O_s}) = (1 - \sigma)\omega(p^{I_s}, p^{O_s}) + \sigma l^{I_s}[(k^{J_s})^\eta - \omega(p^{I_s}, p^{O_s})],$$

the sum of labor and rental income. Compared to the basic model, the income of group J_s now depends on the legal protection of the other group as well, through the endogenous equilibrium wage. The latter has a positive effect on wage-earning group members (the first term on the right-hand side of (20)), but a negative effect on those earning quasi-rents on capital (the second term on the right-hand side).

3.2. Policy and State Capacity

The remainder of the model works exactly as in Section 1. To analyze the incumbent's optimal policy, we replace $w(p^{I_s})$ in (3) by the new income function $y^{J_s}(p^{I_s}, p^{O_s})$ in (20). The main consequence is that if σ is high enough, then an incumbent group I_s may prefer to keep wages low. Moreover, the ruling group can engineer a lower wage by blocking the opposition group's access to legal capacity and hence driving down the demand for labor.

The Role of Taxation

Going through similar steps as in Section 1.2, we can show another proposition.

PROPOSITION 7: *If $\tau_s = 1$, then legal capacity is always fully utilized for both groups. Otherwise, there exists a threshold value $\hat{\tau}_K$ when the value of the public good is α_K with $K \in \{L, H\}$, such that the legal protection of the opposition group is minimal: $p^{O_s} = 0$ for all $\tau_s < \hat{\tau}_K$. Moreover, $\hat{\tau}_L > \hat{\tau}_H$.*

This result says that there is always production efficiency when fiscal capacity is high enough. However, when fiscal capacity is below a critical threshold, an incumbent may prefer an inefficient policy which lowers the level of national income. In this specific example, maximizing (gross) income and using the tax system for redistribution may be less useful to the incumbent than distorting production and raising quasi-rents by maintaining a supply of low-wage labor.²¹

²¹There is an analogy here with Diamond and Mirrlees (1971) who argued that production efficiency is desirable if a tax system is sufficiently rich. One of the assumptions required in their framework is that there be 100% taxation of pure profits. In our model, all income is taxed at the same rate and hence $\tau_s = 1$ is effectively equivalent to full taxation of pure profits (the rents on capital).

Proposition 7 also states that the critical threshold for fiscal capacity to generate an efficient use of legal capacity is lower in the common interest state than in the redistributive state.²²

The observation that limited powers to use taxation for redistribution can lead to distorted factor markets is not new. In particular, this line of argument was developed by Acemoglu (2006). However, to provide a complete explanation, we need to understand why the state lacks the power to tax. This can be addressed only if fiscal capacity is endogenous as it is in the approach taken here.

An Investment Trap for Fiscal Capacity?

The results in Section 1, particularly Proposition 2, give us a stepping stone for the analysis. We now apply this logic to understand why fiscal capacity τ can remain low (below the threshold required for production efficiency). Our key result is the following:

PROPOSITION 8: *Suppose that $\tau_1 < \hat{\tau}_L$. Then, for ϕ close enough to zero, in a range of $\gamma > 1/2$, $\tau_2 = \tau_1$ and investment in legal capacity is lower than it would be if $\tau_1 > \hat{\tau}_L$.*

An immediate corollary of Propositions 7 and 8 is that whenever initial fiscal capacity fulfills $\tau_1 < \hat{\tau}_L$, the opposition group in each period is not fully protected by the legal system. When political instability is high, the incumbent in period 1 does not want to expand the ability to tax, because it fears that such ability will be used to redistribute against its own group. As a result of the weak state, any period-2 incumbent uses inefficient legal protection to generate rents to the capital owners of its own group.

Proposition 8 thus describes an “investment trap” in state capacity. Political instability makes an incumbent group expect that larger state capacity will be used against its interests. That expectation perpetuates an ineffective apparatus for raising taxes, which then causes inefficiencies in production. The situation persists because the probability of the common-interest state ($\alpha_s = \alpha_H$) is low.

²²A previous version of this paper (Besley and Persson (2009c)), included the inclusiveness of political institutions, parametrized by θ as in Besley and Persson (2009a). In that richer setting, the critical threshold for fiscal capacity also depends on institutions, with a lower threshold for more inclusive institutions. Moreover a utilitarian planner would always choose full protection for both groups.

3.3. Implications

These results have implications for growth rates and the level of income. To see this, define the nonresource part of GDP as

$$Y_s = Y(p^{I_s}, p^{O_s}) = \frac{y^{I_s}(p^{I_s}, p^{O_s}) + y^{O_s}(p^{I_s}, p^{O_s})}{2}.$$

With an inefficient regulatory policy in period s , income becomes $Y(\pi_s, 0)$, where, by symmetry, $Y(\pi_s, 0) = Y(0, \pi_s)$. This is clearly lower than the level with efficient legal protection $Y(\pi_s, \pi_s)$.

Consider two economies S and L , where Propositions 7 and 8 apply. Assume the same initial legal capacity $\pi_1^S = \pi_1^L = \pi_1$ prevails in both, but $\tau_1^S < \hat{\tau}(\alpha_L)$ and $\tau_1^L > \hat{\tau}(\alpha_L)$ so that the economies find themselves at opposite sides of the fiscal-capacity threshold, because of different initial fiscal capacities, $\tau_1^S < \tau_1^L$.

Let us compare income levels in periods 1 and 2. By Proposition 7,

$$Y_1^L - Y_1^S = Y(\pi_1, \pi_1) - Y(\pi_1, 0) > 0,$$

that is, in period 1, economy S has a lower income level due to the inefficient legal protection of the opposition group. As the conditions in Proposition 8 hold, we have

$$Y_2^L - Y_2^S = Y(\pi_2^L, \pi_2^L) - Y(\pi_2^S, 0) > Y(\pi_1, \pi_1) - Y(\pi_1, 0),$$

where the inequality follows from the fact that $\pi_2^L > \pi_2^S$. Due to its low fiscal capacity, economy S pursues a policy of less efficient legal protection than economy L in period 2, whichever group is in power. But Proposition 8 tells us that economy S has also invested less in legal capacity than economy L . The larger state not only has the higher GDP level, but its income advantage to the smaller state is growing over time.

These implications of Proposition 7 and 8 suggest another possible interpretation of the correlations in Figure 1. Using the results in Section 1, we may observe a weak government together with low income because the two are jointly determined by other factors or because low income causes weak government (recall Proposition 2). The results in this section suggest that a weak state can actually *cause* low income, to the extent that it encourages policies that distort production.²³

It is interesting to think about ways out of inefficient legal protection in an investment trap. Propositions 7 and 8 suggest that political reform as well as exogenous circumstance may play a role. Reform that diminished political instability (lower value of γ) may induce first-period investment.²⁴ Circumstance,

²³Of course, our caveat noted above about not considering tax distortions still applies.

²⁴In the richer model of Besley and Persson (2009c), political reform that increased the inclusiveness of political institutions may achieve the same goal.

such as a higher likelihood or expected severity of external conflict (higher ϕ or α_H), may make it too costly to pursue inefficient legal protection by raising the prospect of a future common-interest state.

Let us also relate the results to some recent work on the political origins of financial development, which argues that a desire to create or preserve rents can prevent a ruling elite from building the institutions needed for well functioning financial markets (see, e.g., Rajan and Zingales (2003) or Pagano and Volpin (2005)). This work generally considers the financial sector without reference to the tax system. Hence, the political-origins argument may implicitly assume a lack of fiscal capacity, which makes it unattractive for the incumbents to invest in private markets, maximize income, and instead carry out its desired redistribution via taxes and transfers. As stressed by Acemoglu (2003, 2006), it is important to pose the political Coase theorem question explicitly, and our analysis here suggests a new way to do so. But the key innovation is to think of both aspects of state capacity as evolving endogenously together and influencing policy incentives.

We believe that the argument is much more general than the specific example in this section. Further research might consider the joint determination of weak states and other policy-induced production distortions that lead to low income, such as tariffs or red-tape regulation.

4. FINAL REMARKS

In politics, history, and sociology, state capacity is viewed as an important object of study. We have illustrated some simple ways to bring the study of state capacity and its determinants into mainstream economics.

In the development community, a lack of state capacity as manifested in weak states is often cited as a major obstacle to development. We have shown that low legal capacity can be conducive to lackluster economic growth (in Section 1) or might contribute (through wages) to the likelihood of civil war (in Section 2), and that lack of fiscal capacity can yield (through production distortions) low income (in Section 3). These observations make it essential to understand, therefore, where low state capacities come from and all three sections discuss the factors that shape investment incentives.

Our analysis also suggests an important complementarity between these two forms of state capacity. Such complementarity is a natural way to think about the clustering of institutions that appears to be a common feature of weak and strong states at different levels of economic development.

A few common themes emerge from our analysis in Sections 1–3. First, the level of economic development at a point in time affects policy outcomes, but also feeds dynamic state development. Second, realized and prospective shocks to resource rents and public-good preferences have both static and dynamic effects on policies, as well as state development. Third, we have made a distinction between circumstances where the state is mainly used to pursue

common-interest goals and where it is mainly used to redistribute income, and have shown how this distinction between common-interest and redistributive states help us understand why (threats of) external and internal conflict have opposite effects on the incentives to invest in state institutions. These themes, together with the complementarity of state capacities, help us understand why some states stay weak while others grow strong, and why we find weak states mainly at low levels of income.

Although our theory has already helped us approach the data in novel ways, the model variations we have presented are very simple. To better understand the long-run forces of development, it would be valuable to add private capital accumulation and a full-fledged dynamic framework. Another natural extension would be to introduce and endogenize political institutions. Given the history of today's developed states, it is a reasonable conjecture—in line with some work in political science—that demand for more representative government increases with state capacity. This suggests another complementarity, that between political and economic institutions, a possibility which deserves further study.

Its simplicity notwithstanding, we view the research presented here as a first step toward disentangling some of the complex interactions between state capacity, conflict, and development.

APPENDIX

PROOF OF PROPOSITION 2: Part (i) refers to a multiplicative upward shift of the wage function $w(\cdot)$, as this raises both $w(\pi_2)$ and $w_p(\pi_2)$ for any given π . Part (iii) follows from $(\partial E(\lambda_2))/\partial \phi = \alpha_H - 2(1 - \gamma) > 0$, and part (iv) follows from $(\partial E(\lambda_2))/\partial \gamma = -(1 - \phi)2$. Finally, part (v) refers to a multiplicative downward shift of either cost function $L(\cdot)$ or $F(\cdot)$. *Q.E.D.*

PROOF OF PROPOSITION 3: The relevant objective functions when $\alpha_s = \alpha_H > 2$ ((10) and (12)), are strictly decreasing in $\delta^{I_{s-1}}$ and $\delta^{O_{s-1}}$, respectively. *Q.E.D.*

PROOF OF PROPOSITION 4: First, observe that (by (11)) the incumbent will set $\delta^{I_{s-1}} = A^I$ if

$$(\gamma(\delta^{O_{s-1}}, 0) - \gamma(\delta^{O_{s-1}}, A^I))Z(z_s; \pi_s, \tau_s) \geq A^I(1 - \gamma(\delta^{O_{s-1}}, A^I)).$$

If $\delta^{O_{s-1}} = 0$, this condition holds by the definition of \underline{Z} . If $\delta^{O_{s-1}} = A^O$, the condition for $\delta^{I_{s-1}} = A^I$ can be written

$$Z \geq \frac{1 - \gamma(A^O, A^I)}{\gamma(A^O, 0) - \gamma(A^O, A^I)} A^I.$$

Since the expression on the right-hand side is smaller than \underline{Z} by the first part of the inequality in Assumption 1, whenever $Z(z_s; \pi_s, \tau_s) \geq \underline{Z}$, it is optimal for I to set $\delta^{I_{s-1}} = A^I$ independently of what O does.

Next, we show that if $Z(z_s; \pi_s, \tau_s) \geq \underline{Z}$, so that $\delta^{I_{s-1}} = A^I$, then $\delta^{O_{s-1}} = A^O$. From (13), this requires

$$[\gamma(A^O, A^I) - \gamma(0, A^I)](Z - A^I) \geq A^O,$$

which is equivalent to $Z \geq \bar{Z}$. We also need to show that when $Z < \underline{Z}$, then indeed $\delta^{O_{s-1}} = 0$. By (13), the condition is

$$Z < \frac{A^O}{2(\gamma(A^O, 0) - \gamma(0, 0))}.$$

Evaluated at the left-hand side maximum \underline{Z} , the condition becomes

$$\frac{1 - \gamma(0, A^I)}{\gamma(0, 0) - \gamma(0, A^I)} A^I < \frac{A^O}{2(\gamma(A^O, 0) - \gamma(0, 0))},$$

which is fulfilled by the second part of the inequality in Assumption 1. Moreover, the second part of the inequality in Assumption 1 also implies $\bar{Z} > \underline{Z}$. Hence, the above argument rules out the possibility of an undefended insurgency and Proposition 4 follows. *Q.E.D.*

PROOF OF PROPOSITION 7: To prove Proposition 7, first observe that

$$l^{I_s} = \frac{w(p^{I_s})2(1-\sigma)}{[w(p^{I_s}) + w(p^{O_s})]\sigma}.$$

Hence, for all $p^{I_s} > p^{O_s}$

$$\begin{aligned} \frac{\partial y^{I_s}(p^{I_s}, p^{O_s})}{\partial p^{I_s}} \\ = & \left\{ \left[\frac{[(1-\sigma) - \sigma l^{I_s}]}{2(1-\sigma)} \eta + 1 \right] \right. \\ & \times (1-\eta)(k(p^{I_s}, p^{O_s}))^{\eta-1} \sigma w_p(p^{I_s}) \Big\} \\ = & \left\{ \left[\left[\frac{1}{2} - \frac{w(p^{I_s})}{[w(p^{I_s}) + w(p^{O_s})]} \right] \eta + 1 \right] \right. \\ & \times (1-\eta)(k(p^{I_s}, p^{O_s}))^{\eta-1} \sigma w_p(p^{I_s}) \Big\} > 0 \end{aligned}$$

and

$$\begin{aligned} \frac{\partial y^{I_s}(p^{I_s}, p^{O_s})}{\partial p^{O_s}} &= \left\{ \frac{[(1-\sigma) - \sigma l^{I_s}]}{2(1-\sigma)} \eta(1-\eta) k(p^{I_s}, p^{O_s}) \sigma w_p(p^{O_s}) \right\} \\ &= \left\{ \left[\frac{1}{2} - \frac{w(p^{I_s})}{[w(p^{I_s}) + w(p^{O_s})]} \right] \right. \\ &\quad \left. \times \eta(1-\eta) k(p^{I_s}, p^{O_s}) \sigma w_p(p^{O_s}) \right\} < 0. \end{aligned}$$

Thus, there is a conflict of interest between creating property rights for the ruling group and the nonruling group.

In general, we can write the part of the government's objective function that depends on (p^{I_s}, p^{O_s}) as

$$V^{I_s}(p^{I_s}, p^{O_s}; \psi) = \psi y^{I_s}(p^{I_s}, p^{O_s}) + y^{O_s}(p^{I_s}, p^{O_s}),$$

where

$$\psi = \psi(\tau, \alpha) = \begin{cases} \frac{1 + \tau \left(\frac{\alpha}{2} - 1 \right)}{\tau \left(\frac{\alpha}{2} \right)}, & \text{if } \alpha \geq 2, \\ \frac{1}{\tau}, & \text{otherwise.} \end{cases}$$

It is easy to check that $\psi(\tau, \alpha)$ is decreasing in τ and also decreasing in α if $\alpha \geq 2$. Moreover, as $\tau \rightarrow 1$, $\psi \rightarrow 1$ and as $\tau \rightarrow 0$, $\psi \rightarrow \infty$ (independently of the value of α). In general, the condition for choosing p_s^J is

$$\psi \frac{\partial y^{I_s}(p^{I_s}, p^{O_s})}{\partial p_s^J} + \frac{\partial y^{O_s}(p^{I_s}, p^{O_s})}{\partial p_s^J} \stackrel{>}{\leq} 0.$$

Observe that

$$\begin{aligned} \frac{\partial [y^{I_s}(p^{I_s}, p^{O_s}) + y^{O_s}(p^{I_s}, p^{O_s})]}{\partial p_s^J} &= (1-\eta)(k(p^{I_s}, p^{O_s}))^\eta \sigma w_p(p^{J_s}) \\ &> 0. \end{aligned}$$

From this, we conclude that as $\tau \rightarrow 1$ and $\psi \rightarrow 1$, $p^{I_s} = p^{O_s} = \pi_s$, that is, production efficiency obtains, since the incumbent maximizes total income $y^{I_s}(p^{I_s}, p^{O_s}) + y^{O_s}(p^{I_s}, p^{O_s})$. Moreover, as $\tau \rightarrow 0$ and $\psi \rightarrow \infty$ the incumbent

maximizes its own group's income $y^{I_s}(p^{I_s}, p^{O_s})$, such that $p^{I_s} = \pi_s$ and $p^{O_s} = 0$. The existence of the critical threshold now follows from the intermediate value theorem, given that $\psi(\alpha)$ is continuous in τ for any value of α . When $\alpha = \alpha_L$, the threshold value is given by $\frac{1}{\hat{\tau}_L} \in [1, \infty)$ with $\hat{\tau}_H$ defined by

$$\hat{\tau}_H = \left[\left(\frac{1 - \hat{\tau}_L}{\hat{\tau}_L} \right) \frac{\alpha_H}{2} + 1 \right]^{-1} < \hat{\tau}_L,$$

since $\alpha_H > 2$, as claimed.

Q.E.D.

PROOF OF PROPOSITION 8: To prove the proposition, we note some useful preliminaries. It is straightforward to check that the income function is

$$\hat{y}^{I_s}(\pi_s, \alpha) = \left[(1 - \sigma)(1 - \eta) + \sigma \frac{\omega_s^I(p_s^I(\pi_s, \alpha))}{\hat{k}(\pi_s, \alpha)} \right] (\hat{k}(\pi_s, \alpha))^{\eta},$$

where $\hat{k}(\pi_s, \alpha) = k(p_s^I(\pi_s, \alpha), p^{O_s}(\pi_s, \alpha))$. Observe that

$$\frac{\hat{y}^{I_s}(\pi_s, \alpha) + \hat{y}^{O_s}(\pi_s, \alpha)}{2} = (1 - \sigma)(\hat{k}(\pi_s, \alpha))^{\eta}.$$

Now let $k_H = \hat{k}(\pi_s, \pi_s) > \hat{k}(\pi_s, 0) = k_L = k_H/2$.

The incumbent maximizes the expected period-2 benefits

$$\begin{aligned} \Gamma(\pi_2, \tau_2) &= (1 - \gamma) \left\{ \begin{array}{l} \phi \left(\left(1 + \tau_2 \left(\frac{\alpha_H}{2} - 1 \right) \right) \hat{y}^{I_2}(\pi_2, \alpha_H) \right) \\ \quad + \left(\tau_2 \left(\frac{\alpha_H}{2} \right) \right) \hat{y}^{O_2}(\pi_2, \alpha_H) \\ \quad + (1 - \phi)[\hat{y}^{I_2}(\pi_2, \alpha_L) + \tau_2 \hat{y}^{O_2}(\pi_2, \alpha_L)] \end{array} \right\} \\ &\quad + \gamma \left\{ \begin{array}{l} \phi \left(\left(1 + \tau_2 \left(\frac{\alpha_H}{2} - 1 \right) \right) \hat{y}^{O_2}(\pi_2, \alpha_H) \right) \\ \quad + \left(\tau_2 \left(\frac{\alpha_H}{2} \right) \right) \hat{y}^{I_2}(\pi_2, \alpha_H) \\ \quad + (1 - \phi)[1 - \tau_2] \hat{y}^{O_2}(\pi_2, \alpha_L) \end{array} \right\} \end{aligned}$$

less the investment costs in period 1. As $\phi \rightarrow 0$, the marginal benefits with regard to the two choice variables are

$$\Gamma_\tau(\pi_2, \tau_2) = (1 - 2\gamma) \hat{y}^{O_2}(\pi_2, \alpha_L)$$

and

$$\Gamma_\pi(\pi_2, \tau_2) = (1 - \gamma)\hat{y}_\pi^{I_2}(\pi_2, \alpha_L) + [\gamma + \tau_2(1 - 2\gamma)]\hat{y}_\pi^{O_2}(\pi_2, \alpha_L).$$

For $\gamma \geq 1/2$, it is clear that $\Gamma_\pi(\pi_2, \tau_2) < 0$, so that $\tau_2 = \tau_1$. Moreover, since $\tau_1 < \hat{\tau}_L$, then as $\gamma \rightarrow 1/2$,

$$\begin{aligned} \Gamma_\pi(\pi_2, \tau_2) &= (1 - \gamma)\hat{y}_\pi^{I_2}(\pi_2, \alpha_L) + [\gamma - \tau_2(1 - 2\gamma)]\hat{y}_\pi^{O_2}(\pi_2, \alpha_L) \\ &= \frac{1}{2}(\hat{y}_\pi^{I_2}(\pi_2, \alpha_L) + \hat{y}_\pi^{O_2}(\pi_2, \alpha_L)) \\ &= (1 - \sigma)\eta[k_L]^{\eta-1}\frac{\sigma w_p(\pi_2)}{2(1 - \sigma)} \\ &= \eta\left[\frac{\sigma}{2(1 - \sigma)}\right]^\eta [w(\pi_2)]^{\eta-1}w_p(\pi_2) \\ &< \eta 2^\eta\left[\frac{\sigma}{2(1 - \sigma)}\right]^\eta [w(\pi_2)]^{\eta-1}w_p(\pi_2) \\ &= (1 - \sigma)\eta[k_H]^{\eta-1}\frac{w_p(\pi_2)}{(1 - \sigma)}, \end{aligned}$$

where the last expression is equal to $\Gamma_\pi(\pi_2, \tau_2)$ when $\tau_2 > \hat{\tau}_L$. This, along with the fact that the state capacity investments are complements, proves the result. *Q.E.D.*

REFERENCES

- ACEMOGLU, D. (2003): "Why Not a Political Coase Theorem: Social Conflict, Commitment, and Politics," *Journal of Comparative Economics*, 31, 620–652. [27]
- (2005): "Politics and Economics in Weak and Strong States," *Journal of Monetary Economics*, 52, 1199–1226. [4]
- (2006): "Modeling Inefficient Institutions," in *Advances in Economic Theory and Econometrics: Proceedings of the Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge, U.K.: Cambridge University Press. [22,25,27]
- ACEMOGLU, D., D. TICCHI, AND A. VINDIGNI (2009): "Emergence and Persistence of Inefficient States," *Journal of the European Economic Association* (forthcoming). [4]
- BARRO, R. J. (1990): "Government Spending in a Simple Model of Endogenous Growth," *Journal of Political Economy*, 98, 103–125. [22]
- BARRO, R. J., AND X. SALA-I-MARTIN (1992): "Public Finance in Models of Economic Growth," *Review of Economic Studies*, 59, 645–661. [23]
- BENABOU, R. (1997): "Inequality and Growth," in *NBER Macroeconomics Annual 1996*. Cambridge, MA: MIT Press. [12]
- BESLEY, T., AND T. PERSSON (2008a): "Wars and State Capacity," *Journal of the European Economic Association*, 6, 522–530. [13]
- (2008b): "The Incidence of Civil War: Theory and Evidence," Working Paper 14585, NBER. [14,17,18]

- _____. (2009a): "The Origins of State Capacity: Property Rights, Taxation and Politics," *American Economic Review*, 99, 1218–1244. [3-5,7-9,11,12,25]
- _____. (2009b): "Repression or Civil War?" *American Economic Review*, 99, 292–297. [18]
- _____. (2009c): "State Capacity, Conflict and Development," Working Paper 15088, NBER. [25,26]
- BLATTMAN, C., AND E. MIGUEL (2009): "Civil War," *Journal of Economic Literature* (forthcoming). [13]
- BREWER, J. (1989): *The Sinews of Power: War, Money and the English State, 1688–1783*. New York: Knopf. [3]
- COLLIER, P., AND A. HOEFFLER (2004): "Greed and Grievance in Civil War," *Oxford Economic Papers*, 56, 563–595. [13]
- CUKIERMAN, A., S. EDWARDS, AND G. TABELLINI (1992): "Seigniorage and Political Instability," *American Economic Review*, 82, 537–555. [4]
- DIAMOND, P., AND J. MIRRLEES (1971): "Optimal Taxation and Public Production: I Production Efficiency," *American Economic Review*, 61, 8–27. [8,24]
- FEARON, J., AND D. LAITIN (2003): "Ethnicity, Insurgency and Civil War," *American Political Science Review*, 97, 75–90. [13]
- HALL, R., AND C. JONES (1999): "Why Do Some Countries Produce so Much More Output per Worker Than Others?" *Quarterly Journal of Economics*, 114, 83–116. [12]
- KING, R. G., AND R. LEVINE (1993): "Finance and Growth: Schumpeter Might Be Right," *Quarterly Journal of Economics*, 108, 717–737. [12]
- LACINA, B. A., AND N. P. GLEDITSCH (2005): "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths," *European Journal of Population*, 21, 145–165. [13]
- LA PORTA, R., F. LOPEZ DE SILANES, A. SHLEIFER AND R. VISHNY (1998): "Law and Finance," *Journal of Political Economy*, 106, 1113–1155. [11]
- LEVI, M. (1988): *Of Rule and Revenue*. University of California Press. [3]
- MATHIAS, P., AND P. O'BRIEN (1976): "Taxation in Britain and France 1715–1810: A Comparison of the Social and Economic Consequences of Taxes Collected for the Central Governments," *Journal of European Economic History*, 5, 601–650. [10]
- MIGUEL, E., S. SATYANATH, AND E. SERGENTI (2004): "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 112, 725–753. [13]
- MIGDAL, J. S. (1988): *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton, NJ: Princeton University Press. [3]
- MILGROM, P., AND C. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180. [10]
- O'BRIEN, P. (2005): "Fiscal and Financial Preconditions for the Rise of British Naval Hegemony, 1485–1815," Working Paper 91/05, London School of Economics and Political Science. [11]
- PAGANO, M., AND P. VOLPIN (2005): "The Political Economy of Corporate Governance," *American Economic Review*, 95, 1005–1030. [27]
- RAJAN, R., AND L. ZINGALES (2003): "The Great Reversal: The Politics of Financial Development in the Twentieth Century," *Journal of Financial Economics*, 69, 5–50. [27]
- RICE, S., AND S. PATRICK (2008): *Index of State Weakness in the Developing World*. Washington, DC: The Brookings Institution. [3]
- ROSS, M. (2004): "What Do We Know About Natural Resources and Civil War?" *Journal of Peace Research*, 41, 337–356. [13]
- SAMBANIS, N. (2002): "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War," *Defense and Peace Economics*, 13, 215–243. [13]
- STASAVAGE, D. (2007): "Partisan Politics and Public Debt: The Importance of the 'Whig Supremacy' for Britain's Financial Revolution," *European Review of Economic History*, 11, 123–153. [11]
- SVENSSON, J. (1998): "Investment, Property Rights and Political Instability: Theory and Evidence," *European Economic Review*, 42, 1317–1341. [4]

TILLY, C. (1985): "Warmaking and State Making as Organized Crime," in *Bringing the State Back In*, ed. by P. Evans, D. Rueschemeyer, and T. Skocpol. Cambridge, U.K.: Cambridge University Press. [3,10]

London School of Economics, Houghton Street, London WC2A 2AE, U.K.;
t.besley@lse.ac.uk; http://econ.lse.ac.uk/staff/tbesley/index_own.html
and

*Institute for International Economic Studies, Stockholm University, SE-106 91
Stockholm, Sweden; torsten.perrson@iies.su.se; http://www.iies.su.se/~persson/.*

Manuscript received August, 2008; final revision received October, 2009.

WHAT DRIVES MEDIA SLANT? EVIDENCE FROM U.S. DAILY NEWSPAPERS

BY MATTHEW GENTZKOW AND JESSE M. SHAPIRO¹

We construct a new index of media slant that measures the similarity of a news outlet's language to that of a congressional Republican or Democrat. We estimate a model of newspaper demand that incorporates slant explicitly, estimate the slant that would be chosen if newspapers independently maximized their own profits, and compare these profit-maximizing points with firms' actual choices. We find that readers have an economically significant preference for like-minded news. Firms respond strongly to consumer preferences, which account for roughly 20 percent of the variation in measured slant in our sample. By contrast, the identity of a newspaper's owner explains far less of the variation in slant.

KEYWORDS: Bias, text categorization, media ownership.

1. INTRODUCTION

GOVERNMENT REGULATION OF NEWS MEDIA ownership in the United States is built on two propositions. The first is that news content has a powerful impact on politics, with ideologically diverse content producing socially desirable outcomes. According to the U.S. Supreme Court (1945), “One of the most vital of all general interests [is] the dissemination of news from as many different sources, and with as many different facets and colors as is possible. That interest... presupposes that right conclusions are more likely to be gathered out of a multitude of tongues, than through any kind of authoritative selection.”

The second proposition is that unregulated markets will tend to produce too little ideological diversity. The highly influential Hutchins Commission report identified cross-market consolidation in newspaper ownership as a major obstacle to the emergence of truth in the press (Commission on Freedom of

¹We are grateful to Attila Ambrus, David Autor, Gary Becker, Gary Chamberlain, Raj Chetty, Tim Conley, Liran Einav, Edward Glaeser, Tim Groseclose, Christian Hansen, Justine Hastings, Chris Hayes, Daniel Hojman, Matt Kahn, Larry Katz, John List, Kevin M. Murphy, Ben Olken, Ariel Pakes, Andrea Prat, Riccardo Puglisi, Sam Schulhofer-Wohl, Andrei Shleifer, Monica Singhal, Jim Snyder, Wing Suen, Catherine Thomas, Abe Wickelgren, and numerous seminar and conference participants for helpful comments. We especially wish to thank Renata Voccia, Paul Wilt, Todd Fegan, and the rest of the staff at ProQuest for their support and assistance at all stages of this project. Mike Abito, Steve Cicala, Hays Golden, James Mahon, Jennifer Paniza, and Mike Sinkinson provided outstanding research assistance and showed tireless dedication to this project. We also thank Yujing Chen, Alex Fogel, Lisa Furchtgott, Ingrid Gonçalves, Hayden Haralson Hudson, and Hannah Melnicoe for excellent research assistance. This research was supported by National Science Foundation Grant SES-0617658, as well as the Stigler Center for the Study of the State and the Economy, the Initiative on Global Markets, and the Centel Foundation/Robert P. Reuss Faculty Research Fund, all at the University of Chicago Booth School of Business.

the Press (1947)). The Federal Communications Commission (FCC) “has traditionally assumed that there is a positive correlation between viewpoints expressed and ownership of an outlet. The Commission has sought, therefore, to diffuse ownership of media outlets among multiple firms in order to diversify the viewpoints available to the public” (FCC (2003)). This belief has justified significant controls on cross-market consolidation in broadcast media ownership, on foreign ownership of media, and on cross-media ownership within markets, and has motivated a sizable academic literature arguing that current media ownership is too concentrated (Bagdikian (2000)).

That news content can have significant effects on political attitudes and outcomes has been documented empirically by Strömberg (2004), Gentzkow and Shapiro (2004), Gentzkow (2006), Gerber, Karlan, and Bergan (2009), DellaVigna and Kaplan (2007), and others. In contrast, evidence on the incentives that shape ideological content and on the role of ownership, in particular, is limited. Existing studies have generally relied on hand collection and coding of news content, and so have been restricted to small numbers of sources (e.g., Glasser, Allen, and Blanks (1989), Pritchard (2002)). Groseclose and Milyo (2005) made an important contribution, proposing a new measure of ideological content based on counts of think-tank citations. However, their index was calculated only for a small number of outlets, and has not been used to analyze the determinants of slant.

In this paper, we propose a new index of ideological slant in news coverage and compute it for a large sample of U.S. daily newspapers. We estimate a model of newspaper demand that incorporates slant explicitly, estimate the slant that would be chosen if newspapers independently maximized their own profits, and compare these profit-maximizing points with firms’ actual choices. We estimate the contributions of consumer and owner heterogeneity to cross-market diversity in slant and develop tentative implications for ownership regulation.

Our slant index measures the frequency with which newspapers use language that would tend to sway readers to the right or to the left on political issues. We focus on newspapers’ news (rather than opinion) content, because of its centrality to public policy debates and its importance as a source of information to consumers.² To measure news slant, we examine the set of all phrases used by members of Congress in the 2005 *Congressional Record*, and identify those that are used much more frequently by one party than by another. We then index newspapers by the extent to which the use of politically charged phrases in their news coverage resembles the use of the same phrases in the speech of a congressional Democrat or Republican. The resulting index allows us to

²Nearly two-thirds of Americans report getting news several times a week or daily from local newspapers (Harris Interactive (2006)). Independent evidence suggests that almost 90 percent of readers of daily newspapers read the main news section, with over 80 percent reading the local news section (Newspaper Association of America (2006)).

compare newspapers to one another, though not to a benchmark of “true” or “unbiased” reporting.

Two key pieces of evidence suggest that our methodology produces a meaningful measure of slant. First, many of the phrases that our automated procedure identifies are known from other sources to be chosen strategically by politicians for their persuasive impact. Examples include “death tax,” “tax relief,” “personal account,” and “war on terror” (which we identify as strongly Republican), and “estate tax,” “tax break,” “private account,” and “war in Iraq,” (which we identify as strongly Democratic). Second, the index that we construct using counts of these phrases in news coverage is consistent with readers’ subjective evaluation of newspapers’ political leanings (data on which are available for several large papers in our sample).

We use our measure to estimate a model of newspaper demand, in which a consumer’s utility from reading a newspaper depends on the match between the newspaper’s slant and the consumer’s own ideology (Mullainathan and Shleifer (2005), Gentzkow and Shapiro (2006)). Using zip code-level data on newspaper circulation, we show that right-wing newspapers circulate relatively more in zip codes with a higher proportion of Republicans, even within a narrowly defined geographic market. Left-wing newspapers show the opposite pattern. Because we only use within-market variation to identify our model, our estimates are consistent even though slant is endogenous to the average political tastes in a market. We show that our results are also robust to correcting for measurement error (and for a subtler form of endogeneity bias) using an identification strategy in the spirit of George and Waldfogel (2003).

Treating newspapers as local monopolists, we compute the slant that each newspaper would choose if it independently maximized its own profits. The average profit-maximizing slant is close to the newspapers’ actual slant. This finding is relevant to theories in which supply-side forces cause distortions in slant at the aggregate level. For example, if either the party identity of national incumbent politicians (Besley and Prat (2006)) or the distribution of political views among journalists in the country as a whole (Baron (2006)) were important drivers of slant, we would have expected to see deviation from profit maximization on average.

We also estimate a model of the supply of slant, in which we allow slant to respond both to the ideology of a newspaper’s customers and also to the identity of its owner.

Variation in slant across newspapers is strongly related to the political makeup of their potential readers and thus to our estimated profit-maximizing points. The relationship between slant and consumer ideology remains when we compare different newspapers with the same owner or different newspapers in the same state. Overall, variation in consumer political attitudes explains roughly 20 percent of the variation in measured slant in our sample.

An obvious concern in interpreting the relationship between slant and consumer attitudes is that it may reflect causation running from slant to consumer

beliefs rather than the reverse. To address this, we show that the relationship survives when we instrument for consumer political attitudes using religiosity—a strong predictor of political preferences that is unlikely to be affected by newspaper content. These results do not mean that newspapers do not affect beliefs; indeed, our study is motivated in part by evidence that they do. Rather, our findings suggest that the effect of slant on ideology accounts for only a small part of the cross-sectional variation in ideology that identifies our model.

We find little evidence that the identity of a newspaper's owner affects its slant. After controlling for geographic clustering of newspaper ownership groups, the slant of co-owned papers is only weakly (and statistically insignificantly) related to a newspaper's political alignment. Direct proxies for owner ideology, such as patterns of corporate or executive donations to political parties, are also unrelated to slant. Estimates from a random effects model confirm a statistically insignificant role for owners, corresponding to approximately 4 percent of the variance in measured slant.

In the final section of the paper, we present additional evidence on the role of pressure from incumbent politicians (Besley and Prat (2006)), and the tastes of reporters and editors (Baron (2006)). The evidence we present suggests that neither of these forces is likely to explain a large share of the variation in slant.

This paper presents the first large-scale empirical evidence on the determinants of political slant in the news,³ and informs the theoretical literature on demand-side (Mullainathan and Shleifer (2005), Gentzkow and Shapiro (2006), Suen (2004)) and supply-side (Besley and Prat (2006), Balan, De-Graba, and Wickelgren (2009), Baron (2006)) drivers of slant. Our findings contribute to the literature on product positioning in the mass media (Sweeting (2007, 2008), Myers (2008), George (2007)), as well as to research on product differentiation more generally (Mazzeo (2002a, 2002b), Dranove, Gron, and Mazzeo (2003), Seim (2006), Dubé, Hitsch, and Manchanda (2005), Einav (2007)).

Our work also advances the measurement of media slant (Groseclose and Milyo (2005), Puglisi (2008), Larcinese, Puglisi and Snyder (2007), Gentzkow, Glaeser, and Goldin (2006)).⁴ Groseclose and Milyo (2005) use Congressional citations to estimate the political positions of think tanks, and then use data on media mentions of the same set of think tanks to measure the bias of 20 news outlets. Our automated procedure allows us to measure the slant of a much wider range of outlets, including over 400 daily newspapers representing over

³Hamilton (2004) presented an important overview of many of the issues we explore. An existing literature explores the determinants of newspaper endorsements of political candidates, rather than news content (see, e.g., Akhavan-Majid, Rife, and Gopinath (1991) or Ansolabehere, Lessem, and Snyder (2006)).

⁴Our approach borrows tools from the computer science literature on text categorization (see Aas and Eikvil (1999) for a review), which social scientists have applied to the measurement of sentiment (e.g., Antweiler and Frank (2004)) and politicians' platforms (Laver, Benoit, and Garry (2003)), but not (to our knowledge) to the political slant of the news media.

70 percent of total daily circulation in the United States. Moreover, rather than imposing a list of likely partisan phrases (such as names of think tanks), we use data from Congress to isolate the phrases that have the most power to identify the speaker’s ideology.

The remainder of the paper is organized as follows. Section 2 discusses our data sources. Section 3 describes the computation of our measure of newspaper slant and validates the measure using alternative rankings of newspapers’ political content. Section 4 presents our model, and Section 5 discusses identification and estimation. Sections 6, 7, and 8 present our core results. Section 9 tests two prominent theories of the determinants of media slant. Section 10 concludes.

2. DATA

2.1. *Congressional Record and Congressperson Data*

Our approach to measuring slant requires data on the frequency with which individual members of Congress use particular phrases. We use the text of the 2005 *Congressional Record*, downloaded from [thomas.loc.gov](#) and parsed using an automated script that identifies the speaker of each passage. To increase the efficiency of our text analysis algorithm, we apply a standard preprocessing procedure that removes extremely common words (such as “to,” “from,” and “the”) and strips words down to shared linguistic roots (so that, for example, “tax cut” and “tax cuts” are identified as the same phrase). A final script produces counts by speaker and party of two- and three-word phrases in the *Congressional Record*. Appendix A contains additional details on this process.

For each congressperson (member of the House or Senate), we obtain data on party identification, as well as the share of the 2004 two-party presidential vote total going to George W. Bush in the congressperson’s constituency (congressional district for representatives; state for senators). This vote share (which comes from [polidata.org](#) in the case of congressional districts) serves as our primary measure of a congressperson’s ideology. We show in the online Appendix B (Gentzkow and Shapiro (2010)) that it is highly correlated with two commonly used roll-call measures of congressional ideology and that our results are robust to using these alternative measures of ideology as the basis for our analysis.

2.2. *Newspaper Text and Characteristics*

As an input to our slant measure, we obtain counts of the frequency with which phrases appear in news coverage from two sources: the NewsLibrary data base ([newslibrary.com](#)) and the ProQuest Newsstand data base ([proquest.com](#)). For each data base, we use an automated script to calculate the number of articles containing each phrase in each newspaper during calendar year 2005. Whenever possible, we exclude opinion content. Also, because

some newspapers do not archive reprinted wire stories with ProQuest, we exclude articles from the Associated Press, focusing instead on content originating with the newspaper. Appendix A provides additional details on the mechanics of these searches.

We compute slant for all English language daily newspapers available in either ProQuest or NewsLibrary for a total sample of 433 newspapers.⁵ These newspapers together represented 74 percent of the total circulation of daily newspapers in the United States in 2001.

To measure the ownership and market characteristics of the newspapers in our sample, we first match every newspaper to data from the 2001 Editor and Publisher (E&P) International Yearbook CD-ROM. The E&P data set identifies the owner of each newspaper as of 2000.

The E&P data set also identifies the zip code of each newspaper's headquarters, which we match to counties using the United States 5-Digit ZIP Code Database from Quentin Sager Consulting. We match counties to primary metropolitan statistical areas (PMSAs) using definitions from the 1990 census. We define each newspaper's geographic market as the PMSA in which it is headquartered. If a newspaper is not located inside a PMSA, we define its market to be the county in which it is located. For the median newspaper, this market definition includes more than 90 percent of the newspaper's total circulation (among newspapers for which we have zip code-level circulation data). For four newspapers—the *New York Times*, the *Wall Street Journal*, the *Christian Science Monitor*, and *USA Today*—the notion of a geographic market is ill defined. We exclude these papers from our analysis, leaving a sample of 429 newspapers with well defined geographic markets.

For each newspaper, we obtain a wide range of demographic characteristics of the paper's market from the 2000 U.S. Census. We also obtain data from David Leip's Atlas of U.S. Presidential Elections (uselectionatlas.org) on the share of votes in each market going to Bush in the 2004 presidential election; this is used as a proxy for the market's political leanings. Last, we use the DDB Needham Life Style Survey (Putnam (2000)), available on bowlingalone.com, to compute a measure of the share of survey respondents from 1972 to 1998 who reported attending church monthly or more. This measure serves as a plausibly exogenous shifter of the political leanings of the market in that it is unlikely to be directly affected by the slant of area newspapers.

As a potential proxy for a media firm's ideological leanings, we obtain data from the Center for Public Integrity (publicintegrity.org) on the share of each newspaper firm's corporate political contribution dollars going to Republicans. We also searched the Federal Election Commission (FEC) disclosure data base

⁵One additional newspaper—the *Chicago Defender*—is present in the news data bases, but is excluded from our analysis because it is an extreme outlier (more than 13 standard deviations away from the mean) in the distribution of slant. A large share of hits for this paper are for a single phrase, “African American,” which is strongly predictive of liberal ideology in Congress.

for information on the personal contributions of the Chief Executive Officer, President, Chairman, and Managing Director of each firm that owns two or more U.S. daily newspapers. For newspapers owned by a firm with no other daily newspaper holdings, we conducted an analogous search, but collected data on executives of the newspaper itself.

2.3. Newspaper Circulation and Consumer Characteristics

For our study of the effects of slant on newspaper demand, we use zip code-level data on newspaper circulation from the Audit Bureau of Circulation's (ABC) Newspaper GeoCirc data set. We include all zip code–newspaper pairs with positive circulation. We match each zip code to a news market using the market definition above.

To adjust for nonpolitical differences across zip codes, we make use of a set of zip code demographics taken from the 2000 U.S. Census (census.gov): log of total population, log of income per capita, percent of population urban, percent white, percent black, population per square mile, share of houses that are owner occupied, and the share of population 25 and over whose highest level of schooling is college.

Measuring each zip code's ideology is complicated by the fact that voting data are not available at the zip code level. To circumvent this problem, we use the Federal Election Commission's (FEC) 2000, 2002, and 2004 Individual Contributions Files. These files, which are available for download at fec.gov, contain a record of every individual contribution to a political party, candidate, or political action committee registered with the FEC. Each donor record includes a complete address, allowing us to identify donors' zip codes. For each zip code, we compute the share of donations (denominated in number of donations, not dollars) received by a Republican affiliate among donations received by either Republican- or Democrat-affiliated entities. To reduce the noise in the measure, we restrict attention to zip codes with 20 or more donors.

To test the validity of this proxy for ideology, we take advantage of data on the number of registered Democrats and Republicans by zip code in California as of March 2006.⁶ The donation measure has a correlation of 0.65 with the two-party share of Republican registrants.

Of course, the sample of donors to political causes is not fully representative of the entire population of a zip code. Donors tend to be older, richer, and more educated than nondonors (Gimpel, Lee, and Kaminski (2006)). However, these are also the demographic characteristics of likely readers of newspapers (Gentzkow (2007)) and, therefore, if anything, may tend to make our measure more representative of the population relevant for studying newspaper demand.

⁶We are grateful to Marc Meredith for providing these data.

Our analysis of newspaper demand is restricted to the 290 newspapers in our primary sample for which we observe at least one zip code with both positive circulation in the ABC data and sufficiently many donors in the FEC data.

3. MEASURING SLANT

Our approach to measuring the slant of a newspaper will be to compare phrase frequencies in the newspaper with phrase frequencies in the 2005 *Congressional Record* to identify whether the newspaper's language is more similar to that of a congressional Republican or a congressional Democrat.

For a concrete illustration of our approach to measuring slant, consider the use of the phrases "death tax" and "estate tax" to describe the federal tax on assets of the deceased. The phrase "death tax" was coined by the tax's conservative opponents. According to a high-level Republican staffer, "Republicans put a high level of importance on the death/estate tax language—they had to work hard to get members to act in unison, including training members to say 'death tax'... Estate tax sounds like it only hits the wealthy but 'death tax' sounds like it hits everyone" (Graetz and Shapiro (2005)). In Congress in 2005, Republicans used the phrase "death tax" 365 times and the phrase "estate tax" only 46 times. Democrats, by contrast, had the reverse pattern, using the phrase "death tax" only 35 times and the phrase "estate tax" 195 times.

The relative use of the two phrases in newspaper text conforms well to prior expectations about political slant. Compare, for example, the *Washington Post* and the *Washington Times*. The *Post* is widely perceived to be more liberal than the *Times*.⁷ In 2005, the *Post* used the phrase "estate tax" 13.7 times as often as it used the phrase "death tax," while the *Times* used "estate tax" 1.3 times as often. As we show below, this case is not unusual: there is a significant correlation between popular perceptions of a newspaper's political leanings and its propensity to use words and phrases favored by different political parties in Congress. Our measure of media slant exploits this fact by endogenously identifying politically charged phrases like "death tax" and "estate tax," and computing their frequencies in daily newspapers throughout the United States.

In principle, we could base our measure on counts of *all* phrases that appear in the *Congressional Record*. A simple procedure would be as follows. First, for each politician, we compute a vector that gives the number of times each phrase appeared in their speeches. Second, we compute a mapping from the vector of counts to a measure of a politician's ideology. Finally, we generate counts of each phrase in a newspaper's text and apply the same mapping to generate an index of the newspaper's ideology.

⁷The website mondotimes.com presents an index of newspapers' political leanings based on user ratings. The *Post* is rated as "leans left," while the *Times* is rated as "conservative." Groseclose and Milyo (2005) also rated the *Post* as significantly to the left of the *Times*.

Because the total number of phrases that appear in the *Congressional Record* is in the millions, this simple procedure is computationally infeasible. We therefore add a “feature selection” step in which we use simple computations to identify a set of phrases that are highly diagnostic of the speaker’s political party. We use this restricted phrase set for the more computationally burdensome step of mapping phrase counts to a continuous measure of ideology, counting occurrences in newspapers, and estimating newspaper ideology.

3.1. Selecting Phrases for Analysis

Let f_{pld} and f_{plr} denote the total number of times phrase p of length l (two or three words) is used by Democrats and Republicans, respectively. Let $f_{\sim pld}$ and $f_{\sim plr}$ denote the total occurrences of length- l phrases that are *not* phrase p spoken by Democrats and Republicans, respectively. Let χ^2_{pl} denote Pearson’s χ^2 statistic for each phrase:

$$(1) \quad \chi^2_{pl} = \frac{(f_{plr}f_{\sim pld} - f_{pld}f_{\sim plr})^2}{(f_{plr} + f_{pld})(f_{plr} + f_{\sim plr})(f_{pld} + f_{\sim pld})(f_{\sim plr} + f_{\sim pld})}.$$

We select the phrases for our analysis as follows:

- (i) We compute the total number of times that each phrase appeared in newspaper headlines and article text in the ProQuest Newsstand data base from 2000 to 2005. We restrict attention to two-word phrases that appeared in at least 200 but no more than 15,000 newspaper headlines, and three-word phrases that appeared in at least 5 but no more than 1000 headlines. We also drop any phrase that appeared in the full text of more than 400,000 documents.
- (ii) Among the remaining phrases, we select the 500 phrases of each length l with the greatest values of χ^2_{pl} , for a total of 1000 phrases.

The first step eliminates phrases that are not likely to be useful for diagnosing newspaper partisanship. For example, procedural phrases such as “yield the remainder of my time” are commonly employed in the *Congressional Record*—especially by the majority party—but are almost never used in newspapers. Extremely common phrases such as “third quarter” or “exchange rate” are also unlikely to be diagnostic of ideology, but impose a high burden on our procedure for extracting phrase counts in newspaper text. The cutoffs we impose are arbitrary. In (online) Appendix B, we show that our results are robust to tightening these cutoffs.

The second step identifies phrases that are diagnostic of the speaker’s political party. If the counts f_{pld} and f_{plr} are drawn from (possibly different) multinomial distributions, χ^2_{pl} is a test statistic for the null hypothesis that the propensity to use phrase p of length l is equal for Democrats and Republicans. This statistic conveniently summarizes the political asymmetry in the use of the phrase. (More naive statistics, such as the ratio of uses by Republicans to uses by Democrats, would tend to select phrases that are used only once by

TABLE I
MOST PARTISAN PHRASES FROM THE 2005 CONGRESSIONAL RECORD^a

Panel A: Phrases Used More Often by Democrats		
<i>Two-Word Phrases</i>		
private accounts	Rosa Parks	workers rights
trade agreement	President budget	poor people
American people	Republican party	Republican leader
tax breaks	change the rules	Arctic refuge
trade deficit	minimum wage	cut funding
oil companies	budget deficit	American workers
credit card	Republican senators	living in poverty
nuclear option	privatization plan	Senate Republicans
war in Iraq	wildlife refuge	fuel efficiency
middle class	card companies	national wildlife
<i>Three-Word Phrases</i>		
veterans health care	corporation for public	cut health care
congressional black caucus	broadcasting	civil rights movement
VA health care	additional tax cuts	cuts to child support
billion in tax cuts	pay for tax cuts	drilling in the Arctic National
credit card companies	tax cuts for people	victims of gun violence
security trust fund	oil and gas companies	solvency of social security
social security trust	prescription drug bill	Voting Rights Act
privatize social security	caliber sniper rifles	war in Iraq and Afghanistan
American free trade	increase in the minimum wage	civil rights protections
central American free	system of checks and balances	credit card debt
	middle class families	

(Continues)

Republicans and never by Democrats, even though pure sampling error could easily generate such a pattern.) χ^2_{pl} is also simple to compute, in the sense that it requires only two calculations per phrase: the number of uses by Republicans and the number of uses by Democrats.

Table I shows the top phrases (arranged in order of descending χ^2_{pl} by length) in our final set of 1000. Panel A shows phrases used more often by congressional Democrats. Panel B shows phrases used more often by congressional Republicans.

Our procedure identifies many phrases that both intuition and existing evidence suggest are chosen strategically for their partisan impact. For example, a widely circulated 2005 memo by Republican consultant Frank Luntz advised candidates on the language they should use to describe President Bush's proposed Social Security reform (Luntz (2005)):

Never say 'privatization/private accounts.' Instead say 'personalization/personal accounts.' Two-thirds of America want to personalize Social Security while only one-third would privatize it. Why? Personalizing Social Security suggests ownership and control over your retirement savings, while privatizing it suggests a profit motive and winners and losers.

TABLE I—Continued

Panel B: Phrases Used More Often by Republicans		
<i>Two-Word Phrases</i>		
stem cell	personal accounts	retirement accounts
natural gas	Saddam Hussein	government spending
death tax	pass the bill	national forest
illegal aliens	private property	minority leader
class action	border security	urge support
war on terror	President announces	cell lines
embryonic stem	human life	cord blood
tax relief	Chief Justice	action lawsuits
illegal immigration	human embryos	economic growth
date the time	increase taxes	food program
<i>Three-Word Phrases</i>		
embryonic stem cell	Circuit Court of Appeals	Tongass national forest
hate crimes legislation	death tax repeal	pluripotent stem cells
adult stem cells	housing and urban affairs	Supreme Court of Texas
oil for food program	million jobs created	Justice Priscilla Owen
personal retirement accounts	national flood insurance	Justice Janice Rogers
energy and natural resources	oil for food scandal	American Bar Association
global war on terror	private property rights	growth and job creation
hate crimes law	temporary worker program	natural gas natural
change hearts and minds	class action reform	Grand Ole Opry
global war on terrorism	Chief Justice Rehnquist	reform social security

^aThe top 60 Democratic and Republican phrases, respectively, are shown ranked by χ^2_{pl} . The phrases are classified as two or three word after dropping common “stopwords” such as “for” and “the.” See Section 3 for details and see Appendix B (online) for a more extensive phrase list.

We identify “personal accounts,” “personal retirement accounts,” and “personal savings accounts” as among the most Republican phrases in the *Congressional Record*, while “private accounts,” “privatization plan,” and other variants show up among the most Democratic phrases. Similarly, we identify “death tax” (whose partisan pedigree we discussed above) as the third most Republican phrase. We identify “tax relief”—a term also advocated by Luntz (2005)—as strongly Republican, while “tax breaks” is strongly Democratic. On foreign policy, we identify variants on the phrase “global war on terror” as among the most strongly Republican phrases, while “war in Iraq” and “Iraq war” are Democratic, again consistent with accounts of party strategy (e.g., Stevenson (2005)).

The phrases in our sample arise regularly in news content. The average newspaper in our sample used these phrases over 13,000 times in 2005. Even newspapers in the bottom quartile of daily circulation (in our newspaper sample) use these phrases over 4000 times on average. The contexts in which these phrases appear include local analogues of national issues, local impact of federal legislation, and the actions of legislators from local districts. In Ap-

pendix A, we present more systematic evidence on the contexts in which our phrases appear. Most occurrences are in independently produced news stories.

3.2. Mapping Phrases to Ideology

Re-index the phrases in our sample by $p \in \{1, \dots, 1000\}$. (Ignore phrase length for notational convenience.) For each congressperson $c \in C$, we observe ideology y_c and phrase frequencies $\{f_{pc}\}_{p=1}^{1000}$. Let $\tilde{f}_{pc} \equiv f_{pc} / \sum_{p=1}^P f_{pc}$ denote the relative frequency of phrase p in the speech of congressperson c .

We have a set of newspapers $n \in N$ for which we observe phrase frequencies $\{f_{pn}\}_{p=1}^{1000}$ but not ideology y_n . We estimate ideology for newspapers as follows:

- (i) For each phrase p , we regress \tilde{f}_{pc} on y_c for the sample of congresspeople, obtaining intercept and slope parameters a_p and b_p , respectively.
- (ii) For each newspaper n , we regress $(\tilde{f}_{pn} - a_p)$ on b_p for the sample of phrases, obtaining slope estimate

$$(2) \quad \hat{y}_n = \frac{\sum_{p=1}^{1000} b_p (\tilde{f}_{pn} - a_p)}{\sum_{p=1}^{1000} b_p^2}.$$

(We also compute an analogous estimate \hat{y}_c for each congressperson c .)

This approach can be understood as follows. First, we use congresspeople—whose ideology is observed—to estimate the relationship between the use of a phrase p and the ideology of the speaker. Second, we use the relationship observed in the first stage to infer the ideology of newspapers by asking whether a given newspaper tends to use phrases favored by more Republican members of Congress. If the use of some phrase p is uncorrelated with a congressperson's ideology ($b_p = 0$), the use of that phrase does not contribute to the estimate \hat{y}_n . If phrase p is used more often by more right-wing congresspeople ($b_p > 0$), the estimator will judge a speaker who uses phrase p often as more right wing. If newspaper phrase frequencies are given by $\tilde{f}_{pn} = a_p + b_p y_n + e_{pn}$, with $E(e_{pn} | b_p) \equiv 0 \forall n$, then $E(\hat{y}_n) = y_n \forall n$.

The estimates \hat{y}_c have a correlation of 0.61 with true ideology y_c among our sample of congresspeople. This correlation provides in-sample evidence for the validity of our estimates, but also implies that our estimates are likely to contain a significant amount of noise. Taking the square of the correlation coefficient, 37 percent of the variation in slant is attributable to variation in ideology, with the rest coming from noise. Therefore, a useful benchmark is that, assuming the same share of noise among congresspeople and newspapers, 63 percent of the variation in slant among newspapers is likely to be noise.

Validating our approach among newspapers is more difficult. The estimate \hat{y}_n attempts to answer the question, “If a given newspaper were a congressperson,

how Republican would that congressperson's district be?" By definition, the true answer to this question is unobservable for newspapers, but a crude proxy is available. The media directory website Mondo Times (mondotimes.com) collects ratings of newspapers' political orientation from its users.⁸ Note that we would not necessarily expect these correlations to be perfect, both because most papers receive only a few ratings and because Mondo Times users are rating the opinion as well as news content of the papers, whereas our slant measure focuses on news content. Nevertheless, in Figure 1 we show that these

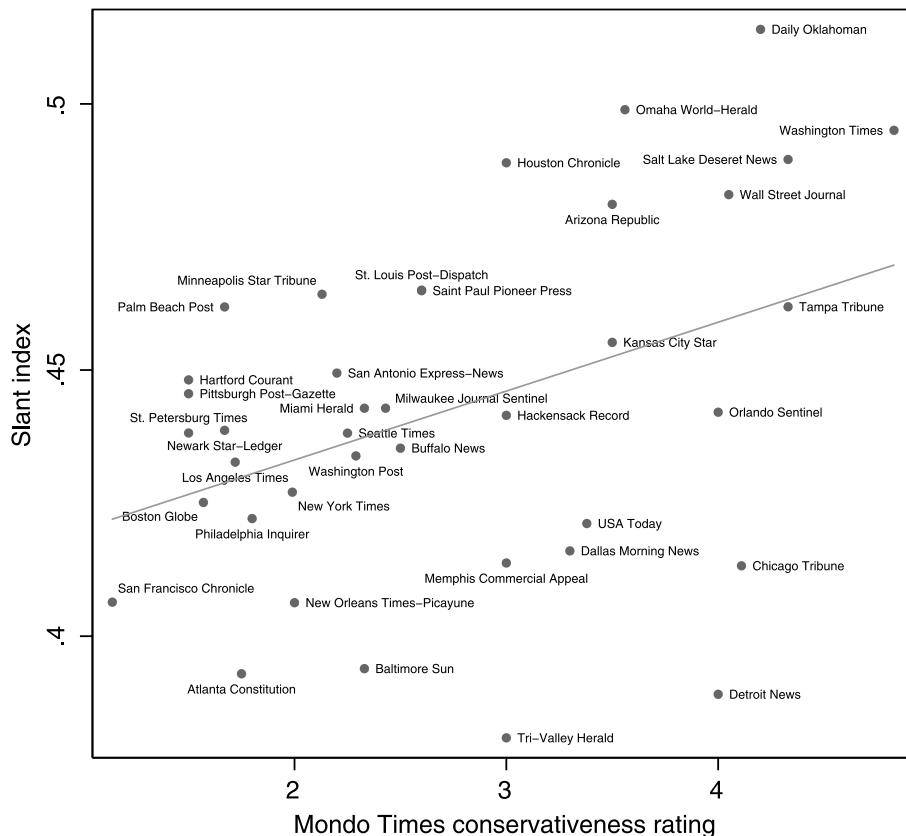


FIGURE 1.—Language-based and reader-submitted ratings of slant. The slant index (y axis) is shown against the average Mondo Times user rating of newspaper conservativeness (x axis), which ranges from 1 (liberal) to 5 (conservative). Included are all papers rated by at least two users on Mondo Times, with at least 25,000 mentions of our 1000 phrases in 2005. The line is predicted slant from an OLS regression of slant on Mondo Times rating. The correlation coefficient is 0.40 ($p = 0.0114$).

⁸We wish to thank Eric Kallgren of Mondo Code for graciously providing these data.

ratings are positively related to our slant index with a correlation coefficient of 0.40.

4. AN ECONOMIC MODEL OF SLANT

In this section we define the demand for and supply of slant. Our model is designed to capture three important features of newspaper markets. First, consumers may prefer newspapers whose slant is close to their own ideology. Second, firms will have an incentive to cater to this demand. Finally, owner ideology may also affect firms' choices of slant and this may lead slant to differ from the profit-maximizing level.

4.1. Consumer Problem

Each zip code z contains a continuum of households of mass H_z , with individual households indexed by i . A set of newspapers N_z is available in each zip code, and each household i must choose a subset $N_{iz} \subseteq N_z$ of the available newspapers to read. Household i in zip code z gets value u_{izn} from reading newspaper n , and the utility U_{iz} of household i is

$$U_{iz} \equiv \sum_{n \in N_{iz}} u_{izn}.$$

Consistent with utility maximization, household i in zip code z reads newspaper $n \in N_z$ iff $u_{izn} \geq 0$.

Each zip code z has an exogenous ideology r_z (with higher values meaning more conservative) and a preferred slant

$$\text{ideal}_z \equiv \alpha + \beta r_z.$$

If $\beta > 0$, more conservative zip codes prefer more conservative news, as in Mullainathan and Shleifer (2005).

Household utility u_{izn} is the sum of three components:

$$(3) \quad u_{izn} \equiv \bar{u}_{zn} - \gamma(y_n - \text{ideal}_z)^2 + \varepsilon_{izn}.$$

The term \bar{u}_{zn} is the exogenous taste of consumers in zip code z for newspaper n , possibly related to observables, but not affected by slant y_n . The term $-\gamma(y_n - \text{ideal}_z)^2$ captures the distaste for reading a newspaper whose slant y_n deviates from the preferred slant ideal_z . The error term ε_{izn} is a household-specific taste shock which we assume has a logistic distribution. We assume that \bar{u}_{zn} is known to firms (but not necessarily to the econometrician).

The share of households in zip code z reading newspaper n is then

$$(4) \quad S_{zn} = \frac{\exp[\bar{u}_{zn} - \gamma(y_n - \text{ideal}_z)^2]}{1 + \exp[\bar{u}_{zn} - \gamma(y_n - \text{ideal}_z)^2]}$$

if $n \in N_z$ and 0 otherwise.

If $\gamma, \beta > 0$, it is straightforward to show that equation (4) implies two key testable hypotheses:

HYPOTHESIS D1: Fixing \bar{u}_{zn}

$$\frac{\partial^2}{\partial y_n \partial r_z} \left(\ln \frac{S_{zn}}{1 - S_{zn}} \right) > 0.$$

More conservative zip codes have a relatively greater taste for more conservatively slanted news.

HYPOTHESIS D2: Fixing \bar{u}_{zn} and y_n

$$\frac{\partial^2}{\partial r_z^2} \left(\ln \frac{S_{zn}}{1 - S_{zn}} \right) < 0.$$

Demand has an inverted-U relationship to zip code ideology, peaking at $r_z = (y_n - \alpha)/\beta$.

4.2. Firm Problem

Assume that circulation revenue, advertising revenue, and variable costs are proportional to circulation, so that each newspaper earns a fixed markup for each copy sold. Let ideal_n be the value of y_n that maximizes newspaper n 's circulation. If all newspapers were operated by profit-maximizing firms, equilibrium slant would be $y_n^* = \text{ideal}_n$.

We allow for deviations from profit maximization. Each newspaper n is owned by a firm g , which has an ideology μ_g . Equilibrium slant is given by

$$(5) \quad y_n^* = \rho_0 + \rho_1 \text{ideal}_n + \mu_g.$$

When $\rho_0 = 0$, $\rho_1 = 1$, and $\mu_g = 0$, equation (5) is equivalent to profit maximization. Equation (5) can therefore be thought of as an approximation to a model in which a newspaper owner maximizes a utility function that includes dollar profits as well as nonpecuniary ideological motivations. In Gentzkow and Shapiro (2007), we derived an expression analogous to equation (5) from a set of primitive assumptions on consumers' and firms' utility functions.

We highlight two testable hypotheses of the model:

HYPOTHESIS S1: $\partial y_n / \partial \text{ideal}_n > 0$. Slant is increasing in consumer Republicanism.

HYPOTHESIS S2: $\partial y_n / \partial \mu_g > 0$. Slant is increasing in owner Republicanism.

4.3. *Discussion*

Our model is restrictive in a number of respects.

First, we do not explicitly model the fact that consumer ideology r_z may itself be a function of slant. Evidence suggests that slant does affect political behavior; this is an important motivation for our study. However, we expect that *most* of the variation in consumer ideology is related to consumer characteristics such as geography, race, and religiosity that are not affected by newspapers, making the potential bias in our estimates from ignoring reverse causality relatively small. In Section 7.1, we support this interpretation directly using an instrumental variables strategy in a cross-market regression of slant on consumer ideology. It is worth stressing, however, that we do not have an analogous instrument for the within-market (cross zip code) variation in ideology that identifies our demand model. Our demand estimates therefore rely more heavily than our supply estimates on the assumption that most variation in ideology is exogenous with respect to newspaper content.

Second, we assume that ideology does not vary across consumers within a zip code. This assumption approximates a model in which the average Republican in a heavily Republican zip code is further to the right than the average Republican in a more liberal zip code. In Gentzkow and Shapiro (2007), we showed that our main findings survive in a model that allows explicitly for within-zip code heterogeneity in political ideology.

Third, we assume that consumer utility is additive over newspapers, thus eliminating complementarity or substitutability in demand, and ruling out strategic interactions among newspapers. Since only a handful of papers in our sample face same-city competitors, we view a model without strategic interactions as a reasonable approximation. Excluding newspapers with same-city competitors does not change our results regarding the supply of slant (see online Appendix B). Our model does, however, ignore some potentially important strategic interactions, such as between newspapers and local television stations or newspapers in neighboring cities.

Fourth, we normalize the outside option to zero for all consumers. The outside option captures the value of all alternatives not written into the model, including television news, Internet news, and so forth. Because we will include market–newspaper fixed effects (FE) in our demand estimation, we in fact allow the utility of the outside option to vary nonparametrically by market. We do not, however, allow its utility to vary across zip codes; in particular, we rule out variation that is correlated with r_z . That assumption is important for our tests of Hypothesis D2 and for our structural estimates. It is not important for our tests of Hypothesis D1 and, indeed, we find evidence for Hypothesis D1 in a zip code fixed effects specification that allows arbitrary variation in the outside option across zip codes.

Finally, we assume that the markup newspapers earn is the same for each unit of circulation, whereas in reality advertisers prize some readers more than others. We show in the online Appendix B that allowing advertising revenues

per reader to vary across zip codes as a function of demographic characteristics does not change our conclusions.

5. IDENTIFICATION AND ESTIMATION

5.1. Demand Parameters

To estimate the demand model of equation (3), we specify the zip code–newspaper taste parameter \bar{u}_{zn} as

$$(6) \quad \bar{u}_{zn} = X_z \phi^0 + W_{zn} \phi^1 + \xi_{mn} + \nu_{zn},$$

where ϕ^0 and ϕ^1 are parameter vectors, X_z is a vector of zip code demographics, W_{zn} is a vector of interactions between the zip code demographics in X_z and the average level of the corresponding demographics in the newspaper’s market, ξ_{mn} is an unobservable product characteristic that is allowed to vary at the market level, and ν_{zn} is a zip code–newspaper-level unobservable.

Substituting for \bar{u}_{zn} and ideal_z in equation (4), and combining terms that do not vary within market–newspaper pairs, we have our estimating equation

$$(7) \quad \ln \frac{S_{zn}}{1 - S_{zn}} = \delta_{mn} + \lambda_0^d y_n r_z + \lambda_1^d r_z + \lambda_2^d r_z^2 + X_z \phi^0 + W_{zn} \phi^1 + \nu_{zn},$$

where $\lambda_0^d = 2\gamma\beta$, $\lambda_1^d = -2\gamma\alpha\beta$, and $\lambda_2^d = -\gamma\beta^2$, and where we treat the market–newspaper term

$$(8) \quad \delta_{mn} = -\gamma\alpha^2 - \gamma y_n^2 + 2\gamma\alpha y_n + \xi_{mn},$$

as a fixed effect.

We adopt an instrumental variables strategy to allow for measurement error in \hat{y}_n . We let R_n be the overall share of Republicans in newspaper n ’s primary market, measured using the Republican share of the 2004 two-party vote for president. We make the following assumptions:

- (i) $E[(\hat{y}_n - y_n) | R_n, r_z, X_z, W_{zn}, \delta_{mn}] = 0$.
- (ii) $E[\nu_{zn} | R_n, r_z, X_z, W_{zn}, \delta_{mn}] = 0$.

Under these assumptions, we consistently estimate the parameters of equation (7) via two-stage least squares, treating $r_z \hat{y}_n$ as an endogenous regressor, $r_z R_n$ as an excluded instrument, and δ_{mn} as a fixed effect. We allow for correlation in the error term ν_{zn} across observations for a given newspaper n .

Our instrumental variables strategy builds on George and Waldfogel’s (2003) insight that because fixed costs lead newspapers to cater to the average tastes of their readers, individuals will tend to read more when their tastes are similar to the average. By the same logic, our model predicts that if slant is an important component of demand, (i) newspapers with high R_n should choose high values of y_n and (ii) newspapers with high R_n should consequently be read relatively

more in zip codes with high r_z . The strength of these relationships will identify the coefficient on $y_n r_z$. Note that assuming that R_n is correlated with y_n is not equivalent to assuming that $y_n = y_n^*$ or that $y_n = \text{ideal}_n$. That is, for the purposes of our demand analysis, we do not assume that slant is chosen to maximize profits, only that it is correlated with consumer ideology in the newspaper's home market.

This strategy requires that the noise in our search-based measure of slant is unrelated to the characteristics of a newspaper's market. It also requires that we have controlled for zip code-specific factors that affect demand and are correlated with r_z or the interaction $r_z R_n$. Note that we do not need to assume that the market-newspaper taste shock ξ_{mn} is orthogonal to R_n : we allow for ξ_{mn} to be endogenous to R_n by treating δ_{mn} as a fixed effect.

Although our main reason for instrumenting is to correct for measurement error in \hat{y}_n , our instrument also addresses a subtle form of endogeneity bias. Note that the most obvious kind of endogeneity—that slant y_n may be a function of the unobserved product characteristic ξ_{mn} —would not affect even ordinary least squares (OLS) estimates because both the main effect of y_n and the unobservable ξ_{mn} are absorbed in δ_{mn} . However, slant could be endogenous, not to overall demand for the newspaper, but to the correlation between zip code ideology r_z and demand. More precisely, if the error term were written as $\tilde{\xi}_{mn} r_z + \nu_{zn}$, where $\tilde{\xi}_{mn}$ is a random coefficient, then slant y_n might tend to be higher for newspapers receiving a higher draw of $\tilde{\xi}_{mn}$, because such newspapers have (exogenously) greater presence in highly Republican zip codes. Such a force would bias OLS estimates upward (absent measurement error), but would be addressed by our instrumental variables strategy provided that $E[\tilde{\xi}_{mn} | R_n, r_z, X_z, W_{zn}, \delta_{mn}] = 0$.

Our controls address a range of other possible confounds. Including fixed effects δ_{mn} at the market-newspaper level will control for unobserved newspaper characteristics, unobserved market-level tastes, and heterogeneity in the “fit” between the newspaper and the market (say, because of physical distance). Zip code-level controls X_z account for the fact that demographics like education and race affect readership and may be correlated with political tastes. The interactions W_{zn} account for the fact that these other characteristics may have different effects on readership depending on the average characteristics of a newspaper's market (George and Waldfoegel (2003)). For example, the percent black in a zip code may relate positively to readership of newspapers from predominantly black markets, and negatively on readership of newspapers from predominantly white neighborhoods.

5.2. Supply Parameters

To estimate the supply model of equation (5), we assume that true slant $y_n = y_n^*$, but allow that measured slant $\hat{y}_n \neq y_n$.

Because we can only calculate the profit-maximizing level of slant ideal_n directly for the 290 of newspapers in our demand sample, we approximate ideal_n as a linear function of the Republican vote share in a newspaper's market: $\widehat{\text{ideal}}_n = \eta_0 + \eta_1 R_n + \zeta_n$. This allows us to use our complete sample of 429 newspapers for the supply analysis.

Substituting $\widehat{\text{ideal}}_n$ in place of ideal_n , we then have the estimating equation

$$(9) \quad \hat{y}_n = \lambda_0^s + \lambda_1^s R_n + \mu_g + \omega_n,$$

where $\lambda_0^s = \rho_0 + \rho_1 \eta_0$, $\lambda_1^s = \rho_1 \eta_1$, and $\omega_n = \rho_1 \zeta_n + (\hat{y}_n - y_n)$.

We assume that $\omega_n \sim N(\theta_s, \sigma_\omega^2)$, where s is the newspaper's home state. Here, θ_s is a state-specific measurement error component, with $E(\theta_s) \equiv 0$. We assume that $\mu_g \sim N(\bar{\mu}, \sigma_\mu^2)$, with μ_g , R_n , and ω_n orthogonal conditional on θ_s .

Equation (9) is then a random effects (RE) model. We will control for θ_s flexibly using state fixed effects. Variation in slant that is common to newspapers with the same owner is attributed to variation in μ_g . Newspaper-level variation that is not correlated across newspapers with the same owner is attributed to variation in ω_n .

We include the state-specific measurement error component θ_s in the model because the strong geographic clustering of ownership groups (Lacy and Simon (1997), Martin (2003)) means that any geographic component of measurement error, due to regional patterns of speech or news, could otherwise be spuriously attributed to owner tastes. Inclusion of this component means that variation in owner tastes is identified from correlation in deviations across newspapers with the same owner, after accounting for state effects. Identification therefore relies on the significant number of owners with geographically diverse holdings. Half of the ownership groups with multiple papers in our sample span more than two states. For example, the markets where the New York Times Company owns newspapers range from New York City to Sarasota, FL and Spartanburg, SC.

Our main specifications require that there is no causality running from \hat{y}_n to R_n . We address the possibility of reverse causality below by instrumenting for R_n with consumer religiosity—a characteristic we expect to be a strong predictor of R_n but unaffected by \hat{y}_n .

6. EVIDENCE ON THE DEMAND FOR SLANT

Figure 2 presents evidence on Hypothesis D1. For each newspaper, we regress demand $\ln(S_{zn}/(1 - S_{zn}))$ on zip code ideology r_z , with fixed effects for market. We plot the resulting coefficients against measured slant \hat{y}_n for the 59 newspapers that circulate in markets containing more than 200 zip codes (where coefficients are reasonably well identified). As predicted, the estimated

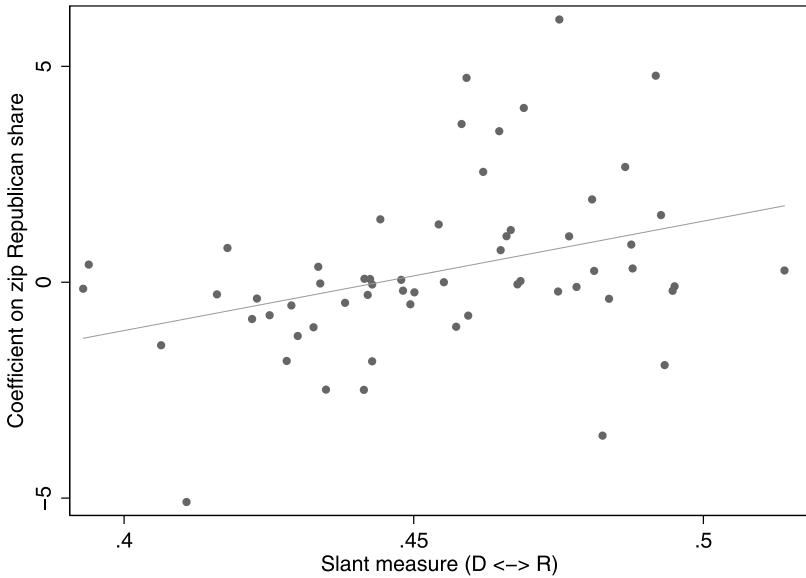


FIGURE 2.—Newspaper slant and coefficients on zip code ideology. The *y* axis shows the estimated coefficient in a regression of the share of households in the zip code reading each newspaper on the zip code share Republican, for newspapers circulating in more than 200 zip codes. The *x* axis shows slant measure.

effect of zip code Republicanism on demand has a clear positive relationship with the newspaper's slant.

Figure 3 presents evidence on Hypothesis D2. Each panel shows, for newspapers in a given quartile of the distribution of measured slant \hat{y}_n , the coefficients on dummies for deciles of zip code ideology r_z , in a regression of demand on decile dummies and market–newspaper fixed effects, weighted by H_z . The graphs are noisy but consistent with an inverted-U relationship, peaking further to the right at higher values of \hat{y}_n .

The first column of Table II presents these findings quantitatively. We regress $\ln(S_{zn}/(1 - S_{zn}))$ on $r_z \hat{y}_n$, r_z , and r_z^2 , and adjust standard errors for correlation at the newspaper level. Consistent with Hypothesis D1, the coefficient on the interaction term $r_z \hat{y}_n$ is positive and statistically significant. Consistent with Hypothesis D2, the coefficient on r_z is negative and statistically significant, and the coefficient on r_z^2 is negative and marginally statistically significant.

The second column of Table II adds controls for zip code demographics X_z and zip code demographics interacted with market demographics W_{zn} . Our findings survive and, if anything, the evidence for Hypothesis D2 becomes stronger statistically.

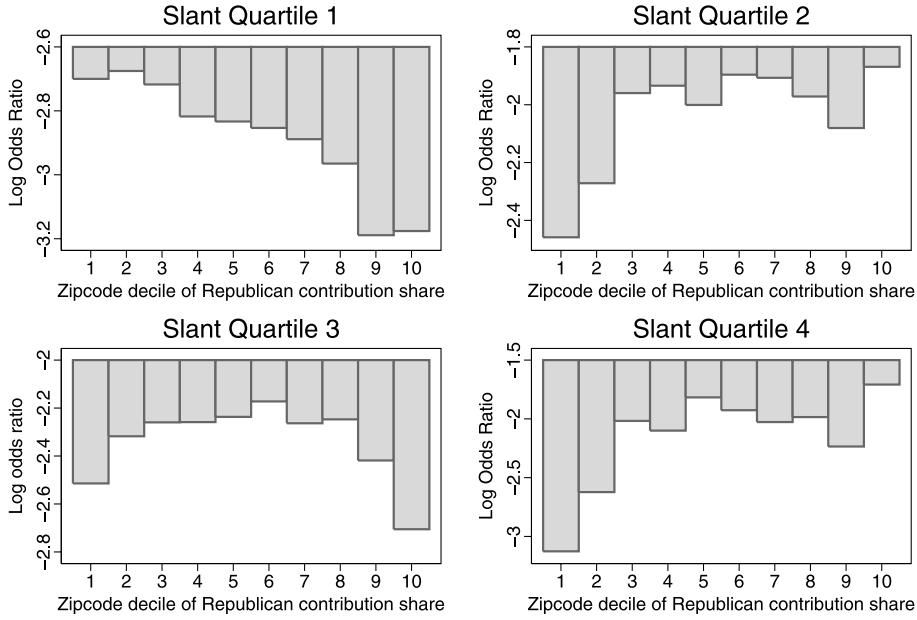


FIGURE 3.—Newspaper demand and zip code ideology by quartiles of newspaper slant. The coefficients on decile dummies in regressions of the share of households in a zip code reading a newspaper on dummies for decile of share donating to Republicans in the 2000–2004 election cycle are shown with market–newspaper fixed effects and weighted by zip code population. The equation is estimated separately for newspapers in each quartile of the distribution of measured slant.

The third column of Table II adds controls for zip code fixed effects. This model is identified from zip codes where two or more newspapers circulate. It allows for unobserved zip code characteristics that affect the overall propensity to read newspapers. In particular, it allows for the possibility that the utility of the outside option varies across zip codes in a way that is correlated with r_z . By definition, we cannot test Hypothesis D2 in this specification, but the evidence for Hypothesis D1 survives.

The last column of Table II presents estimates of our preferred demand model—estimating equation (7) under the assumptions of Section 5.1. We instrument for $r_z \hat{y}_n$ with $r_z R_n$ to address measurement error in \hat{y}_n . As expected, the coefficient on $r_z \hat{y}_n$ increases. The change in magnitude is quantitatively plausible: given that about 63 percent of the variation in \hat{y}_n is measurement error, we would expect its coefficient to be attenuated by a factor of $\frac{1}{1-0.63} \approx 2.7$. In fact, the coefficient in the last column is about 2.6 times that in the second column.

TABLE II
EVIDENCE ON THE DEMAND FOR SLANT^a

Description	Model			
	OLS	OLS	OLS	2SLS
(Zip share donating to Republicans) × Slant	10.66 (3.155)	9.441 (2.756)	14.61 (6.009)	24.66 (7.692)
Zip share donating to Republicans	-4.376 (1.529)	-3.712 (1.274)	—	-10.41 (3.448)
(Zip share donating to Republicans) ²	-0.4927 (0.2574)	-0.5238 (0.2237)	—	-0.7103 (0.2061)
Market-newspaper FE?	X	X	X	X
Zip code demographics?		X	X	X
Zip code X market characteristics?		X	X	X
Zip code FE?			X	
Number of observations	16,043	16,043	16,043	16,043
Number of newspapers	290	290	290	290

^aThe dependent variable is log odds ratio $\ln(S_{zn}) - \ln(1 - S_{zn})$. Standard errors (in parentheses) allow for correlation in the error term across observations for the same newspaper. Zip code demographics are log of total population, log of income per capita, percent of population urban, percent white, percent black, population per square mile, share of houses that are owner occupied, and the share of population aged 25 and over whose highest level of schooling is college, all as of 2000. “Zip code X market characteristics” refers to a vector of these characteristics interacted with their analogue at the level of the newspaper’s market. An excluded instrument in the model in the last column is an interaction between zip share donating to Republicans and share of Republican in the newspaper’s market in 2004. The first-stage *F*-statistic on the excluded instrument is 8.79.

7. EVIDENCE ON THE SUPPLY OF SLANT

7.1. Does Consumer Ideology Affect Slant?

Consistent with Hypothesis S1, slant is highly related to consumer ideology. Figure 4 plots estimated slant \hat{y}_n against the share voting Republican R_n in the newspaper’s market. The graph shows clearly that in more Republican markets, newspapers adopt a more right-wing slant. The first column of Table III shows that in an OLS regression, an increase of 10 percentage points in the share voting Republican translates into an increase in slant of 0.015. This coefficient is highly statistically significant, and variation in consumer preferences explains nearly 20 percent of the variation in slant in this specification.

The relationship between slant and consumer ideology is robust to corrections for possible reverse causality from slant to consumer ideology. The second column of Table III (2SLS (two-stage least squares)) shows that the estimated effect of consumer ideology on slant is similar (though less precise) when we instrument for slant with an estimate of the share of the newspaper’s market attending church monthly or more during 1972–1998. This variable has a large effect on a market’s political leaning (Glaeser, Ponzetto, and Shapiro (2005)), and our estimates using this instrument are valid if the religiosity of

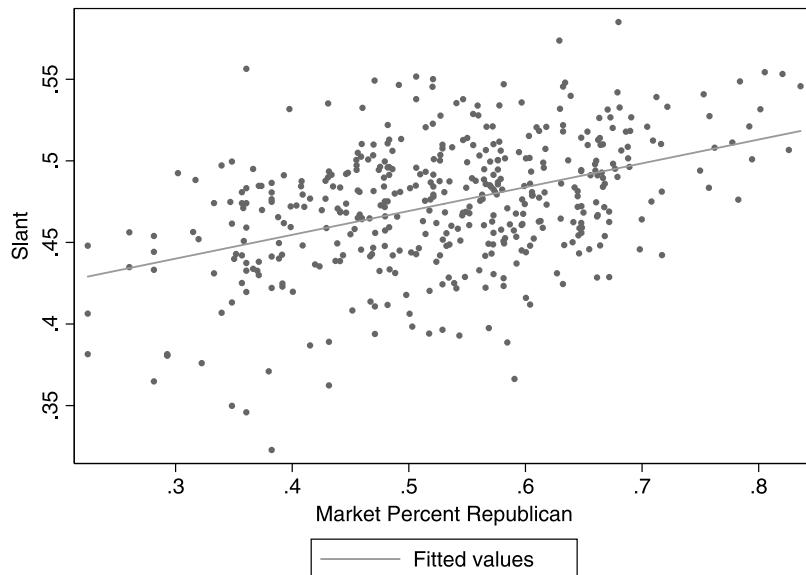


FIGURE 4.—Newspaper slant and consumer ideology. The newspaper slant index against Bush's share of the two-party vote in 2004 in the newspaper's market is shown.

a geographic market is exogenous to the political slant of the market's daily newspaper.

TABLE III
DETERMINANTS OF NEWSPAPER SLANT^a

	OLS	2SLS	OLS	RE
Share Republican in newspaper's market	0.1460 (0.0148)	0.1605 (0.0612)	0.1603 (0.0191)	0.1717 (0.0157)
Ownership group fixed effects?			X	
State fixed effects?				X
Standard deviation (SD) of ownership effect			0.0062 (0.0037)	
Likelihood ratio test that SD of owner effect is zero (<i>p</i> value)			0.1601	
Number of observations	429	421	429	429
R ²	0.1859	—	0.4445	—

^aThe dependent variable is slant index (\hat{y}_n). Standard errors are given in parentheses. An excluded instrument in the 2SLS model is share attending church monthly or more in the newspaper's market during 1972–1998, which is available for 421 of our 429 observations. The first-stage has coefficient 0.2309 and standard error 0.0450. The RE model was estimated via maximum likelihood. See Section 7.2 for details.

The third column of Table III shows that the estimated effect of consumer ideology is similar when we include fixed effects for ownership groups. This confirms that our result is not driven by a tendency of owners to buy papers in markets where consumers' ideology is similar to their own.

In Gentzkow and Shapiro (2007), we reported a number of additional robustness checks. First, we include controls for several measures of newspaper quality (following Berry and Waldfogel (2003)): the log of the newspaper's number of employees, the log of the number of pages, and the number of Pulitzer prizes from 1970 to 2000. Second, we instrument for consumer ideology with a vector of market demographics predictive of voting: log population, percent black, percent with a college degree, percent urban, and log income per capita. Third, we use a preliminary version of our slant measure for the years 2000 and 2004, along with voting data for both years, to estimate a model with newspaper fixed effects. In all cases, the estimated effect of consumer ideology on slant remains large and statistically significant.

7.2. Does Ownership Affect Slant?

Turning to Hypothesis S2, once we account for the propensity of owners to own newspapers in politically and geographically similar markets, we find no evidence that two jointly owned newspapers have a more similar slant than two randomly chosen newspapers. Panel A of Figure 5 plots each newspaper's slant against the average slant of other newspapers with the same owner, revealing a positive and statistically significant correlation. Panel B plots the residual from a regression of slant on the Republican vote share in a paper's market and state fixed effects against the average of this residual among other papers with the same owner. In this panel, there is no visible correlation between the two variables, and the relationship between the variables is no longer significant.

The last column of Table III presents estimates of our preferred supply model—equation (9) under the assumptions of Section 5.2. Our estimate of the variance of the owner effect is small, and we cannot reject the null hypothesis that the variance of the owner effect is zero.

We find no evidence that slant is related to owner ideology, as proxied by political donations. In Figure 6, we plot the relationship between slant and the share of contributions going to Republican candidates for three categories of contributions: (i) those from executives at firms that own multiple U.S. newspapers, (ii) those from executives at independent newspapers (not jointly owned with any other U.S. paper), and (iii) corporate contributions by newspaper firms. The correlation between slant and contributions is weak and statistically insignificant. This remains true in regressions that control for the percent voting Republican in each paper's market (see online Appendix B, Table B.II). Taking donations as a proxy for owner ideology, then, we do not find evidence for Hypothesis S2.

In Gentzkow and Shapiro (2007), we reported additional evidence on the role of ownership in determining slant. We show in a range of random effects

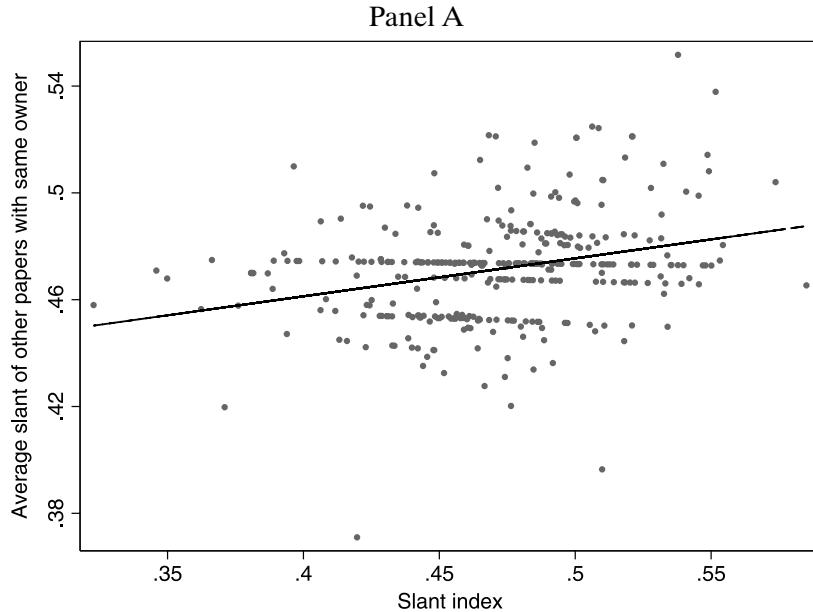


FIGURE 5.—Newspaper slant and ownership. Panel A shows average slant of co-owned newspapers graphed against a newspaper's own slant (correlation = 0.29, $p < 0.001$). Panel B parallels Panel A, but measures slant using residuals from a regression of slant on percent Republican in market and dummies for the state in which the newspaper is located (correlation = 0.09, $p = 0.11$).

models that the owner effect diminishes as we control more tightly for geography, and that it is largely eliminated by controlling for the Republican vote share and Census division fixed effects. In contrast, the role of consumer characteristics grows stronger as we focus on variation in slant within geographic areas. We also examine three important ownership changes that occur during a period (2000–2005) for which we have computed a preliminary slant index. We find no clear evidence that acquired newspapers' slant moves closer to the mean slant of newspapers in the acquiring group.

8. IMPLICATIONS OF THE MODEL

Table IV presents a series of calculations that expose the model's economic implications.

The first row of Table IV presents the observed slant of the average newspaper in the sample. The second row of Table IV presents the profit-maximizing slant of the average newspaper in the sample. Though statistically distinguishable, the two are close in magnitude. At our point estimate, the average news-

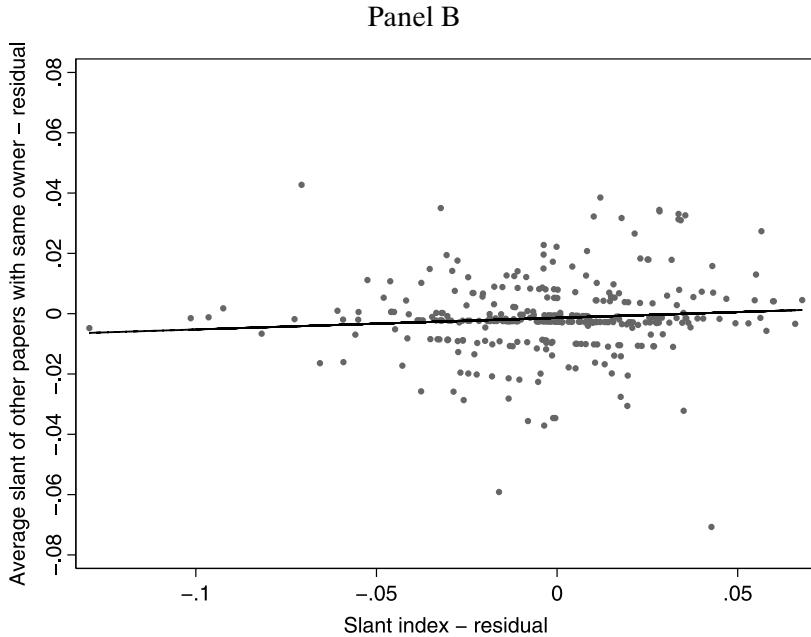


FIGURE 5.—(Continued.)

paper would move slightly to the left in a counterfactual world in which all newspapers choose exactly the profit-maximizing value of slant.

Newspapers could deviate systematically from profit maximization on average due to owner ideology (Balan, DeGraba, and Wickelgren (2009)), pressure from incumbent politicians (Besley and Prat (2006)), or the tastes of reporters (Baron (2006)). A large popular literature has argued that such forces create an overall conservative (Alterman (2003), Franken (2003)) or liberal (Coulter (2003), Goldberg (2003)) bias in the media. Our data do not show evidence of an economically significant bias relative to the benchmark of profit maximization.

The third row of Table IV presents the percent loss in circulation that the average newspaper would experience if it were to deviate by 1 standard deviation from the profit-maximizing level of slant. We estimate an economically large effect of about 18 percent, though the precision of this estimate is limited.

The last two rows of Table IV present the shares of the within-state variation in slant that can be explained by variation in consumer and owner ideology, respectively. At our point estimates, consumer ideology explains 22 percent of the within-state variation in slant, while owner ideology explains only 4 percent. Put differently, our point estimates imply that eliminating cross-market

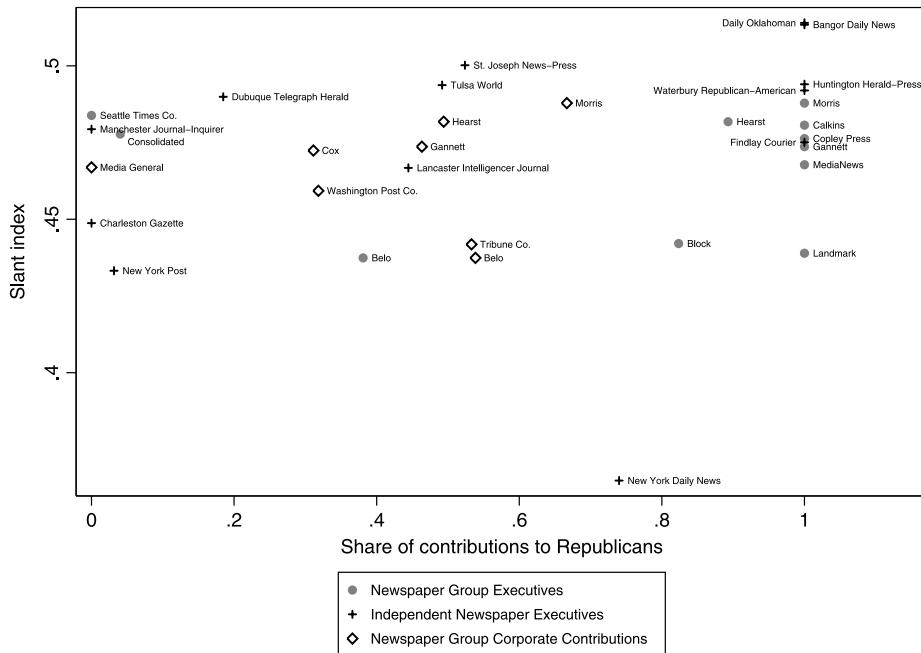


FIGURE 6.—Newspaper slant and political contributions. The average slant of newspapers owned by a firm is graphed against the share of total dollars going to Republicans within each category of contributions. Correlation coefficients are -0.04 ($p = 0.90$) for newspaper group executives, 0.29 ($p = 0.34$) for independent newspaper executives, and 0.01 ($p = 0.97$) for newspaper group corporate contributions.

TABLE IV
ECONOMIC INTERPRETATION OF MODEL PARAMETERS^a

Quantity	Estimate
Actual slant of average newspaper	0.4734 (0.0020)
Profit-maximizing slant of average newspaper	0.4600 (0.0047)
Percent loss in variable profit to average newspaper from moving 1 SD away from profit-maximizing slant	0.1809 (0.1025)
Share of within-state variance in slant from consumer ideology	0.2226 (0.0406)
Share of within-state variance in slant from owner ideology	0.0380 (0.0458)

^aStandard errors, given in parentheses, are from the delta method. The sample in the first three rows includes 290 newspapers in the demand sample. The sample in the last two rows includes 429 newspapers in the supply sample. The calculation in the fourth row is $(\hat{\lambda}_1^s)^2$ times the within-state variance in R_n , divided by the within-state variance of \hat{y}_n . The calculation in the last row is $\hat{\sigma}_{\mu}^2$ divided by the within-state variance of \hat{y}_n .

diversity in consumer ideology would reduce the variance of measured slant by 22 percent, whereas eliminating heterogeneity in owner ideology (say by having all newspapers jointly owned) would reduce it by only 4 percent. We can reject the hypothesis that the share of variance explained by consumers and owners is the same ($p = 0.003$).

9. OTHER DETERMINANTS OF SLANT

We have interpreted the observed relationship between slant and consumer ideology as evidence that newspapers cater to their readers. Here, we consider two alternative explanations:

- (i) Incumbent politicians influence news content (Besley and Prat (2006)), and incumbent politicians' ideology is correlated with consumer ideology.
- (ii) Reporters and editors are drawn from the local population, have ideologies correlated with those of local consumers, and are willing to sacrifice wage income to represent their own views in the newspaper (Baron (2006)).

Ideology of Incumbent Politicians

If incumbent politicians influence news content, then any correlation between incumbent politicians' ideology and consumer ideology could bias our results. In regression models reported in online Appendix B, we find no evidence that slant is related to the party affiliation of local elected officials. Controlling for consumer ideology, having a Republican governor (as of the end of 2005) is associated with a statistically insignificant leftward shift in slant of about 0.9 percentage points, with a confidence interval that rules out a rightward shift larger than about 0.5 percentage points (1/8 of a standard deviation). We also find that, controlling for consumer ideology, the Republican share of representatives to the U.S. House from districts in the newspaper's market (as of the 109th Congress) has a statistically insignificant negative effect on slant. The coefficient implies that moving from a completely Democratic to a completely Republican delegation reduces newspaper slant by 0.004, with a confidence interval that excludes substantial positive effects.

Ideology of Local Reporters and Editors

If local reporters/editors always had the same ideological preferences as consumers, a model where slant responds only to consumers and a model where it also responds to reporters/editors would be observationally equivalent. The important economic question is therefore how slant would be chosen in the event that reporters' and editors' ideologies diverged from those of consumers. For a number of reasons, we believe that it is unlikely that reporter/editor ideology would exert a significant influence in such a case.

Consider a case where consumers' preferred slant is 1 standard deviation to the right of that of local editors and reporters. The local newspaper considers

whether to choose reporters' or consumers' preferred slant. (For simplicity, suppose this choice is either/or.) The cost of satisfying consumer demand is that the newspaper must pay more to bring in qualified reporters and editors from elsewhere and possibly train them in local knowledge, or convince local staff to deviate from their personal ideologies. According to our demand estimates, the benefit is an increase of 18 percent in variable profits. A crude estimate is that the salaries of editors and reporters are on the order of 10 percent of variable profits for a typical newspaper.⁹ Therefore, for reporters' tastes to overwhelm consumer demand, equally qualified reporters willing to report as consumers wish would need to cost 18 percent/10 percent = 180 percent more than those drawn from the local population.

That the cost of qualified reporters could be so high seems especially unlikely given that the market for editors and reporters is not highly localized. In a regression model using Census microdata, we find that reporters and editors are 8 percentage points *more* likely than other professionals to live in a state other than the one in which they were born, controlling for education, age, gender, and race.¹⁰ These "outside" reporters and editors are not of lower quality: reporters and editors born outside their current state of residence earn, if anything, somewhat *more* than those working in their states of nativity. Survey data also show that the average college-educated journalist has nearly a 40 percent chance of working in a Census division other than the one in which he or she attended college (Weaver and Wilhoit (1996)), considerably higher than the average among other college-educated workers.¹¹

Put differently, the elasticity of reporters and editors of different types into a given local market is likely to be very high, as each market draws from the same large national pool of talent. Given consumers' strong demand for like-minded slant, if the tastes of local readers and potential local reporters varied independently, we would expect the tastes of readers to dominate in the determination of equilibrium slant.

As a separate test of the influence of local reporters' ideology, we have constructed a version of our slant measure using only stories written by newspa-

⁹Gentzkow (2007) estimated that the *Washington Post*'s variable profit per *daily* copy sold was \$1.83 in 2004. Applying the same profit rate to Sunday copies (probably an understatement) gives a total yearly variable profit of \$539 million. *Burrelle's/Luce Media Directory 2001* (Burrelle's Information Services (2001)) lists 222 reporters and 175 editors working for the *Post*. If we assume that the average reporter's salary is \$90,000 per year and the average editor's salary is \$125,000 per year, we estimate the *Post*'s wage bill for reporters and editors to be about \$42 million per year, or about 8 percent of variable profits.

¹⁰They are also three percentage points more likely to have moved in the past five years. These figures are coefficients on reporter/editor dummies in regressions using data from the 1980, 1990, and 2000 Censuses (Ruggles et al. (2004)). The sample is restricted to 25- to 55-year-old workers in professional occupations (1950 occupation codes 000–099). Wage regressions reported below are restricted to prime-age male reporters and editors working full time.

¹¹We are extremely grateful to Lisa Kahn for providing the appropriate calculations from the 1979 National Longitudinal Study of Youth (NLSY).

pers' Washington DC bureaus. The reporters and editors of these stories typically live and work in Washington and not in their newspapers' home markets. If slant were determined largely by the geographic home of the editorial staff, we would expect much more homogeneous slant in Washington bureau stories than in locally written stories. In fact, a regression of the slant of Washington bureau stories on consumer ideology yields a positive and statistically significant coefficient, with a value not statistically distinguishable from the coefficient we obtain when we use the overall slant measure. (We note, however, that many papers do not have Washington bureaus, which limits the statistical power of this test.)

Note that the preceding argument is fully consistent with an equilibrium correlation between consumers' and reporters' ideologies; indeed, we would expect such a correlation if reporters have a comparative advantage in writing with a slant consistent with their own views. While we do not have direct evidence on the institutional mechanism through which newspapers "choose" their slant, the choice of editorial staff (along with choice of topics and explicit style policies) seems like a plausible channel through which newspaper content is calibrated to the views of the local population.

10. CONCLUSIONS

In this paper, we develop and estimate a new measure of slant that compares the use of partisan language in newspapers with that of Democrats and Republicans in Congress. Our measure is computable with a minimum of subjective input, is related to readers' subjective ratings of newspaper slant, and is available for newspapers representing over 70 percent of the daily circulation in the United States.

Combining our measure with zip code-level circulation data, we show that consumer demand responds strongly to the fit between a newspaper's slant and the ideology of potential readers, implying an economic incentive for newspapers to tailor their slant to the ideological predispositions of consumers. We document such an effect and show that variation in consumer preferences accounts for roughly one-fifth of the variation in measured slant in our sample.

By contrast, we find much less evidence for a role of newspaper owners in determining slant. While slant is somewhat correlated across co-owned papers, this effect is driven by the geographic clustering of ownership groups. After controlling for the geographic location of newspapers, we find no evidence that the variation in slant has an owner-specific component. We also find no evidence that pressure from incumbent politicians or the tastes of reporters are important drivers of slant.

Taken together, our findings suggest that ownership diversity may not be a critical precondition for ideological diversity in the media, at least along the dimension we consider. This conclusion has broad implications for the regulation of ownership in the media.

We wish to stress three important caveats, however.

First, our measure of slant is a broad aggregate that includes coverage of many different topics over a reasonably long window of time. Owners, politicians, or reporters may still exert significant influence on coverage of specific domains in which their interests are especially strong. For example, Gilens and Hertzman (2009) showed that the 1996 Telecommunications Act received more favorable coverage from newspapers whose parent companies stood to gain from the act's passage. In such areas, where the financial interest of the owner is strong relative to the likely interest of the reader, it is not surprising to see an important effect of ownership, even in light of our finding that ownership is not predictive of our broad index of slant.

Second, our results may not extend to settings with significantly different legal or institutional environments—less developed markets, more state ownership, less freedom of the press. Silvio Berlusconi's influence on Italian media is a case in point (Anderson and McLaren (2009), Durante and Knight (2009)).

Finally, finding that ownership is not an important driver of content diversity does not imply that the market produces the *optimal* level of diversity. In particular, it remains true that virtually all local newspaper markets are monopolies, and the number of independent sources for local news is many cities is correspondingly small. How diversity and welfare are affected by the degree of local newspaper competition remains an important area for future research.

APPENDIX A: DETAILS ON NEWS SEARCHES

A.1. Mechanics of Congressional Record

We use an automated script to download the *Congressional Record* from thomas.loc.gov. Our data base of *Congressional Record* text is incomplete, mostly due to errors in the website that archives the *Congressional Record*. These errors affect a relatively small share of documents in the *Congressional Record* (roughly 15 percent).

We apply a second script to the downloaded text to ascertain the speaker of each passage. We wish to focus on floor speeches rather than text that is primarily procedural, so we exclude speech by officers such as the Clerk, the Speaker of the House, and the President of the Senate. We also exclude block quotations, text that is inserted into the *Record* from other sources such as reports or letters, and nonspeech items like records of roll-call votes.

Before producing phrase counts, we remove extremely common words (“stopwords”). We use the list from Fox (1990), augmented with a list of proper nouns that appear frequently in procedural text—days of the week, the Hart

Senate Office Building, and the Dirksen Senate Office Building. We also exclude the names of major newspapers.

We use the Porter Stemmer (tartarus.org/martin/PorterStemmer/) to strip words down to their linguistic roots. This means that phrases in the *Congressional Record* that differ only in either stopwords or suffixes are equivalent in our algorithm. For example, “war on terror,” “war against terror,” and “wars on terror” would all appear in the preprocessed *Congressional Record* as “war terror” and thus be treated as the same phrase.

A.2. Mechanics of Newspaper Searches

Following the steps outlined in Section 3.1, we identify 1000 phrases to use in our analysis. We wish to count the number of times each of these 1000 phrases appears in each of our sample of newspapers using the ProQuest and NewsLibrary data bases.

Among our 433 newspapers, data are available for 394 from NewsLibrary and for 164 from ProQuest, with an overlap of 125 newspapers. Among the newspapers that overlap between the two data bases, the correlation between the counts for our 1000 phrases is 0.85. In cases of overlap, we use the NewsLibrary counts for analysis.

The two data bases do not agree perfectly for several reasons, including differences in the set of articles newspapers choose to post to each data base and differences in how the two data bases permit us to identify editorials and opinion pieces (see below). An important third reason is that the data bases are dynamic: content is added over time, so that searches conducted at different times may produce different results. As a consequence, one potential source of disagreement between ProQuest and NewsLibrary is a difference in the posting lag between the two data bases.

Because of the preprocessing steps above (stopword removal and stemming), each of our 1000 phrases thus corresponds to a group of one, two, or several original phrases, and it is these original phrases that we search for in the data bases.

The set of original phrases we search is slightly restricted for two reasons. First, the ProQuest data base limits search strings to 75 characters. We therefore drop any original phrase longer than 75 characters. Second, our data base of *Congressional Record* text has improved over time as we have adjusted for errors in the source website and improved our parsing algorithm. The set of original phrases included in each group is based on a slightly older version of the *Congressional Record* text than the one used for our main analysis, so it omits some relatively rare original phrases.

We search for each group of original phrases (connected with the OR operator) in the All Text field (NewsLibrary) or Document Text field (ProQuest), restricted to 2005 and with the following terms excluded from the Headline and Author fields: “editor,” “editorial,” “associated press,” “ap,” “opinion,” “op-ed,” and “letter.”

A.3. Audit Study

Our searches are designed to isolate the slant of news content produced independently by each paper. The way stories are archived and classified in the data bases means that we can only imperfectly separate these stories from other kinds of content such as opinion pieces and wire stories. To provide a more precise picture of the kinds of content we are measuring, we have audited the results for seven phrases chosen from Table I. For each phrase, we looked at the full set of hits for the papers included in the NewsLibrary data base and recorded whether they appeared to be (i) independently produced news stories, (ii) AP wire stories, (iii) other wire stories, (iv) letters to the editor, or (v) opinion pieces (including unsigned editorials). Because we do not have access to the full text of articles in NewsLibrary, this classification is based on the headline and first paragraph of the story.

In a separate exercise, we use results from the papers we can search in the ProQuest data base (for which we can retrieve full text articles) to record the number of times each phrase appears in quotation.

The results are shown in Table A.I. Overall, approximately 71 percent of our hits are independently produced news stories. Of the remainder, 22 percent are either clearly or possibly opinion, 3 percent are letters to the editor, and 3 percent are wire stories. The table also shows that these shares are heterogeneous across phrases. For example, the share of opinion pieces ranges from 12 percent for “global war on terrorism” to 51 percent for “death tax.” The results also show that only 10 percent of our hits appear in quotations, with the share ranging from 3 percent for “child support enforcement” to 36 percent for “death tax.” We have also spot checked the articles that are being excluded from our search results and verified that virtually all of them are, as desired, either wire stories or opinion pieces.

As a final check, we have also computed the share of phrases appearing in direct quotes of local congresspeople, which could cause a mechanical correlation between slant and the political leanings of local markets. Among 10 randomly chosen papers (representing different levels of circulation), we hand coded the frequency of uses of the top 50 phrases in direct quotes of congresspeople. On average, such quotes account for only 0.3 percent of the phrase hits in this sample.

Taken together, the results confirm that our measure is primarily picking up the slant of independently produced news stories, with some weight given to opinion pieces.

REFERENCES

- AAS, K., AND L. EIKVIL (1999): “Text Categorisation: A Survey,” Report 941, Norwegian Computing Center. [38]
- AKHAVAN-MAJID, R., A. RIFE, AND S. GOPINATH (1991): “Chain Ownership and Editorial Independence: A Case Study of Gannett Newspapers,” *Journalism and Mass Communication Quarterly*, 68, 59–66. [38]

TABLE A.I
AUDIT OF SEARCH RESULTS^a

Phrase	Total Hits	Share of Hits in Quotes	Share of Hits That Are					
			AP Wire Stories	Other Wire Stories	Letters to the Editor	Maybe Opinion	Clearly Opinion	Independently Produced News
Global war on terrorism	2064	16%	3%	4%	1%	2%	10%	80%
Malpractice insurance	2190	5%	0%	0%	1%	3%	12%	84%
Universal health care	1523	9%	1%	0%	7%	8%	28%	56%
Assault weapons	1411	9%	3%	12%	4%	1%	25%	56%
Child support enforcement	1054	3%	0%	0%	1%	2%	11%	86%
Public broadcasting	3375	8%	1%	0%	2%	4%	22%	71%
Death tax	595	36%	0%	0%	2%	5%	46%	47%
Average (hit weighted)		10%	1%	2%	3%	3%	19%	71%

^aAuthors' calculations based on ProQuest and NewsLibrary data base searches. See Appendix A for details.

- ALTERMAN, E. (2003): *What Liberal Media? The Truth About Bias and the News*. Basic Books. [60]
- ANDERSON, S. P., AND J. MCLAREN (2009): "Media Mergers and Media Bias With Rational Consumers," Working Paper, University of Virginia. [65]
- ANSOLABEHERE, S., R. LESSEM, AND J. M. SNYDER JR. (2006): "The Orientation of Newspaper Endorsements in U.S. Elections, 1940–2002," *Quarterly Journal of Political Science*, 1, 393–404. [38]
- ANTWEILER, W., AND M. Z. FRANK (2004): "Is All That Talk Just Noise? The Information Content of Internet Message Boards," *Journal of Finance*, 59, 1259–1294. [38]
- BAGDIKIAN, B. H. (2000): *The Media Monopoly* (Sixth Ed.). Boston: Beacon Press. [36]
- BALAN, D. J., P. DEGRABA, AND A. L. WICKELGREN (2009): "Ideological Persuasion in the Media," Mimeo, Federal Trade Commission. [38,60]
- BARON, D. P. (2006): "Persistent Media Bias," *Journal of Public Economics*, 90, 1–36. [37,38,60,62]
- BERRY, S., AND J. WALDFOGEL (2003): "Product Quality and Market Size," Working Paper 9675, NBER. [58]
- BESLEY, T., AND A. PRAT (2006): "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability," *American Economic Review*, 96, 720–736. [37,38,60,62]
- BURRELLE'S INFORMATION SERVICES (2001): *Burrelle's/Luce Media Directory—2001 Edition*. Livingston, NJ: Burrelle's Information Services. [63]
- COMMISSION ON FREEDOM OF THE PRESS (1947): *A Free and Responsible Press: A General Report on Mass Communication: Newspapers, Radio, Motion Pictures, Magazines, and Books*. Chicago, IL: The University of Chicago Press. [36]
- COULTER, A. (2003): *Slander: Liberal Lies About the American Right*. New York: Three Rivers Press. [60]
- DELLAVIGNA, S., AND E. KAPLAN (2007): "The Fox News Effect: Media Bias and Voting," *Quarterly Journal of Economics*, 122, 1187–1234. [36]
- DRANOVA, D., A. GRON, AND M. J. MAZZEO (2003): "Differentiation and Competition in HMO Markets," *Journal of Industrial Economics*, 51, 433–454. [38]
- DUBÉ, J.-P., G. J. HITSCH, AND P. MANCHANDA (2005): "An Empirical Model of Advertising Dynamics," *Quantitative Marketing and Economics*, 3, 107–144. [38]
- DURANTE, R., AND B. KNIGHT (2009): "Partisan Control, Media Bias, and Viewer Responses: Evidence From Berlusconi's Italy," Working Paper 14762, NBER. [65]
- EINAV, L. (2007): "Seasonality in the US Motion Picture Industry," *RAND Journal of Economics*, 38, 127–145. [38]
- FEDERAL COMMUNICATIONS COMMISSION (2003): *Report and Order and Notice of Proposed Rule-making*. Washington, DC: Federal Communications Commission. [36]
- FOX, C. (1990): "A Stop List for General Text," *SIGIR FORUM*, 24, 19–35. [65]
- FRANKEN, A. (2003): *Lies and the Lying Liars Who Tell Them: A Fair and Balanced Look at the Right*. Boston, MA: E. P. Dutton. [60]
- GENTZKOW, M. (2006): "Television and Voter Turnout," *Quarterly Journal of Economics*, 121, 931–972. [36]
- _____. (2007): "Valuing New Goods in a Model With Complementarity: Online Newspapers," *American Economic Review*, 97, 713–744. [41,63]
- GENTZKOW, M. A., AND J. M. SHAPIRO (2004): "Media, Education, and Anti-Americanism in the Muslim World," *Journal of Economic Perspectives*, 18, 117–133. [36]
- _____. (2006): "Media Bias and Reputation," *Journal of Political Economy*, 114, 280–316. [37, 38]
- _____. (2007): "What Drives Media Slant? Evidence From U.S. Daily Newspapers," Working Paper 12707, NBER. [49,50,58]
- _____. (2010): "Supplement to 'What Drives Media Slant? Evidence From U.S. Daily Newspapers,'" *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/7195_tables-figures.pdf. [39]
- GENTZKOW, M. A., E. L. GLAESER, AND C. D. GOLDIN (2006): "The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered," in *Corruption and Reform*:

- Lessons From America's Economic History*, ed. by E. L. Glaeser and C. Goldin. Chicago, IL: University of Chicago Press, 187–230, Chapter 6. [38]
- GEORGE, L. (2007): “What’s Fit to Print: The Effect of Ownership Concentration on Product Variety in Newspaper Markets,” *Information Economics and Policy*, 19, 285–303. [38]
- GEORGE, L., AND J. WALDFOGEL (2003): “Who Affects Whom in Daily Newspaper Markets?” *Journal of Political Economy*, 111, 765–784. [37,51,52]
- GERBER, A. S., D. KARLAN, AND D. BERGAN (2009): “Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions,” *American Economic Journal: Applied Economics*, 1 (2), 35–52. [36]
- GILENS, M., AND C. HERTZMAN (2009): “Corporate Ownership and News Bias: Newspaper Coverage of the 1996 Telecommunications Act,” *Journal of Politics*, 62, 369–386. [65]
- GIMPEL, J. G., F. E. LEE, AND J. KAMINSKI (2006): “The Political Geography of Campaign Contributions in American Politics,” *Journal of Politics*, 68, 626–639. [41]
- GLAESER, E. L., G. A. M. PONZETTO, AND J. M. SHAPIRO (2005): “Strategic Extremism: Why Republicans and Democrats Divide on Religious Values,” *Quarterly Journal of Economics*, 120, 1283–1330. [56]
- GLASSER, T. L., D. S. ALLEN, AND S. E. BLANKS (1989): “The Influence of Chain Ownership on News Play: A Case Study,” *Journalism Quarterly*, 66, 607–614. [36]
- GOLDBERG, B. (2003): *Bias: A CBS Insider Exposes How the Media Distort the News*. New York: Perennial. [60]
- GRAETZ, M. J., AND I. SHAPIRO (2005): *Death by a Thousand Cuts: The Fight Over Taxing Inherited Wealth*. Princeton, NJ: Princeton University Press. [42]
- GROSECLOSE, T., AND J. MILYO (2005): “A Measure of Media Bias,” *Quarterly Journal of Economics*, 120, 1191–1237. [36,38,42]
- HAMILTON, J. T. (2004): *All the News That's Fit to Sell: How the Market Transforms Information Into News*. Princeton, NJ: Princeton University Press. [38]
- HARRIS INTERACTIVE (2006): “Seven in 10 U.S. Adults Say They Watch Broadcast News at Least Several Times a Week,” *The Harris Poll*, 20. [36]
- LACY, S., AND T. F. SIMON (1997): “Intercounty Group Ownership of Daily Newspapers and the Decline of Competition for Readers,” *Journalism and Mass Communication Quarterly*, 74, 814–825. [53]
- LARCINESE, V., R. PUGLISI, AND J. M. SNYDER (2007): “Partisan Bias in Economic News: Evidence on the Agenda-Setting Behavior of U.S. Newspapers,” Working Paper 13378, NBER. [38]
- LAVER, M., K. BENOIT, AND J. GARRY (2003): “Extracting Policy Positions From Political Texts Using Words as Data,” *American Political Science Review*, 97, 311–331. [38]
- LUNTZ, F. (2005): *Learning From 2004 ... Winning in 2006*. Washington, DC: Luntz Research Companies. [44,45]
- MARTIN, H. J. (2003): “Some Effects From Horizontal Integration of Daily Newspapers on Markets, Prices, and Competition,” in *Proceedings of the Annual Meeting of the Association for Education in Journalism and Mass Communication, Media Management and Economics Division*, Kansas City, MO, July 30–August 2, 2003. [53]
- MAZZEO, M. J. (2002a): “Product Choice and Oligopoly Market Structure,” *RAND Journal of Economics*, 33, 221–242. [38]
- (2002b): “Competitive Outcomes in Product-Differentiated Oligopoly,” *Review of Economics and Statistics*, 84, 716–728. [38]
- MULLAINATHAN, S., AND A. SHLEIFER (2005): “The Market for News,” *American Economic Review*, 95, 1031–1053. [37,38,48]
- MYERS, C. K. (2008): “Discrimination as a Competitive Device: The Case of Local Television News,” *The B.E. Journal of Economic Analysis and Policy*, 8, Article 28. [38]
- NEWSPAPER ASSOCIATION OF AMERICA (2006): “The Source: Newspapers by the Numbers.” [36]
- PRITCHARD, D. (2002): “Viewpoint Diversity in Cross-Owned Newspapers and Television Stations: A Study of News Coverage of the 2000 Presidential Campaign,” Working Paper, FCC Media Ownership Working Group. [36]

- PUGLISI, R. (2008): "Being the New York Times: The Political Behavior of a Newspaper," Mimeo, ECARES-ULB. [38]
- PUTNAM, R. D. (2000): *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster. [40]
- RUGGLES, S., M. SOBEK, T. ALEXANDER, C. A. FITCH, R. GOEKEN, P. K. HALL, M. KING, AND C. RONNANDER (2004): *Integrated Public Use Microdata Series: Version 3.0*. Minneapolis, MN: Minnesota Population Center. Available at <http://www.ipums.org>. [63]
- SEIM, K. (2006): "An Empirical Model of Firm Entry With Endogenous Product-Type Choices," *RAND Journal of Economics*, 37, 619–640. [38]
- STEVENSON, R. W. (2005): "President Makes It Clear: Phrase Is 'War on Terror,'" *New York Times*, August 4. [45]
- STRÖMBERG, D. (2004): "Radio's Impact on Public Spending," *Quarterly Journal of Economics*, 119, 189–221. [36]
- SUEN, W. (2004): "The Self-Perpetuation of Biased Beliefs," *Economic Journal*, 114, 377–396. [38]
- SWEETING, A. (2007): "Dynamic Product Repositioning in Differentiated Product Industries: The Case of Format Switching in the Commercial Radio Industry," Working Paper 13522, NBER. [38]
- _____. (2008): "The Effects of Horizontal Mergers on Product Positioning: Evidence From the Music Radio Industry," Mimeo, Duke University. [38]
- U.S. SUPREME COURT (1945): "Associated Press v. United States," Washington, DC. [35]
- WEAVER, D. H., AND G. C. WILHOIT (1996): *The American Journalist in the 1990s: U.S. News People at the End of an Era*. Mahwah, NJ: Lawrence Erlbaum Associates. [63]

*University of Chicago Booth School of Business, 5807 S. Woodlawn Avenue,
Chicago, IL 60637, U.S.A. and NBER; gentzkow@chicagobooth.edu
and*

*University of Chicago Booth School of Business, 5807 S. Woodlawn Avenue,
Chicago, IL 60637, U.S.A. and NBER; jesse.shapiro@chicagobooth.edu.*

Manuscript received May, 2007; final revision received August, 2009.

LARGE RISKS, LIMITED LIABILITY, AND DYNAMIC MORAL HAZARD

BY BRUNO BIAIS, THOMAS MARIOTTI, JEAN-CHARLES ROCHE
AND STÉPHANE VILLENEUVE¹

We study a continuous-time principal–agent model in which a risk-neutral agent with limited liability must exert unobservable effort to reduce the likelihood of large but relatively infrequent losses. Firm size can be decreased at no cost or increased subject to adjustment costs. In the optimal contract, investment takes place only if a long enough period of time elapses with no losses occurring. Then, if good performance continues, the agent is paid. As soon as a loss occurs, payments to the agent are suspended, and so is investment if further losses occur. Accumulated bad performance leads to downsizing. We derive explicit formulae for the dynamics of firm size and its asymptotic growth rate, and we provide conditions under which firm size eventually goes to zero or grows without bounds.

KEYWORDS: Principal–agent model, limited liability, continuous time, Poisson risk, downsizing, investment, firm size dynamics.

1. INTRODUCTION

INDUSTRIAL AND FINANCIAL FIRMS are subject to large risks: the former are prone to accidents and the latter are exposed to sharp drops in the value of their assets. Preventing these risks requires managerial effort. Systematic analyses of industrial accidents point to the role of human deficiencies and inadequate levels of care.² A striking illustration is offered by the explosion at

¹We thank a co-editor, and three anonymous referees for very thoughtful and detailed comments. We also thank Ron Anderson, Dirk Bergemann, Peter DeMarzo, Ivar Ekeland, Eduardo Faingold, Michael Fishman, Jean-Pierre Florens, Xavier Gabaix, Christian Gollier, Alexander Gümbel, Zhiguo He, Christian Hellwig, Augustin Landier, Ali Lazrak, Erzo Luttmer, George Mailath, Roger Myerson, Thomas Philippon, Bernard Salanié, Yuliy Sannikov, Hyun Song Shin, Dimitri Vayanos, Nicolas Vieille, and Wei Xiong for very valuable feedback. Finally, we thank seminar audiences at Imperial College London, New York University, Oxford-Man Institute of Quantitative Finance, Princeton University, Universidad Carlos III de Madrid, Université Paris 1, Universiteit van Tilburg, University College London, University of Edinburgh, University of Warwick, Wissenschaftszentrum Berlin für Sozialforschung, and Yale University, as well as conference participants at the 2007 CEPR Corporate Finance and Risk Management Conference, the 2007 IDEI-LERNA Conference on Environment and Resource Economics, the 2007 Oxford Finance Summer Symposium, the 2008 Cowles Summer Conference, the 2008 Pacific Institute for the Mathematical Sciences Summer School in Finance, the 2009 Institut Finance Dauphine Workshop on Dynamic Risk Sharing, and the 2009 Paul Woolley Centre for Capital Market Dysfunctionality Conference for many useful discussions. Financial support from the Chaire Finance Durable et Investissement Responsable, the Chaire Marchés des Risques et Création de Valeur, the ERC Starting Grant 203929-ACAP, the Europlace Institute of Finance, and the Fédération Française des Sociétés d'Assurances is gratefully acknowledged.

²See, for instance, Leplat and Rasmussen (1984), Gordon, Flin, Mearns, and Fleming (1996), or Hollnagel (2002).

the BP Texas refinery in March 2005. After investigating the case, the Baker panel concluded that “BP executive and corporate refining management have not provided effective process safety leadership.”³ Similarly, the large losses incurred by banks and insurance companies during the recent financial crisis were in part due to insufficient risk control. These large risks present a major challenge to firms, investors, and citizens. This paper studies the design of incentives to mitigate them.

One way to stimulate the prevention of large risks would be to make managers and firms bear the social costs that they generate. Yet, this is often impossible in practice, because total damages often exceed the wealth of managers and even the net worth of firms, while the former are protected by limited liability and the latter by bankruptcy laws.⁴ This curbs managers’ incentives to reduce the risk of losses that exceed the value of their own assets.⁵ Of course, if the risk prevention activities undertaken by managers were observable, it would be straightforward to design compensation schemes that would induce them to take socially optimal levels of risk. To a large extent, however, these activities are unobservable by external parties, which leads to a moral hazard problem.

In addition to informational asymmetries, another important aspect of large risks lies in their timing. Large losses are relatively rare events that contrast with day-to-day firm operations and cash flows.⁶ It is, therefore, natural to study large risk prevention in a dynamic setup, where the timing of losses differs from that of operations. To do so, we focus on the simplest model: operating cash flows are constant per unit of time, while losses occur according to a Poisson process whose intensity depends on the level of risk prevention.

In this context, we study the optimal contract between a principal and an agent that provides the latter with appropriate incentives to reduce the risk of losses under dynamic moral hazard. The agent, who can be thought of as an entrepreneur or a manager running a business, is risk-neutral and protected by limited liability. She can exert effort to reduce the instantaneous probability

³“The Report of the BP U.S. Refineries Independent Safety Review Panel,” January 16, 2007. Also, Chemical Safety Board Chairman Carolyn W. Merritt stated that “BP’s global management was aware of problems with maintenance, spending, and infrastructure well before March 2005. [...] Unsafe and antiquated equipment designs were left in place, and unacceptable deficiencies in preventative maintenance were tolerated” (“CSB Investigation of BP Texas City Refinery Disaster Continues as Organizational Issues Are Probed,” CSB News Release, October 30, 2006).

⁴For instance, Katzman (1988) reported that “In *Ohio v. Kovacs* (U.S.S.C. 83-1020), the U.S. Supreme Court unanimously ruled that an industrial polluter can escape an order to clean up a toxic waste site under the umbrella of federal bankruptcy.” Similarly, the social losses created by the recent financial crisis exceeded by far the assets one could withhold from financial executives.

⁵Shavell (1984, 1986) discussed how a party’s inability to pay for the full magnitude of harm done dilutes its incentives to reduce risk.

⁶From now on, we generically refer to any realization of a large risk as a loss.

of losses.⁷ Effort is costly to the agent and unobservable by other parties. The project run by the agent can expand, through investment, or shrink, through downsizing. While downsizing is unconstrained, we assume that the pace of investment is limited by adjustment costs, in the spirit of Hayashi (1982) or Kydland and Prescott (1982). We also assume constant returns to scale, in that downsizing and investment affect, by the same factor, the operating profits of the project, the social costs of accidents, and the private benefits that the agent derives from shirking. This assumption implies that the principal's value function is homogeneous in size and enables us to characterize the optimal contract explicitly. However, as discussed in the paper, some of our key qualitative results are robust to relaxing the constant returns to scale assumption.

The optimal contract maximizes the expected value that the principal derives from an incentive feasible risk prevention policy. It relies on two instruments: positive payments to the agent and project size management through downsizing and investment. While these decisions are functions of the entire past history of the loss process, this complex history dependence can be summarized by two state variables: the size of the project and the continuation utility of the agent. The former reflects the history of past downsizing and investment decisions, while the latter reflects the prospect of future payments to the agent. The evolution of the agent's continuation utility mirrors the dynamics of losses and thus serves as a track record of the agent's performance.⁸ We characterize the compensation and size management policy that arise in the optimal contract.

First consider the compensation policy. To motivate the agent, the optimal contract relies on the promise of payments after good performance and the threat of reductions in her continuation utility after losses. When the track record of the agent is relatively poor, there is a probation phase during which she does not receive any payment. As long as no loss occurs, the size-adjusted continuation utility of the agent increases until it reaches a threshold at which she receives a constant wage per unit of time and size of the project, such that her size-adjusted continuation utility remains constant. As soon as a loss occurs, the continuation utility of the agent undergoes a sharp reduction and the contract reverts to the probation phase. The magnitude of that reduction in the agent's continuation utility is pinned down by the incentive compatibility constraint. The more severe the moral hazard problem and the larger the project, the greater the punishment. The induced sensitivity of the agent's continuation utility to the random occurrence of losses is socially costly because the principal's value function is concave in that state variable. Therefore, it is optimal

⁷Unlike in Shapiro and Stiglitz (1984) or Akerlof and Katz (1989), effort in our model merely makes losses less likely, but does not eliminate them altogether. As a result, losses do occur on the equilibrium path, and it is no longer optimal to systematically terminate the principal–agent relationship following a loss.

⁸That the optimal contract exhibits memory is a standard feature of dynamic moral hazard models, see for instance Rogerson (1985).

to set the reduction in the agent's continuation utility following a loss to the minimum level consistent with incentive compatibility.

Next consider the dynamics of the size of the project. In the first-best case, there is no need for downsizing. Since the project has positive net present value, investment then always takes place at the highest feasible rate so as to maximize the size of the project. In the second-best case, however, the size of the project is lower than in the first-best. The intuition is the following. As mentioned above, the agent is partly motivated by the threat of reductions in her continuation utility in case of bad performance. Yet, when the continuation utility of the agent is low, the threat to reduce it further has limited bite, because of limited liability. To cope with this limitation, it can be necessary to lower the agent's temptation to shirk by reducing the scale of operations after losses. Apart from such circumstances, and in particular when no loss occurs, the project is never downsized. In addition to downsizing, moral hazard also affects the size of the project through its impact on investment. Since increases in the size of the project raise the temptation to shirk, investment can take place only when the agent has enough at stake in the project, that is, when her track record has been good enough for her continuation utility to reach a given threshold. While payments when they occur are costly for the principal, investment benefits both parties. As long as investment takes place, the total size of the pie grows, which in turn makes delaying the compensation of the agent less costly. Thus it is efficient to invest before actually compensating the agent. Note that the sequencing of compensation and investment is reversed in the first-best case. This is because the agent, who is assumed to be more impatient than the principal, then receives all her compensation at time zero, before any investment actually takes place.

We obtain an explicit formula that maps the path of the agent's size-adjusted continuation utility into the size of the project. If one interprets the latter as firm size, this formula exactly spells out how firm size grows, stays constant, or declines over time. Relying on asymptotic theory for Markov ergodic processes, we then characterize the long-run growth rate of the firm. In the first-best case, firm size goes to infinity at a constant rate. Our formula for the long-run growth rate of the firm shows how, in the second-best case, this trend in firm size is reduced by downsizing and possibly lower investment rates. When the adjustment costs are high, firm size eventually goes to zero. By contrast, when both the adjustment costs and the frequency of losses are low, firm size eventually goes to infinity, although more slowly than in the first-best case.

Our paper belongs to the rich and growing literature on dynamic moral hazard that uses recursive techniques to characterize optimal dynamic contracts.⁹

⁹See, for instance, Green (1987), Spear and Srivastava (1987), Thomas and Worrall (1990), or Phelan and Townsend (1991) for seminal contributions along these lines. By focusing on the case where the agent is risk-neutral, with limited liability, our model is in line with the recent papers by Clementi and Hopenhayn (2006) and DeMarzo and Fishman (2007a, 2007b).

One of our contributions relative to this literature is to study the case where moral hazard is about large but relatively infrequent risks. As illustrated by recent industrial accidents or by the recent financial crisis, preventing such risks is a major challenge. We show that optimal contracts that mitigate the risk of infrequent but large losses differ markedly from those that prevail when fluctuations in the output process are frequent but infinitesimal. In the latter, as illustrated by the Brownian motion models of DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin, and Rochet (2007), or Sannikov (2008), the continuation utility of the agent continuously fluctuates until it reaches zero, an event that is predictable. At this point, the project is liquidated. In contrast, with Poisson risk, the continuation utility of the agent increases smoothly most of the time, but incurs sharp decreases when losses occur. In this context, incentive compatibility together with limited liability imply unpredictable downsizing, unlike in the Brownian case.

Another contribution of this paper relative to the literature is to analyze the interplay between incentive considerations and firm size dynamics, and in particular to study the long-run impact of downsizing and investment on firm size under moral hazard. Our analysis of the interactions between incentives and investment is in line with DeMarzo and Fishman (2007a). In a finite horizon, discrete-time framework, they derived a number of predictions regarding the relationship between current investment, current and past cash flows, and agent's compensation. They showed that these predictions are relatively insensitive to the specific nature of the agency problem, provided its static version has a certain structure. Thanks to the finiteness of the horizon, these results are derived recursively, starting from the final period. Our analysis first differs from DeMarzo and Fishman's (2007a) in that our starting point is a stationary continuous-time model, which raises further conceptual and technical difficulties. Second, to derive sharper implications from the analysis, we consider a particular type of informational friction, namely a moral hazard problem with Poisson uncertainty. This modeling approach enables us to precisely characterize the properties of the optimal contract, to provide an explicit formula for the dynamics of firm size, and ultimately to conduct an asymptotic analysis of its long-run evolution and that of the agent's utility. In particular, a key insight of our analysis is that, when investment is taken into account, it need not be the case that the firm eventually vanishes and that the agent's utility eventually goes to zero. This contrasts with the classic immiserization result of Thomas and Worrall (1990). This also contrasts with the contemporaneous work by DeMarzo, Fishman, He, and Wang (2008), who studied the dynamics of average and marginal q in a Brownian model of agency and investment with convex adjustment costs and constant returns to scale. In their model, as in ours, the agent's continuation utility and the current capital stock are sufficient statistics for the optimal contract. An important difference is that, in DeMarzo, Fishman, He, and Wang (2008), the firm will eventually be liquidated when the agent's size-adjusted utility reaches zero, which occurs with

probability 1. By contrast, in our Poisson model, the size-adjusted utility of the agent is bounded away from zero, and incentives are provided by partial downsizing instead of outright liquidation. As a result, the firm can grow without bounds when adjustment costs are low enough so that investment outweighs downsizing.

In the context of a political economy model, Myerson (2008) contemporaneously offered an analysis of dynamic moral hazard in a Poisson framework. A distinctive feature of our paper is that we analyze the impact of investment on the principal–agent relationship. Moreover, Myerson (2008) considered the case where the principal and the agent have identical discount rates. This case, however, is not conducive to continuous-time analysis, as an optimal contract does not exist. To cope with this difficulty, Myerson (2008) imposed an exogenous upper bound on the continuation utility of the agent. By contrast, we do not impose such a constraint on the set of feasible contracts. Instead, we consider the case where the principal is less impatient than the agent. While this makes the formal analysis more complex, it also restores the existence of an unconstrained optimal contract.

Sannikov (2005) also used a Poisson payoff structure. A key difference with our analysis lies in the way output is affected by the jumps of the Poisson process. In Sannikov (2005), jumps correspond to positive cash-flow shocks, while in our model they correspond to losses that are less likely to occur if the agent exerts effort.¹⁰ This leads to qualitatively very different results. While downsizing is a key feature of our optimal contract, as it ensures that incentives can still be provided following a long sequence of losses, it plays no role in Sannikov (2005). Liquidation in his model is still required to provide incentives, but it corresponds to a predictable event: if a sufficiently long period of time elapses during which the agent reports no cash flow, the firm is liquidated. By contrast, downsizing in our model is unpredictable.¹¹

Our paper is also related to the literature on accident law. Shavell (1986, 2000) argued that the desirability of liability insurance depends on the ability of insurers to monitor the firm's prevention effort and to link insurance premia to the observed level of care. If insurers cannot observe the firm's level of care, making full liability insurance mandatory results in no care at all being taken.¹² In our dynamic analysis, the optimal contract ties the firm's allowed activity level to its performance record: following a series of losses, the firm can be forbidden to engage at full scale in its risky activity. These instruments provide

¹⁰Thus jumps in our model are bad news in the sense of Abreu, Milgrom, and Pearce (1991).

¹¹Poisson processes have also proved useful in the theory of repeated games with imperfect monitoring; see, for instance, Abreu, Milgrom, and Pearce (1991), Kalesnik (2005), and Sannikov and Skrzypacz (2010). Our focus differs from theirs in that we consider a full commitment contracting environment, in which we explicitly characterize the optimal incentive compatible contract.

¹²See Jost (1996) and Polborn (1998) for important extensions and qualifications of this argument.

the manager of the firm with dynamic incentives to exert the appropriate risk prevention effort, although the latter is not observed by the principal.

The paper is organized as follows. Section 2 presents the model. Section 3 formulates the incentive compatibility and limited liability constraints. Section 4 characterizes the optimal contract under maximal risk prevention. Based on this analysis, Section 5 studies the dynamics of firm size. Section 6 discusses the robustness of our results. Section 7 derives some empirical implications of our theoretical analysis. Section 8 concludes. Sketches of proofs are provided in the [Appendix](#). Complete proofs are available in the Supplemental Material ([Biais, Mariotti, Rochet, and Villeneuve \(2010\)](#)).

2. THE MODEL

There are two players: a principal and an agent. The agent can run a potentially profitable project for which she has unique necessary skills.¹³ However, this project entails costs, and the agent has limited liability and no initial cash. By contrast, the principal has unlimited liability and is able to cover the costs. Think of the agent as an entrepreneur or a manager running a business, and think of the principal as a financier, an insurance company, or society at large.

Time is continuous and the project can be operated over an infinite horizon. The two players are risk-neutral. The principal discounts the future at rate $r > 0$ and the agent discounts at rate $\rho > r$, which makes her more impatient than the principal. This introduces a wedge between the valuation of future transfers by the principal and the agent, and rules out indefinitely postponing payments to the latter. Without loss of generality, we normalize to 0 the setup cost of the project.

At any time t , the size X_t of the project can be scaled up or down. There are no constraints on downsizing: any fraction of the assets between 0 and 1 can be instantaneously liquidated. For simplicity, we normalize the maximal possible initial size of the project to 1 and assume that the liquidation value of the assets is 0. The project can also be expanded at unit cost $c \geq 0$. The rate at which such investments can take place is constrained, however. This reflects, for instance, that new plants cannot be built instantaneously or that the inflow of new skilled workers is constrained by search and training. Consistent with this, we assume that the instantaneous growth rate g_t of the project is at most equal to $\gamma \in (0, r)$. This is in line with the macroeconomic literature that emphasizes the delays and costs associated with investment, such as time-to-build constraints ([Kydland and Prescott \(1982\)](#)) or convex adjustment costs ([Hayashi \(1982\)](#)). Our formulation corresponds to a simple version of the adjustment cost model

¹³Empirically, this assumption is particularly relevant in the case of small businesses, where the entrepreneur-manager is often indispensable for operating the firm efficiently ([Sraer and Thesmar \(2007\)](#)).

in which there are no adjustment costs up to an instantaneous size adjustment $X_t \gamma dt$ and infinite adjustment costs beyond this point.

Operating profits per unit of time are equal to $X_t \mu$, where $\mu > 0$ is a constant that represents day-to-day size-adjusted operating profits. While such profits are constant, the project is subject to the risk of large losses. In the case of a manufacturing firm, such losses can be generated by a severe accident. In the case of a financial firm, they can result from a sudden and sharp decrease in the value of the assets in which the firm invested. The occurrence of these losses is modeled as a point process $N = \{N_t\}_{t \geq 0}$, where for each $t \geq 0$, N_t is the number of losses up to and including time t . Denote by $(T_k)_{k \geq 1}$ the successive random times at which these losses occur. A loss generates costs that are borne by the principal rather than by the agent. For example, an oil spill imposes huge damages on the environment and on the inhabitants of the affected region, but has limited direct impact on the manager of the oil company. Alternatively, in the case of financial firms, the losses incurred by many banks in 2007 and 2008 exceeded what they could cope with, and governments and taxpayers had to bear the costs. To capture this in our model, we assume that the agent has limited liability and cannot be held responsible for these losses in excess of her current wealth, so that it is the principal who has to incur the costs. We assume that, like operating profits, losses increase linearly with the size of the project. Thus, if there is a loss at time t , the corresponding cost is $X_t C$, where $C > 0$ is the size-adjusted cost. Overall, the net output flow generated by the project during the infinitesimal time interval $(t, t + dt]$ is $X_t(\mu dt - C dN_t)$.

By exerting effort, the agent affects the probability with which losses occur: a higher effort reduces the probability $\Lambda_t dt$ that a loss occurs during $(t, t + dt]$. For simplicity, we consider only two levels of effort, corresponding to $\Lambda_t = \lambda > 0$ and $\Lambda_t = \lambda + \Delta\lambda$, with $\Delta\lambda > 0$. To model the cost of effort, we adopt the same convention as Holmström and Tirole (1997): if the agent shirks at time t , that is, if $\Lambda_t = \lambda + \Delta\lambda$, she obtains a private benefit $X_t B$; by contrast, if the agent exerts effort at time t , that is, if $\Lambda_t = \lambda$, she obtains no private benefit. This formulation is similar to one in which the agent incurs a constant cost per unit of time and per unit of size of the project when exerting effort, and incurs no cost when shirking.

REMARK: It is natural to assume that operating profits and losses are increasing in the size of the project. It is also natural to assume that the opportunity cost of risk prevention is increasing in the size of the project: it takes more time, effort, and energy to check compliance and monitor safety processes in two plants than in a single plant, or in a large trading room with many traders than in a small one. Observe, however, that we require more than monotonicity, since we assume that operating profits, losses, and private benefits are linear in the size of the project. This constant returns to scale assumption is made for tractability. As shown in Section 4, it implies that the value function solution to the Hamilton–Jacobi–Bellman equation (23) is homogeneous of degree 1,

which considerably simplifies the characterization of the optimal contract. Yet, even without this assumption, some of the qualitative features of our analysis are upheld, as discussed in Section 6.

We assume throughout the paper that

$$(1) \quad \frac{\mu - \lambda C}{r} > c$$

and that

$$(2) \quad \Delta\lambda C > B.$$

The left-hand side of (1) is the present value of the net expected cash flow generated by one unit of capacity over an infinite horizon when the agent always exerts effort. The right-hand side of (1) is the cost of an additional unit of capacity. Condition (1) implies that the project has positive net present value and that investment is desirable when the agent always exerts effort. The left-hand side of (2) is the size-adjusted expected social cost of increased risk when the agent shirks. The right-hand side of (2) is the size-adjusted private benefit from shirking. Condition (2) implies that in the absence of moral hazard, it is socially optimal to require the agent to always exert effort. The first-best policy can therefore be characterized as follows: first, the project is initiated at its maximal capacity of 1 and then it grows at the maximal feasible rate γ with no downsizing ever taking place; second, a maximal risk prevention policy is implemented in which the agent always exerts effort.

From now on, we focus on the case where there is asymmetric information. Specifically, we assume that, unlike profits and losses, the agent's effort decisions are not observable by the principal. This leads to a moral hazard problem, whose key parameters are B and $\Delta\lambda$. The larger the size-adjusted private benefit B is, the more attractive it is for the agent to shirk. The lower $\Delta\lambda$ is, the more difficult it is to detect shirking. The contract between the principal and the agent is designed and agreed upon at time 0. The agent reacts to this contract by choosing an effort process $\Lambda = \{\Lambda_t\}_{t \geq 0}$. We assume that the players can fully commit to a long-term contract.

REMARK: We thus abstract throughout from imperfect commitment problems and focus on a single source of market imperfection: moral hazard in risk prevention. This assumption is standard in the dynamic moral hazard literature; see, for instance, Rogerson (1985), Spear and Srivastava (1987), or Phelan and Townsend (1991). More precisely, our analysis is in line with Clementi and Hoppenhayn (2006), DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin, and Rochet (2007), DeMarzo and Fishman (2007a, 2007b), or Sannikov (2008), where limited liability reduces the ability to punish the agent. This compels the

principal to replace such punishments by actions, such as downsizing or liquidation, that are ex post inefficient.¹⁴ When the principal is more patient than the agent and there is no investment, as in DeMarzo and Sannikov (2006) and Biais, Mariotti, Plantin, and Rochet (2007), this leads to the result that the firm eventually ceases to exist. By contrast, in the present model, this negative trend can be outweighed by investment.

A contract specifies downsizing, investment, and liquidation decisions, as well as payments to the agent, as functions of the history of past losses. The size process $X = \{X_t\}_{t \geq 0}$ is nonnegative, with initial condition $X_0 \leq 1$. The size of the project can be decomposed as

$$(3) \quad X_t = X_0 + X_t^d + X_t^i$$

for all $t \geq 0$, where $X^d = \{X_t^d\}_{t \geq 0}$, the cumulative downsizing process, is decreasing, and $X^i = \{X_t^i\}_{t \geq 0}$, the cumulative investment process, is increasing. Our assumptions imply that X^i is absolutely continuous with respect to time; that is,

$$(4) \quad X_t^i = \int_0^t X_s g_s ds,$$

where the instantaneous growth rate of the project satisfies

$$(5) \quad 0 \leq g_t \leq \gamma$$

for all $t \geq 0$. Because of limited liability, the process $L = \{L_t\}_{t \geq 0}$ which describes the cumulative transfers to the agent, is nonnegative and increasing. The time at which liquidation occurs is denoted by τ . We allow τ to be infinite, and we let $X_t = 0$ and $L_t = L_\tau$ for all $t > \tau$.

At any time t prior to liquidation, the sequence of events during the infinitesimal time interval $[t, t + dt]$ can heuristically be described as follows:

Step 1. The size X_t of the project is determined, that is, there is downsizing or investment or the size remains constant.

Step 2. The agent takes her effort decision Λ_t .

Step 3. With probability $\Lambda_t dt$, there is a loss, in which case $dN_t = 1$; otherwise $dN_t = 0$.

Step 4. The agent receives a nonnegative transfer dL_t .

Step 5. The project is either liquidated or continued.

According to this timing, the downsizing and effort decisions are taken before knowing the current realization of the loss process. Formally, the processes X and Λ are \mathcal{F}^N -predictable, where $\mathcal{F}^N = \{\mathcal{F}_t^N\}_{t \geq 0}$ is the filtration generated by N .

¹⁴For a discussion of renegotiation in this context, see Quadrini (2004), DeMarzo and Sannikov (2006, Section IV.B), or DeMarzo and Fishman (2007a, Appendix B2, 2007b, Section 2.9).

By contrast, payment and liquidation decisions at any time are taken after observing whether there was a loss at this time. Hence L is \mathcal{F}^N -adapted and τ is an \mathcal{F}^N -stopping time.¹⁵ An effort process Λ generates a unique probability distribution \mathbf{P}^Λ over the paths of the process N . Denote by \mathbf{E}^Λ the corresponding expectation operator.

Given a contract $\Gamma = (X, L, \tau)$ and an effort process Λ , the expected discounted utility of the agent is

$$(6) \quad \mathbf{E}^\Lambda \left[\int_0^\tau e^{-\rho t} (dL_t + 1_{\{\Lambda_t=\lambda+\Delta\lambda\}} X_t B dt) \right],$$

while the expected discounted profit of the principal is¹⁶

$$(7) \quad \mathbf{E}^\Lambda \left[\int_0^\tau e^{-rt} \{X_t[(\mu - g_t c) dt - CdN_t] - dL_t\} \right].$$

An effort process Λ is incentive compatible with respect to a contract Γ if it maximizes the agent's expected utility (6) given Γ . The problem of the principal is to find a contract Γ and an incentive compatible effort process Λ that maximize expected discounted profit (7), subject to delivering to the agent a required expected discounted utility level. It is without loss of generality to focus on contracts Γ such that the present value of the payments to the agent is finite, that is,

$$(8) \quad \mathbf{E}^\Lambda \left[\int_0^\tau e^{-\rho t} dL_t \right] < \infty.$$

Indeed, by inspection of (7), if the present value of the payments to the agent were infinite, the fact that $\rho > r$ would imply infinitely negative expected discounted profits for the principal. The latter would be better off proposing no contract altogether.

3. INCENTIVE COMPATIBILITY AND LIMITED LIABILITY

To characterize incentive compatibility, we rely on martingale techniques similar to those introduced by Sannikov (2008). When taking her effort decision at a time t , the agent considers how it will affect her continuation utility, defined as

$$(9) \quad W_t(\Gamma, \Lambda) = \mathbf{E}^\Lambda \left[\int_t^\tau e^{-\rho(s-t)} (dL_s + 1_{\{\Lambda_s=\lambda+\Delta\lambda\}} X_s B ds) \middle| \mathcal{F}_t^N \right] 1_{\{\tau > t\}}.$$

¹⁵See, for instance, Dellacherie and Meyer (1978, Chapter IV, Definitions 12, 49, and 61) for definitions of these concepts.

¹⁶All integrals are of the Lebesgue–Stieltjes kind. For each s and t , we write \int_s^t for $\int_{[s,t]}$ and $\int_s^{t^-}$ for $\int_{[s,t)}$.

Denote by $W(\Gamma, \Lambda) = \{W_t(\Gamma, \Lambda)\}_{t \geq 0}$ the agent's continuation utility process. Note that, by construction, $W(\Gamma, \Lambda)$ is \mathcal{F}^N -adapted. In particular, $W_t(\Gamma, \Lambda)$ reflects whether there was a loss at time t . To characterize how the agent's continuation utility evolves over time, it is useful to consider her lifetime expected utility, evaluated conditionally on the information available at time t , that is,¹⁷

$$(10) \quad U_t(\Gamma, \Lambda) = \mathbf{E}^A \left[\int_0^\tau e^{-\rho s} (dL_s + 1_{\{\Lambda_s=\lambda+\Delta\lambda\}} X_s B ds) \middle| \mathcal{F}_t^N \right]$$

$$= \int_0^{t \wedge \tau^-} e^{-\rho s} (dL_s + 1_{\{\Lambda_s=\lambda+\Delta\lambda\}} X_s B ds) + e^{-\rho t} W_t(\Gamma, \Lambda).$$

Since $U_t(\Gamma, \Lambda)$ is the expectation of a given random variable conditional on \mathcal{F}_t^N , the process $U(\Gamma, \Lambda) = \{U_t(\Gamma, \Lambda)\}_{t \geq 0}$ is an \mathcal{F}^N -martingale under the probability measure \mathbf{P}^A . Its last element is $U_\tau(\Gamma, \Lambda)$, which is integrable by (8).

Relying on this martingale property, we now offer an alternative representation of $U(\Gamma, \Lambda)$. Consider the process $M^A = \{M_t^A\}_{t \geq 0}$ defined by

$$(11) \quad M_t^A = N_t - \int_0^t \Lambda_s ds$$

for all $t \geq 0$. Equation (11) is best understood when Λ is a constant process. In that case, M_t^A is simply the number of losses up to and including time t , minus its expectation. More generally, a basic result from the theory of point processes is that M^A is an \mathcal{F}^N -martingale under \mathbf{P}^A . Changes in the effort process Λ induce changes in the distribution of losses, which essentially amount to Girsanov transformations of the process N . The martingale representation theorem for point processes then implies the following lemma.¹⁸

LEMMA 1: *The martingale $U(\Gamma, \Lambda)$ satisfies*

$$(12) \quad U_t(\Gamma, \Lambda) = U_0(\Gamma, \Lambda) - \int_0^{t \wedge \tau} e^{-\rho s} H_s(\Gamma, \Lambda) dM_s^A$$

for all $t \geq 0$, \mathbf{P}^A -almost surely, for some \mathcal{F}^N -predictable process $H(\Gamma, \Lambda) = \{H_t(\Gamma, \Lambda)\}_{t \geq 0}$.

Along with (11), (12) implies that the lifetime expected utility of the agent evolves in response to the jumps of the process N . At any time t , the change in $U_t(\Gamma, \Lambda)$ is equal to the product between a \mathcal{F}^N -predictable function of the

¹⁷For each x and y , we denote by $x \wedge y$ the minimum of x and y , and denote by $x \vee y$ the maximum of x and y .

¹⁸See, for instance, Brémaud (1981, Chapter III, Theorems T9 and T17, and Chapter VI, Theorems T2 and T3) for the relevant results.

past, namely $e^{-\rho t}H_t(\Gamma, \Lambda)$, and a term $-dM_t^\Lambda$ that reflects the events occurring at time t . This term is in turn equal to the difference between the instantaneous probability $\Lambda_t dt$ of a loss and the instantaneous change dN_t in the total number of losses, which is equal to 0 or 1. Equations (10) and (12) imply that the continuation utility of the agent evolves as

$$(13) \quad dW_t(\Gamma, \Lambda) = [\rho W_t(\Gamma, \Lambda) - 1_{\{\Lambda_t=\lambda+\Delta\lambda\}} X_t B] dt + H_t(\Gamma, \Lambda)(\Lambda_t dt - dN_t) - dL_t$$

for all $t \in [0, \tau]$. Equation (13) states that, net of private benefits and wages, the expected instantaneous change in the continuation utility of the agent is equal to her discount rate ρ , while $H(\Gamma, \Lambda)$ is the sensitivity to losses of this utility. Building on this analysis and letting $b = B/\Delta\lambda$, we obtain the following result, in line with Sannikov (2008, Proposition 2).

PROPOSITION 1: *A necessary and sufficient condition for the effort process Λ to be incentive compatible given the contract $\Gamma = (X, L, \tau)$ is that*

$$(14) \quad \Lambda_t = \lambda \quad \text{if and only if} \quad H_t(\Gamma, \Lambda) \geq X_t b$$

for all $t \in [0, \tau]$, \mathbf{P}^Λ -almost surely.

It follows from (13) that if there is a loss at some time $t \in [0, \tau]$, the agent's continuation utility must be instantaneously reduced by an amount $H_t(\Gamma, \Lambda)$.¹⁹ Proposition 1 states that to induce the agent to choose a high level of risk prevention, this reduction in her continuation utility must be at least as large as $X_t b$. This is because $X_t b$ reflects the attractiveness of the private benefits obtained by the agent when shirking. To reason in size-adjusted terms, let $h_t = H_t/X_t$. The incentive compatibility condition (14) under which $\Lambda_t = \lambda$ then is rewritten as

$$(15) \quad h_t \geq b.$$

It is convenient to introduce the notation $W_{t^-}(\Gamma, \Lambda) = \lim_{s \uparrow t} W_s(\Gamma, \Lambda)$ to denote the left-hand limit of the process $W(\Gamma, \Lambda)$ at $t > 0$. While $W_t(\Gamma, \Lambda)$ is the continuation utility of the agent at time t after observing whether there was a loss at time t , $W_{t^-}(\Gamma, \Lambda)$ is the continuation utility of the agent evaluated before such knowledge is obtained.²⁰ Observe that while the process $W(\Gamma, \Lambda)$ is

¹⁹In full generality, one should also allow for jumps in the transfer process. For incentive reasons, it is, however, never optimal to pay the agent when a loss occurs. Moreover, it will turn out that the optimal transfer process is absolutely continuous, so that payments do not come in lump sums. To ease the exposition, we, therefore, rule out jumps in the transfer process in the body of the paper. The possibility of such jumps is explicitly taken into account below in the verification theorem.

²⁰ $W_{0^-}(\Gamma, \Lambda)$ is defined by (6).

\mathcal{F}^N -adapted, the process $W_{t-}(\Gamma, \Lambda) = \{W_{t-}(\Gamma, \Lambda)\}_{t \geq 0}$ is \mathcal{F}^N -predictable. Combining the fact that the continuation utility of the agent must remain nonnegative according to the limited liability constraint with the fact that it must be reduced by an amount $H_t(\Gamma, \Lambda)$ if there is a loss at time t according to (13), one must have

$$(16) \quad W_{t-}(\Gamma, \Lambda) \geq H_t(\Gamma, \Lambda)$$

for all $t \in [0, \tau]$. To simplify notation, we drop the arguments Γ and Λ in the remainder of the paper.

4. OPTIMAL CONTRACTING WITH MAXIMAL RISK PREVENTION

While in the previous section we considered general effort processes, in the present section we characterize the optimal contract that induces maximal risk prevention, that is, $\Lambda_t = \lambda$ for all $t \in [0, \tau]$. This is in line with most of the literature on the principal–agent model, which offers more precise insights into how to implement given courses of actions at minimal cost than into which course of actions is, all things considered, optimal for the principal.²¹ In Section 6.1, however, we provide sufficient conditions under which it is optimal for the principal to request maximal risk prevention from the agent. The optimal contract that we derive in this section can be described with the help of two state variables: the size of the project, resulting from past downsizing and investment decisions, and the continuation utility of the agent, reflecting future payment decisions. To build intuition, we first provide a heuristic derivation of the principal’s value function and of the main features of the optimal contract. Next, we verify that this candidate value function is indeed optimal and we fully characterize the optimal contract.

4.1. A Heuristic Derivation

In this heuristic derivation, we suppose that transfers are absolutely continuous with respect to time and that no payment is made after a loss, that is

$$(17) \quad dL_t = X_t \ell_t 1_{\{dN_t=0\}} dt,$$

where

$$(18) \quad \ell_t \geq 0$$

for all $t \geq 0$. Here $\{\ell_t\}_{t \geq 0}$ is assumed to be an \mathcal{F}^N -predictable process that represents the size-adjusted transfer flow to the agent. We will later verify that this

²¹See, for instance, Laffont and Martimort (2002, Chapters 4 and 8) for a recent overview of that literature.

conjecture is correct at the optimal contract. Now consider project size. Downsizing is suboptimal in the first-best case and, as we later verify, it remains so in the second-best case as long as no losses occur. After losses, however, downsizing may prove necessary in the second-best case. This reflects that, for incentive purposes, it is necessary to reduce the agent's continuation utility after each loss by an amount that is proportional to her private benefits from shirking; the latter, in turn, are proportional to the size of the project. When the continuation utility of the agent is low, the incentive compatibility constraint is compatible with the limited liability constraint only if the size of the project is itself low enough.

To see this more precisely, suppose that at the outset of time t , the size of the project is X_t and the continuation utility of the agent is W_{t-} . If there is a loss at time t , the agent's continuation utility must be reduced from W_{t-} to $W_t = W_{t-} - X_t h_t$. At this point, the question arises whether this loss implies that the project should be downsized. Denote by $X_{t^+} = \lim_{s \downarrow t} X_s \in [0, X_t]$ the size of the project just after time t . Since effort is still required from the agent, Proposition 1 implies that if there were a second loss arbitrarily close to the first, the continuation utility of the agent would have to be reduced further by at least $X_{t^+} b$. This would be consistent with limited liability only if $W_{t-} - X_t h_t \geq X_{t^+} b$, or, equivalently, letting $w_t = W_{t-}/X_t$ and $x_t = X_{t^+}/X_t$ if

$$(19) \quad \frac{w_t - h_t}{b} \geq x_t.$$

Hence, downsizing is necessary after the first loss, that is, $x_t < 1$, whenever the initial size-adjusted continuation utility w_t of the agent is so low that $(w_t - h_t)/b < 1$.

We are now ready to characterize the evolution of the continuation value $F(X_t, W_{t-})$ of the principal. Since the principal discounts the future at rate r , his expected flow of value at time t is given by

$$(20) \quad rF(X_t, W_{t-}).$$

This must be equal to the sum of the expected instantaneous cash flows and of the expected rate of change in his continuation value. The former are equal to the expected net cash flow from the project minus the cost of investment and the expected payment to the agent. By (4) and (17), this yields

$$(21) \quad X_t [\mu - \lambda C - g_t c - \ell_t (1 - \lambda dt)].$$

To evaluate the expected rate of change in the principal's continuation value, we use the dynamics (3) of the project's size along with the dynamics of the agent's continuation utility, setting $\Lambda_t = \lambda$ in (13). Applying the change of variable formula for processes of bounded variation, which is the counterpart of

Itô's formula for these processes, yields²²

$$(22) \quad [\rho W_{t-} + X_t(\lambda h_t - \ell_t)]F_W(X_t, W_{t-}) + X_t g_t F_X(X_t, W_{t-}) \\ - \lambda[F(X_t, W_{t-}) - F(X_t x_t, W_{t-} - X_t h_t)].$$

The first term arises because of the drift of W_{t-} , the second term corresponds to investment, and the third term reflects the possibility of jumps in the project's size and in the agent's continuation utility due to losses. Adding (22) to (21), identifying to (20), and letting dt go to 0, we obtain that the value function of the principal satisfies the Hamilton–Jacobi–Bellman equation

$$(23) \quad rF(X_t, W_{t-}) = X_t(\mu - \lambda C) \\ + \max\{-X_t \ell_t + [\rho W_{t-} + X_t(\lambda h_t - \ell_t)]F_W(X_t, W_{t-}) \\ + X_t g_t [F_X(X_t, W_{t-}) - c] \\ - \lambda[F(X_t, W_{t-}) - F(X_t x_t, W_{t-} - X_t h_t)]\},$$

where the maximization is over the set of controls (g_t, h_t, ℓ_t, x_t) that satisfy constraints (5), (15), (18), and (19).

To get more insight into the structure of the solution, we impose further restrictions on the function F that we later check to be satisfied at the optimal contract. First, because of constant returns to scale, it is natural to require F to be homogenous of degree 1:

$$F(X, W) = XF\left(1, \frac{W}{X}\right) = Xf\left(\frac{W}{X}\right)$$

for all $(X, W) \in \mathbb{R}_{++} \times \mathbb{R}_+$. Intuitively, f maps the size-adjusted continuation utility w_t of the agent into the size-adjusted continuation value of the principal. Second, we require f to be globally concave. This property, which will be formally established in the verification theorem below, has the following economic interpretation. As argued above, while downsizing is inefficient in the first-best case, it is necessary in the second-best case to provide incentives to the agent when w_t is low. When this is the circumstance, the principal's value reacts strongly to bad performance because the latter significantly raises the risk of costly downsizing. By contrast, when w_t is large, bad performance has a more limited impact on downsizing risk. This greater sensitivity to shocks when w_t is low than when it is large is reflected in the concavity of the size-adjusted value function f . Finally, we set

$$f(w) = \frac{f(b)}{b}w$$

²²See, for instance, Dellacherie and Meyer (1982, Chapter VI, Section 92).

for all $w \in [0, b]$. This is just by convention and is done to simplify the notation, since, by (14) and (16), w_t never enters the interval $[0, b)$.

We can now derive several properties of the optimal controls in the Hamilton–Jacobi–Bellman equation. Optimizing with respect to ℓ_t and using the homogeneity of F yields

$$(24) \quad f'(w_t) = F_W(X_t, W_{t-}) \geq -1,$$

with equality only if $\ell_t > 0$. Intuitively, the left-hand side of (24) is the increase in the principal's continuation value due to a marginal increase in the agent's continuation utility, while the right-hand side is the marginal cost to the principal of making an immediate payment to the agent. It is optimal to delay payments as long as they are more costly than utility promises, that is, as long as the inequality in (24) is strict. The concavity of f implies that this is the case when w_t is below a given threshold, which we denote by w^p . The optimal contract thus satisfies the following property.

PROPERTY 1: Payments to the agent are made only if her size-adjusted continuation utility is at least w^p . The payment threshold w^p satisfies

$$(25) \quad f'(w^p) = -1.$$

In the first-best case, all the payments to the agent would be made at time 0, as she is more impatient than the principal. By contrast, in the second-best case, payments must be delayed and made contingent on a long enough record of good performance, so as to provide incentives to the agent. Since f is concave, it follows from (24) and (25) that $f'(w) = -1$ for all $w \geq w^p$. If one were to start from that region, the optimal contract would entail the immediate payment of a lump sum $w - w^p$ to the agent, counterbalanced by a drop in her size-adjusted continuation utility to w^p .

Suppose that w_t is below the threshold w^p , implying that $\ell_t = 0$. Then, using the homogeneity of F , (23) can be rewritten as

$$(26) \quad \begin{aligned} rf(w_t) &= \mu - \lambda C \\ &+ \max \left\{ (\rho w_t + \lambda h_t) f'(w_t) + g_t[f(w_t) - w_t f'(w_t) - c] \right. \\ &\quad \left. - \lambda \left[f(w_t) - x_t f\left(\frac{w_t - h_t}{x_t}\right) \right] \right\}. \end{aligned}$$

Since f is concave and vanishes at 0, the mapping $x_t \mapsto x_t f((w_t - h_t)/x_t)$ is increasing. It is thus optimal to let x_t be as high as possible in (26), reflecting that downsizing is costly since the project is profitable. Using (19) along with the fact that $x_t \leq 1$ then leads to the second property of the optimal contract.

PROPERTY 2: If there is a loss at time t , the optimal downsizing policy is

$$(27) \quad x_t = \frac{w_t - h_t}{b} \wedge 1.$$

This property of the optimal contract reflects that, for a given level of incentives as measured by h_t , downsizing is imposed only as the last resort. Using our convention that f is linear over $[0, b]$, (27) can be substituted into (26) to obtain

$$(28) \quad \begin{aligned} rf(w_t) &= \mu - \lambda C \\ &+ \max\{(\rho w_t + \lambda h_t)f'(w_t) + g_t[f(w_t) - w_t f'(w_t) - c] \\ &- \lambda[f(w_t) - f(w_t - h_t)]\}. \end{aligned}$$

The concavity of f implies that it is optimal to let h_t be as low as possible in (28), which according to the incentive compatibility condition (15) leads to the third property of the optimal contract.

PROPERTY 3: The sensitivity to losses of the agent's continuation utility is given by

$$(29) \quad h_t = b.$$

Intuitively, (29) reflects that because the principal's continuation value is concave in the agent's continuation utility, it is optimal to reduce the agent's exposure to risk by letting h_t equal the minimal value consistent with her exerting effort. In particular, downsizing takes place following a loss at date t if and only if $w_t < 2b$, that is, if and only if it is absolutely necessary so as to maintain the consistency between the incentive compatibility constraint and the limited liability constraint.

Finally turn to investment decisions. Note that the size-adjusted social value of the project, $f(w) + w$, is increasing in w until w^p and is flat afterward. A necessary and sufficient condition for investment to ever be strictly profitable is that the maximal size-adjusted social value of the project be larger than the unit cost of investment:

$$(30) \quad f(w^p) + w^p > c.$$

If (30) did not hold, the value of investment would be lower than its cost, so that it would be suboptimal to invest.²³ Thus, as will be checked below in the

²³If (30) held as an equality, whether or not investment take place would be indifferent from a social viewpoint. Since, fixing the other parameters of the model, this can only occur for a single value of c , we ignore that possibility in the remainder of the paper.

verification theorem, there is some investment in the optimal contract only if c is not too high. Optimizing in (28) with respect to g_t under constraint (5), we obtain that $g_t = \gamma$ if

$$(31) \quad f(w_t) - w_t f'(w_t) > c,$$

and $g_t = 0$ otherwise. The left-hand side of (31) is the marginal benefit of an additional capacity unit, while the right-hand side is the unit cost of investment. Scale expansion is optimal when the former is greater than the latter. In that case, because of the linearity in the technology, size grows at the maximal feasible rate γ . The concavity of f implies that (31) holds when w_t is above a given threshold, which we denote by w^i . The optimal contract thus satisfies the following property.

PROPERTY 4: Investment takes place, at rate γ , if and only if the size-adjusted continuation utility of the agent is above w^i . The investment threshold w^i satisfies

$$(32) \quad w^i = \inf\{w > b \mid f(w) - wf'(w) > c\}.$$

In the first-best case, because of condition (1), investment always takes place at the maximal rate γ . By contrast, in the second-best case, if c is not too low, this is the case only if a long enough record of good performance has been accumulated. This is because increasing the size of the project raises the private benefits from shirking and thus worsens the moral hazard problem. This jeopardizes incentives, except if the agent has enough at stake to still prefer high effort, that is, only if w_t is large enough. An important alternative scenario arises whenever c is low enough. In that case, inequality (31) is satisfied for all $w_t > b$, so that $w^i = b$ and it is always optimal to invest, even in the second-best case. Formally, this is reflected in the fact that the function f is not differentiable at b , with $f'_-(b) = f(b)/b > f'_+(b)$, so that $f(b) - bf'_+(b) > c$ for c close enough to 0.

The dynamics of the principal's size-adjusted continuation value depends on whether there is investment. In the no investment region $(b, w^i]$,

$$(33) \quad rf = \mu - \lambda C + \mathcal{L}f,$$

where the delay differential operator \mathcal{L} is defined by

$$(34) \quad \mathcal{L}f(w) = (\rho w + \lambda b)f'(w) - \lambda[f(w) - f(w - b)].$$

In the investment region $(w^i, w^p]$,

$$(35) \quad (r - \gamma)f = \mu - \lambda C - \gamma c + \mathcal{L}_\gamma f,$$

where the delay differential operator \mathcal{L}_γ is defined by

$$(36) \quad \mathcal{L}_\gamma f(w) = [(\rho - \gamma)w + \lambda b]f'(w) - \lambda[f(w) - f(w - b)].$$

Comparing equations (35) and (36) to equations (33) and (34) reveals that, in addition to the decrease γc in the size-adjusted cash flow, the impact of investment at rate γ is comparable to that of a decrease γ in both the principal's and the agent's discount rates. Intuitively, this reflects that investment makes delaying payments less costly, because the total size of the pie grows while the players wait. Thus, although incentive considerations imply that both investment and payments should be delayed relative to the first-best case, investment takes place before payments do, as stated now.

PROPERTY 5: If investment is strictly profitable, the investment threshold w^i is strictly lower than the payment threshold w^p .

This follows from evaluating (31) at w^p , which yields $f(w^p) - w^p f'(w^p) > c$ because of (25) and (30). While investment takes place in a region where the size-adjusted social value of the project is strictly increasing, payments are made to the agent when the size-adjusted social value of the project reaches its maximum, so that it is inefficient to delay payments any longer. At the payment threshold w^p , transfers are constructed in such a way that the agent's continuation utility stays constant until there is a loss. That is, they are set to the highest level consistent with the size-adjusted social value remaining at its maximum. This level can be computed as follows. Setting $\Lambda_t = \lambda$ in (13) and making use of (17) and Property 3, we obtain

$$(37) \quad dW_t = (\rho W_t + X_t \lambda b) dt - X_t b dN_t - X_t \ell_t 1_{\{dN_t=0\}} dt.$$

Suppose now that $w_t = w^p$, so that the size-adjusted social value of the project is at its maximum and that $dN_t = 0$, so that there is no loss at time t . Then $W_t = X_t w^p$ and $dX_t = X_t \gamma dt$. Substituting in (37), we obtain the following property.

PROPERTY 6: If there is no loss at time t , the size-adjusted transfer flow is

$$(38) \quad \ell_t = [(\rho - \gamma)w^p + \lambda b]1_{\{w_t=w^p\}}.$$

According to (38), when payments are made at the payment threshold w^p , they come in a steady flow in size-adjusted terms until a loss occurs.

The above conjectures about the structure of the optimal contract are illustrated on Figure 1. Because of constant returns to scale, there are four regimes in the (X_t, W_{t-}) plane separated by straight lines, reflecting that downsizing, investment, or transfers take place, depending on the position of the agent's size-adjusted continuation utility relative to the thresholds b , w^i , and w^p . Because $b \leq w_t \leq w^p$ for all $t > 0$, (X_t, W_{t-}) stays away from the interiors of the downsizing and transfer regions after time 0.

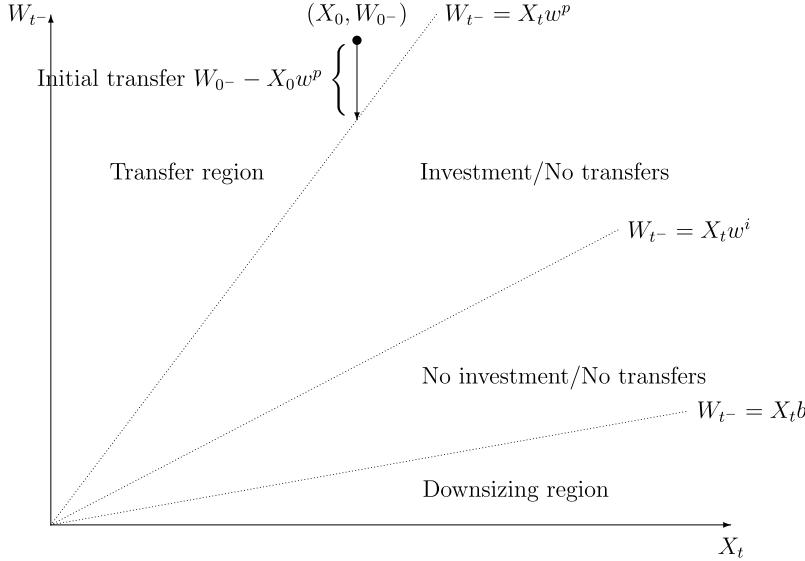


FIGURE 1.—The four regions that characterize the optimal contract. A situation in which the contract is initiated at a point (X_0, W_{0-}) that lies in the interior of the transfer region is represented. When transfers take place later on, the state variables move along the straight line $W_{t-} = X_t w^p$.

4.2. The Verification Theorem

We now show that the above heuristic characterization does correspond to the optimal contract. To do this, we first show that there exists a size-adjusted value function f such that Properties 1–6 hold.

PROPOSITION 2: *Suppose that*

$$(39) \quad \mu - \lambda C > (\rho - r)b \left(2 + \frac{r}{\lambda} \right).$$

Then there exists a constant $\bar{c} > 0$ such that if

$$(40) \quad c < \bar{c},$$

the delay differential equation

$$(41) \quad \begin{aligned} f(w) &= \frac{f(b)}{b}w, \quad \text{if } w \in [0, b], \\ rf(w) &= \mu - \lambda C + \mathcal{L}f(w), \quad \text{if } w \in (b, w^i], \\ (r - \gamma)f(w) &= \mu - \lambda C - \gamma c + \mathcal{L}_\gamma f(w), \quad \text{if } w \in (w^i, w^p], \\ f(w) &= f(w^p) + w^p - w, \quad \text{if } w \in (w^p, \infty) \end{aligned}$$

has a maximal solution f , where the thresholds w^p and w^i are endogenously determined by (25) and (32), with $w^p > w^i$, and the operators \mathcal{L} and \mathcal{L}_γ are defined as in (34) and (36). The function f is globally concave and continuously differentiable except at b .

If condition (39) did not hold, the solution would be degenerate, with downsizing taking place after each loss. This would arise because the private benefits from shirking would be large relative to the expected cash flow from the project, making the agency problem very severe. Condition (40) ensures that the investment cost is low enough so that there are circumstances in which it is strictly optimal to increase the size of the project. If the investment cost c were strictly larger than \bar{c} , the optimal contract would be similar to that described above, except that the investment region would be empty. The threshold value \bar{c} corresponds to the maximum of the size-adjusted social value of the project that arises in this no investment situation.

The next steps of the analysis are to show that the function constructed in Proposition 2 yields the maximal value that can be obtained by the principal and to explicitly construct the optimal contract. To do so, fix an initial project size X_0 and an initial expected utility W_{0-} for the agent, and consider the processes $\{w_t\}_{t \geq 0}$ and $\{l_t\}_{t \geq 0}$ to be solutions to

$$(42) \quad w_t = w_0 + \int_0^{t^-} \left\{ [(\rho - \gamma 1_{\{w_s > w^i\}})w_s + \lambda b] ds \right.$$

$$\left. - b \left(\frac{w_s - b}{b} \wedge 1 \right) dN_s - dl_s \right\},$$

$$(43) \quad l_t = (w_0 - w^p) \vee 0 + \int_0^t [(\rho - \gamma)w^p + \lambda b] 1_{\{w_{s+} = w^p\}} ds$$

for all $t \geq 0$, where $w_0 = W_{0-}/X_0$, and w^i and w^p are defined as in Proposition 2. For the moment, we simply take these processes as given. Yet, consistent with the heuristic derivation of Section 4.1, it eventually turns out in equilibrium that, at any time t , w_t is the initial size-adjusted continuation utility of the agent, while l_t represents cumulative size-adjusted transfers up to and including time t . The following proposition is central to our results.

PROPOSITION 3: *Under conditions (39) and (40), the optimal contract $\Gamma = (X, L, \tau)$ that induces maximal risk prevention and delivers the agent an initial expected discounted utility W_{0-} given initial firm size X_0 is as follows:*

(i) *The project is downsized by a factor $[(w_{T_k} - b)/b] \wedge 1$ at any time T_k at which there is a loss. Moreover, the size of the project grows at rate γ as long as $w_t > w^i$, and grows at rate 0 otherwise. As a result, the size of the project is*

$$(44) \quad X_t = X_0 \prod_{k=1}^{N_t^-} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) \exp \left(\int_0^t \gamma 1_{\{w_s > w^i\}} ds \right)$$

*at any time $t \geq 0$.*²⁴

(ii) *The flow of transfers to the agent is $X_t[(\rho - \gamma)w^p + \lambda b]$ as long as $w_t = w^p$ and no loss occurs. As a result, the cumulative transfers to the agent are*

$$(45) \quad L_t = X_0 l_0 + \int_0^t X_s dl_s$$

*at any time $t \geq 0$.*²⁵

(iii) *Liquidation occurs with probability 0 on the equilibrium path, that is,*

$$(46) \quad \tau = \infty,$$

P-almost surely.

The value to the principal of this contract is $F(X_0, W_{0-}) = X_0 f(W_{0-}/X_0)$, with f constructed as in Proposition 2.

As shown in the proof of Proposition 3, the optimal contract entails at any time t a continuation utility $W_t = \lim_{s \downarrow t} X_s w_s$ for the agent. The process W obtained in this way satisfies (13) with $\Lambda_t = \lambda$ and $H_t = X_t b$, and thus induces maximal risk prevention. As conjectured in Section 4.1, the optimal contract involves two state variables: the size of the project, X_t , and the size-adjusted continuation utility of the agent, w_t , or, equivalently, her beginning-of-period continuation utility $W_{t-} = X_t w_t$.

The main features of the optimal contract are also in line with the heuristic derivation of Properties 1–6. First consider transfers, as given by (43) and (45). If $w_0 > w^p$, an initial lump-sum is immediately distributed to the agent. Then, at time $t > 0$, transfers take place if and only if $w_t = w^p$ and there is no loss, and they are constructed in such a way that the agent's size-adjusted continuation utility stays constant until a loss occurs. This is consistent with Properties 1 and 6.

Next consider the size of the project, as given by (44). The first term on the right-hand side of (44) is the initial size of the firm. The second term on the right-hand side of (44) reflects downsizing, which takes place only after losses occur at the random times T_k and when $(w_{T_k} - b)/b < 1$. This is consistent with

²⁴By convention, $\prod_{\emptyset} = 1$.

²⁵Observe from (42) and (43) that $w_{t+} = w^p$ if and only if $w_t = w^p$ and there is no loss at time t .

Properties 2 and 3. The third term on the right-hand side of (44) reflects that investment takes place, at rate γ , if and only if $w_t > w^i$. This is consistent with Properties 4 and 5.

Finally consider the size-adjusted continuation utility of the agent, as given by (42). Its dynamics is somewhat complicated, as it reflects the joint effect of direct changes in the agent's continuation utility and indirect changes due to the variations in the project's size. It follows from (42) that if a loss occurs at a time T_k such that $w_{T_k} \geq 2b$, no downsizing takes place and the size-adjusted continuation utility of the agent drops by an amount b . This is consistent with Property 3. By contrast, if a loss occurs at a time T_k such that $b \leq w_{T_k} < 2b$, the project is downsized by a factor $(w_{T_k} - b)/b$, and the size-adjusted continuation utility of the agent drops by an amount $w_{T_k} - b$. Thus, in any case, the sensitivity to losses of the agent's size-adjusted continuation utility is $(w_{T_k} - b) \wedge b$.

It should be emphasized that liquidation plays virtually no role in the optimal incentive contract, as reflected by (46). Indeed, as can be seen from (42), $w_t = W_{t-}/X_t$ always remains strictly greater than b . As a result, W_t , which is in the worst case equal to $W_{t-} - X_t b$ if there is a loss at time t , always remains strictly positive.²⁶ This is in sharp contrast to the Brownian models studied by DeMarzo and Sannikov (2006), Biais, Mariotti, Plantin, and Rochet (2007), and Sannikov (2008), in which the optimal contract relies crucially on liquidation and involves no downsizing. Admittedly, even in the context of our Poisson model, an alternative way to provide incentives to the agent in the event of bad performance would be to threaten her to randomly liquidating the project, as is customary in discrete-time models (see, for instance, Clementi and Hopenhayn (2006) or DeMarzo and Fishman (2007b)). But in contrast to what happens in Brownian models, liquidation would then necessarily have to be both *stochastic* (as it would depend on the realization of a lottery at each potential liquidation time) and *unpredictable* (as it would take place only after a loss). When modeled in this way, liquidation allows the principal to achieve the same value as under downsizing. This would, however, be less tractable analytically and less conducive to a realistic implementation of the optimal contract. In addition, and more importantly, allowing for downsizing gives rise to a richer dynamics for the size of the project, which can increase but also decrease over time following good or bad performance.

Proposition 3 describes the optimal contract for a given initial project size X_0 and a given initial expected discounted utility W_{0-} for the agent. In Biais, Mariotti, Rochet, and Villeneuve (2010, Section D.3), we examine how these are determined at time 0 whenever the principal is competitive. That is, we look for a pair (X_0, W_{0-}) that maximizes utilitarian welfare under the constraint that the principal breaks even on average. As soon as f takes strictly positive values, it is optimal to start operating the project at full scale, $X_0 = 1$.²⁷ When the

²⁶Exceptions arise only with probability 0; for instance, if $W_{0-} = X_0 b$ and there is a loss at time 0.

²⁷Otherwise it is optimal to let $X_0 = W_{0-} = 0$ and not to operate the project.

participation constraint of the principal is slack, the contract is initiated at the payment threshold $w_0 = w^p$, so that the agent is immediately compensated. By contrast, when the participation constraint of the principal binds, it is necessary to initiate the contract at a lower level $w_0 < w^p$, so that it is optimal to wait before compensating the agent.

5. FIRM SIZE DYNAMICS

In this section, we build on the above analysis to study size dynamics under maximal risk prevention. Because of downsizing and investment, the scale of operations varies over time in the optimal contract. These variations can be interpreted as the dynamics of firm size. Our model generates a rich variety of possible paths for such dynamics. Over its life cycle, the firm can grow, stagnate, or decline.

To illustrate this point, consider the following typical path, depicted on Figure 2. Firm size starts at the level X_0 . As long as there is no loss, the size-adjusted continuation utility of the agent rises and eventually reaches the investment threshold w^i . From this point on, investment takes place at rate γ and the firm grows. However, if a loss occurs at time T_k , the size-adjusted continuation utility of the agent drops from w_{T_k} to $w_{T_k^+} = (w_{T_k} - b) \vee b$. If this lower utility level is below w^i , investment stops and firm size remains constant. Furthermore, if $w_{T_k^+} < 2b$ and there is another loss shortly afterward, downsizing is necessary. The corresponding path in the (X_t, W_{t^-}) plane is depicted on Figure 3.

While, in the short run, firm size can grow, stagnate, or decline, it is unclear how it is likely to behave in the long run. Will downsizing bring it down to 0 or will the firm grow indefinitely thanks to investment? To address this issue, we study the limit as t goes to ∞ of the average growth rate of the firm until time t . For simplicity, set X_0 to 1. Then Proposition 3 implies that this average growth rate is

$$(47) \quad \frac{\ln(X_t)}{t} = \frac{1}{t} \left[\sum_{k=1}^{N_t^-} \ln\left(\frac{w_{T_k} - b}{b} \wedge 1\right) + \int_0^t \gamma 1_{\{w_s > w^i\}} ds \right].$$

Now, let μ^w be the unique invariant measure associated to the process $\{w_{T_k}\}_{k \geq 1}$ of the agent's size-adjusted continuation utility just before losses, let μ^{w^+} be the unique invariant measure associated to the process $\{w_{T_k^+}\}_{k \geq 1}$ of the agent's size-adjusted continuation utility just after losses, and let Λ be the exponential distribution with parameter λ . Then, using an appropriate law of large numbers for Markov ergodic processes, the following result can be derived.²⁸

²⁸See, for instance, Stout (1974, Theorem 3.6.7).

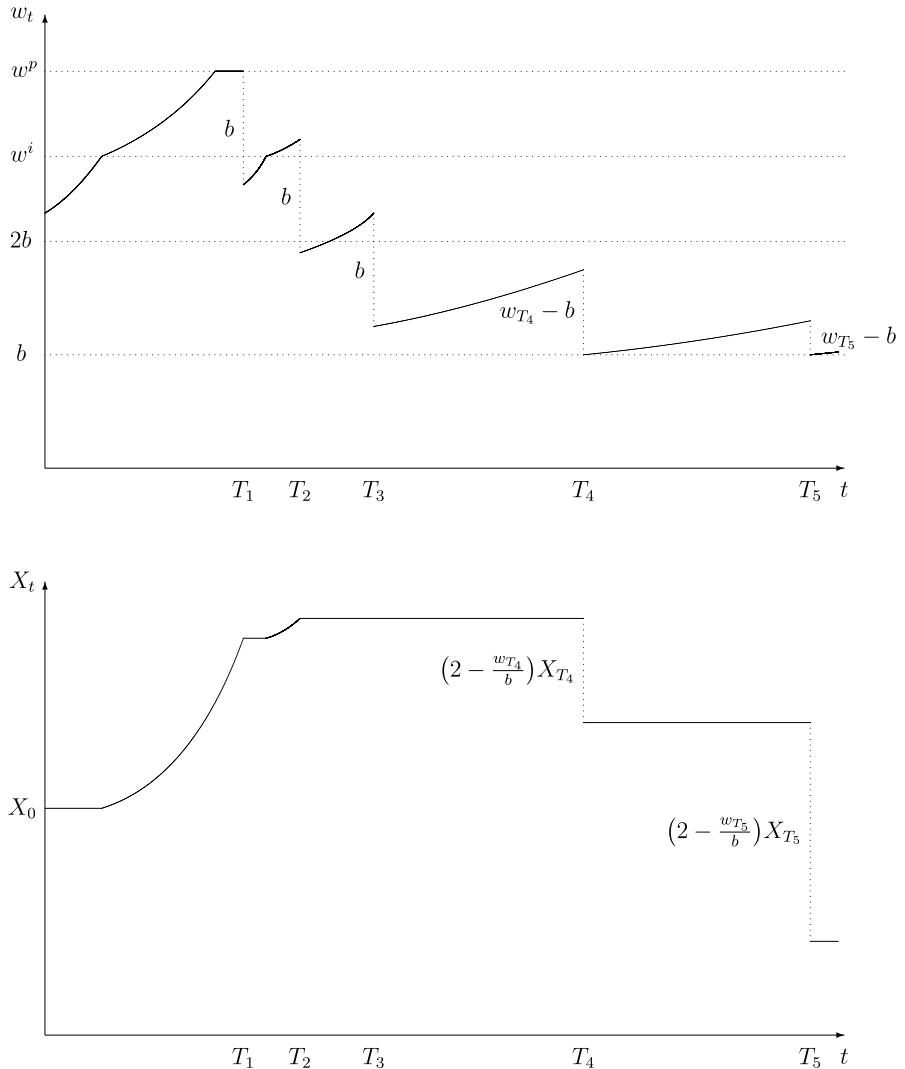


FIGURE 2.—Top panel: A sample path for the agent's size-adjusted continuation utility. Bottom panel: The corresponding path for the evolution of firm size. Investment takes place as long as $w_t > w^i$. Losses occur at times T_1-T_5 . Because $w_{T_k} > 2b$ at T_1, T_2 , and T_3 , losses at these times induce a drop of b in continuation utility and no downsizing. By contrast, $w_{T_k} < 2b$ at T_4 and T_5 , so that losses at these times induce a drop of $w_{T_k} - b$ in continuation utility and downsizing by an amount $X_{T_k} - X_{T_k}^+ = (2 - w_{T_k}/b)X_{T_k}$.

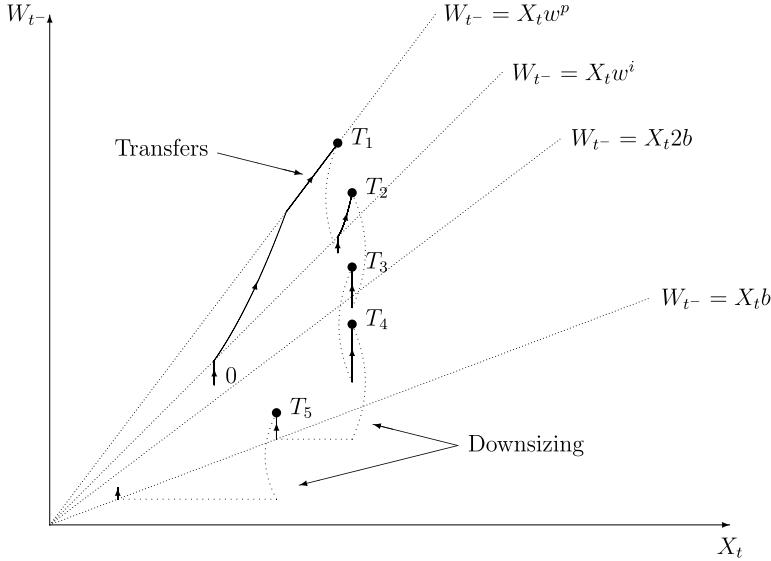


FIGURE 3.—The joint evolution of firm size and of the agent’s continuation utility for the sample path of the agent’s size-adjusted continuation utility illustrated on Figure 2. Dashed curves correspond to downward jumps in the agent’s continuation utility triggered by losses at times T_1-T_5 ; horizontal dashed lines correspond to downsizing at times T_4 and T_5 . Arrows indicate the direction of evolution of the state variables as long as no losses occur.

PROPOSITION 4: *Under conditions (39) and (40), the long-run growth rate of the firm is*

$$(48) \quad \lim_{t \rightarrow \infty} \frac{\ln(X_t)}{t} = \lambda \int_{[b, 2b]} \ln\left(\frac{w-b}{b}\right) \mu^w(dw) + \gamma \left[1 - \lambda \int_{[b, w^i] \times \mathbb{R}_+} (t_{w,w^i} \wedge s) \mu^{w+} \otimes \lambda(dw, ds) \right],$$

P-almost surely, where

$$t_{w,w^i} = \frac{1}{\rho} \ln\left(\frac{\rho w^i + \lambda b}{\rho w + \lambda b}\right)$$

is the time it takes for the agent’s size-adjusted continuation utility to reach w^i when starting from $w \in [b, w^i]$, if there are no losses in the meantime.

The first term on the right-hand side of (48) reflects the impact of downsizing. Downsizing takes place when losses occur, which is more likely if the

intensity λ of the loss process N is high, and when the size-adjusted continuation utility of the agent lies in the region $[b, 2b]$, where downsizing cannot be avoided whenever a loss occurs.

The second term on the right-hand side of (48) reflects the impact of investment. The latter takes place, at rate γ , when the size-adjusted continuation utility of the agent is above the investment threshold w^i . The term within brackets that multiplies γ on the right-hand side of (48) is the frequency with which the size-adjusted continuation utility of the agent is above w^i . To build intuition about this term, consider the time interval $(T_k, T_{k+1}]$ between two consecutive losses. There is no investment during this time interval as long as the size-adjusted continuation utility of the agent stays below w^i . The probability of that event depends on the value of the continuation utility of the agent at the beginning of the time interval, $w_{T_k^+}$, as well as on the length $T_{k+1} - T_k$ of this time interval. This is why there is a double integral in (48), with respect to the invariant measures μ^{w^+} and λ of these two independent random variables. The interpretation of the term in parentheses inside the double integral is that there is no investment during $(T_k, T_{k+1}]$ if $T_{k+1} - T_k < t_{w_{T_k^+}, w^i}$, that is, if a loss occurs before the size-adjusted continuation utility of the agent had the time to reach w^i starting from $w_{T_k^+} < w^i$.

To gain more insights into the long-run behavior of the size of the firm, consider for tractability the case where c is small, so that

$$(49) \quad f(b) - bf'_+(b) \geq c.$$

In that case, the optimal contract stipulates that investment should continuously take place at rate γ , and we obtain the following result.

PROPOSITION 5: *Under conditions (39), (40), and (49), if γ is close to 0, then*

$$(50) \quad \lim_{t \rightarrow \infty} X_t = 0,$$

P-almost surely, while if $\gamma > \lambda^2/(\rho - \gamma + \lambda)$, then

$$(51) \quad \lim_{t \rightarrow \infty} X_t = \infty,$$

P-almost surely.

First consider (50). In that case, the maximal feasible growth rate γ of the firm is low, so that the impact of investment is negligible. Now, to maintain incentive compatibility and limited liability, downsizing must take place when losses occur and the agent's size-adjusted continuation utility is close to its lower bound b . Because the stochastic process that describes the agent's size-adjusted continuation utility is Markov ergodic over $[b, w^p]$, this situation will

prevail an infinite number of times with probability 1. As a result, the size of the firm and the continuation utility of the agent must eventually go to 0.

Next consider (51). In that case, the frequency λ of losses is low relative to the maximal feasible growth rate γ of the firm, so that the positive effect of investment dominates the negative effect of downsizing. Thus, in the long run, the firm becomes infinitely large. Note, however, that, even in this case, the long-run growth rate of the firm remains strictly lower than in the first-best case, because of downsizing.

These two asymptotic results differ from the classic immiserization result of Thomas and Worrall (1990). In their model, the agent's continuation utility eventually diverges to $-\infty$, but the reason why this outcome obtains differs from the reason why, in our model, firm size goes to 0 when γ is low. Indeed, in Thomas and Worrall (1990), the period utility function of the agent is concave and unbounded below. Consequently, providing incentives is cheaper, the lower is the agent's continuation utility. This reflects the fact that the cost of obtaining a given spread in the agent's continuation utility is then lower. The principal thus has an incentive to let the agent's utility drift to $-\infty$. By contrast, in our model, the cost of incentive compatibility is high when the agent's continuation utility is low. This reflects the fact that limited liability makes it then more difficult to induce large variations in the agent's continuation utility. Yet, firm size can go to 0 if γ is low relative to λ , so that the effect of downsizing overcomes that of investment. If γ is high relative to λ , firm size goes to infinity. Now, the continuation utility of the agent is equal to her size-adjusted continuation utility, which by construction lives in $[b, w^p]$, multiplied by firm size. Hence in that case, the continuation utility of the agent grows unboundedly, which is exactly the opposite of the immiserization result.

Proposition 5 provides parameter restrictions under which firm size X_t unambiguously goes to 0 or ∞ with probability 1 when t goes to ∞ . More generally, for all parameter values, including those under which (49) does not hold, the following holds.

PROPOSITION 6: *Under conditions (39) and (40), each of the events $\{\lim_{t \rightarrow \infty} X_t = 0\}$ and $\{\lim_{t \rightarrow \infty} X_t = \infty\}$ has either a probability 0 or 1 of occurring.*

The intuition for this result is twofold. First, as can be seen from (44), the events that firm size X_t goes to 0 or to ∞ are tail events. That is, whether they occur depends on what happens in the long run, not on what happens over any finite horizon. Second, the stochastic processes that drive the evolution of firm size satisfy a mixing property, which implies that tail events have either probability 0 or 1. Note that Proposition 6 does not assert that one of the events $\{\lim_{t \rightarrow \infty} X_t = 0\}$ and $\{\lim_{t \rightarrow \infty} X_t = \infty\}$ must occur with probability 1: both of them may have probability 0. What it rules out, for instance, is a scenario in which, with probability p , the size of the firm eventually vanishes, while with probability $1 - p$ it eventually explodes, for some $p \in (0, 1)$.

This asymptotic result sharply differs from that arising in Clementi and Hopenhayn (2006). In their long-run analysis, either the firm is eventually liquidated or the first-best case is eventually attained and the firm is never liquidated. Each of these absorbing outcomes has a strictly positive probability in the stationary distribution. This difference with our results stems from the fact that, in their model, the principal and the agent have identical discount rates, while in ours, the agent is more impatient than the principal.²⁹ In Clementi and Hopenhayn (2006), because the principal and the agent are equally patient, it is costless to delay the agent's consumption while capitalizing it at the common discount rate. Hence, it is optimal to try and accumulate pledges to the agent until her savings are so high that she can buy the firm and implement the first-best case policy. With some probability, the agent is lucky enough that such high performance is achieved and the first-best case is attained. With the complementary probability, the agent is not as lucky, and liquidation eventually occurs. By contrast, in our model, delaying consumption is costly, since the agent is more impatient than the principal. It is, therefore, optimal to let her consume before the first-best case is attained. This reduces the growth in the accumulated pledge to the agent, which, in turn, raises the risk of downsizing. Whenever the maximal investment rate is low, such downsizing eventually brings firm size to 0 with probability 1. Whenever the maximal investment rate is high, firm size tends to grow so fast that it eventually explodes in spite of downsizing. Note, however, that in that case, the first-best case is not attained, even in the long run, because moral hazard still slows down the rate at which the firm grows.

6. ROBUSTNESS

In this section, we discuss the robustness of our results. We first provide sufficient conditions for the optimality of maximal risk prevention. Then we briefly examine the case of nonconstant returns to scale.

6.1. *Optimality of Maximal Risk Prevention*

So far, our analysis has focused on the optimal contract under maximal risk prevention. We now investigate under which circumstances it is actually optimal for the principal to require such a high level of effort from the agent. For simplicity, we conduct this analysis in the case where there is no investment, that is, $\gamma = 0$.

Note that the contract characterized in Proposition 3 depends on B and $\Delta\lambda$ only through their ratio $b = B/\Delta\lambda$. Hence there is 1 degree of freedom in the

²⁹In Clementi and Hopenhayn's (2006) discrete-time model, unlike in our continuous-time model, identical discount rates for the principal and the agent do not preclude the existence of an optimal contract. A further difference is that they assumed that capital fully depreciates from one period to the next, while there is no capital depreciation in our model.

parameters of the model, as B and $\Delta\lambda$ can be scaled up or down while keeping b constant, leaving the optimal contract under maximal risk prevention unaffected. Intuition suggests that when $\Delta\lambda$ gets large, it is optimal to prevent losses as much as possible. To see why, observe that if a contract induced shirking during some infinitesimal time interval $[t, t + dt]$, the agent's continuation utility would not need to be affected were a loss to occur at time t ; that is, $H_t = 0$ in (13). Since it is optimal to make no transfers over $[t, t + dt]$ as the agent is shirking, (13) then implies that this would result in a change

$$(52) \quad dw_t = (\rho w_t - B) dt$$

in the agent's size-adjusted continuation utility. To determine whether requiring the agent to always exert effort is optimal, we compare the continuation value of the principal under maximal risk prevention to its counterpart when the agent shirks during $[t, t + dt]$ and then reverts to exerting effort. The former is greater than the latter if

$$(53) \quad f(w_t) \geq [\mu - (\lambda + \Delta\lambda)C] dt + e^{-rdt} f(w_t + dw_t),$$

where dw_t is given by (52). The first term on the right-hand side of (53) reflects the increased intensity of losses over $[t, t + dt]$ due to shirking, while the second term corresponds to the continuation value to the principal from requesting maximal risk prevention from time $t + dt$ on. Given (52), a first-order Taylor expansion in (53) leads to

$$(54) \quad rf(w_t) \geq \mu - (\lambda + \Delta\lambda)C + (\rho w_t - B)f'(w_t).$$

Unlike in (33), there is no delay term on the right-hand side of (54), because the agent's continuation utility is not sensitive to losses during the time interval $[t, t + dt]$. Maximal risk prevention is optimal if (54) holds for any value of $w_t > b$. The following result is found.

PROPOSITION 7: *Suppose that $\gamma = 0$, and fix all the parameters of the model except B and $\Delta\lambda$, for which only the ratio $b = B/\Delta\lambda$ is fixed, so that an increase in B is compensated by a proportional increase in $\Delta\lambda$. Then there exists a threshold $\underline{\Delta\lambda} > 0$ such that the optimal contract involves maximal risk prevention for all $\Delta\lambda > \underline{\Delta\lambda}$.*

The intuition for this result is as follows. Both B and $\Delta\lambda$ affect the magnitude of the moral hazard problem and, therefore, the cost of incentives. However, under maximal risk prevention, they do so only via their ratio b ; formally, this is reflected in the fact that the function f depends on B and $\Delta\lambda$ only through b . Now, while an increase in $\Delta\lambda$ makes shirking easier to detect, and raises the value to the principal of a high level of risk prevention effort, an increase in B leaves this value unaffected. Hence, when b and thus the cost of incentives

is kept constant, increasing $\Delta\lambda$ raises the benefit of effort for the principal without affecting its cost. As a result, when $\Delta\lambda$ is sufficiently high, it is optimal for the principal to require the agent to always exert effort.

6.2. Nonconstant Returns to Scale

Our analysis relies on the assumption that there are constant returns to scale. What can be said when one relaxes this assumption? Suppose, for instance, that the private benefits from shirking are equal to some increasing function $B(X)$ of firm size X and, for simplicity, keep all other assumptions unchanged. Incentive compatibility conditions are basically the same in that extension. The continuation utility of the agent is written as

$$W_t(\Gamma, \Lambda) = \mathbf{E}^A \left[\int_t^\tau e^{-\rho(s-t)} [dL_s + 1_{\{\Lambda_s=\lambda+\Delta\lambda\}} B(X_s) ds] \middle| \mathcal{F}_t^N \right] 1_{\{t < \tau\}},$$

and the underlying martingale is still M^A , so that the martingale representation theorem applies and Lemma 1 continues to hold. Similarly, Proposition 1 is essentially unchanged, except that the incentive compatibility condition under which the agent exerts effort is now

$$H_t(\Gamma, \Lambda) \geq \frac{B(X_t)}{\Delta\lambda}.$$

Suppose now that the principal wants to implement maximal risk prevention. Then, as when returns to scale are constant, it will be necessary to downsize the project after a loss if the agent's continuation utility is too low. To see this more precisely, suppose that, at the outset of time t , the size of the project is X_t and the continuation utility of the agent is $W_{t-}(\Gamma, \Lambda)$. If there is a loss at time t , incentive compatibility requires that the continuation utility be reduced by at least $B(X_t)/\Delta\lambda$. Downsizing can be avoided at this point only if the new level of continuation utility is high enough that it is still possible to provide incentives while satisfying the limited liability constraint, that is, if

$$W_{t-}(\Gamma, \Lambda) - \frac{B(X_t)}{\Delta\lambda} \geq \frac{B(X_t)}{\Delta\lambda}.$$

Thus downsizing must take place whenever $W_{t-}(\Gamma, \Lambda) < 2B(X_t)/\Delta\lambda$ and there is a loss at time t . Yet, it is hard to push the analysis of the optimal contract much further without assuming constant returns to scale. Indeed, the Hamilton–Jacobi–Bellman equation now is written as

$$\begin{aligned} (55) \quad rF(X_t, W_{t-}) &= X_t(\mu - \lambda C) \\ &\quad + \max\{-X_t\ell_t + (\rho W_{t-} + \lambda H_t - X_t\ell_t)F_W(X_t, W_{t-}) \\ &\quad + X_t g_t[F_X(X_t, W_{t-}) - c] \\ &\quad - \lambda[F(X_t, W_{t-}) - F(X_t x_t, W_{t-} - H_t)]\}, \end{aligned}$$

where the maximization in (55) is over the set of controls (g_t, H_t, ℓ_t, x_t) that satisfy (5), (18), and the two constraints

$$(56) \quad \begin{aligned} H_t &\geq \frac{B(X_t)}{\Delta\lambda}, \\ W_{t^-} - H_t &\geq \frac{B(X_t x_t)}{\Delta\lambda}. \end{aligned}$$

The first of these constraints is the agent's date t incentive compatibility constraint, while the second, which parallels (19), expresses the fact that if a loss occurs at date t , reducing by H_t the continuation utility of the agent, it must still be possible to provide incentives after this loss, which requires being able to further reduce the agent's utility by $B(X_t x_t)/\Delta\lambda$, where $X_t x_t$ is the size of the firm after the date t loss. Unlike in the constant returns to scale case, the nonlinearity of $B(X)$ with respect to X makes it impossible to reduce the delay partial differential equation (55) to a delay ordinary differential equation.

While it is difficult to rigorously study the system (55)–(56) when $B(X)$ is not linear in X , a heuristic analysis similar to that in Section 4.1 can be performed for a small perturbation of the private benefits function:

$$B^\varepsilon(X) = BX + \varepsilon X \phi(X),$$

where ε is a small number and ϕ is a bounded function. This analysis, which can be found in the supplement (Biais, Mariotti, Rochet, and Villeneuve (2010)), suggests that, under regularity conditions, the qualitative properties of the optimal contract can reasonably be expected to be upheld for such a small perturbation. The optimal contract could then be depicted on a figure similar to Figure 1. The differences would be that the boundary of the downsizing region would be the nonlinear function $B^\varepsilon(X)/\Delta\lambda$ of firm size X instead of the linear function Xb , and that the upper and lower boundaries of the investment/no transfers region would also presumably be nonlinear functions of X .

7. EMPIRICAL IMPLICATIONS

While, in the first-best case, firms in our model should always invest, in the second-best case the optimal contract stipulates that firms can invest only after a long enough record of good performance, at least when the unit cost of investment is not too low.³⁰ Such clauses are consistent with the empirical results of Kaplan and Strömberg (2004), who found that venture capital funding for new investment is contingent on financial and nonfinancial milestones. They also found that such conditioning is more frequent when the proxy for agency problems is more severe.

³⁰Throughout this section, we assume that $f(b) - bf'_+(b) < c < \bar{c}$, so that $w^i > b$.

In our model, the optimal contract specifies that after good performance, agents will be compensated, while after bad performance, the firm will be partially liquidated. This is in line with the contractual clauses documented by Kaplan and Strömborg (2003). The circumstances in which downsizing takes place in the optimal contract can be interpreted as financial distress. This is in line with the empirical findings of Denis and Shome (2005), who reported that financially distressed firms are often downsized.

In our model, small firms tend to be below the investment threshold. They are thus likely to be exposed to financial constraints on investment, as documented by Beck, Demirguc-Kunt, and Maksimovic (2005). Our model also predicts that small firms are relatively more fragile, since a few negative shocks are enough to drive them into the zone where further losses trigger downsizing. Conversely, large firms that have enjoyed long periods of sustained investment are more likely to have long records of good performance, which pushes them away from that zone. Overall, the probability of downsizing is decreasing in firm size. This is in line with the empirical findings of Dunne, Roberts, and Samuelson (1989), who reported that failure rates decline with increases in firm or plant size. Note, however, that the same logic implies that, according to our model, large firms should tend to have higher growth rates than smaller ones, while data suggest that on average the opposite is true; see Evans (1987a, 1987b) and Dunne, Roberts, and Samuelson (1989). Interestingly, though, Dunne, Roberts, and Samuelson (1989) found that this pattern is reversed in the case of multiplant firms: mean growth rates for plants owned by such firms tend to increase with size, reflecting that the tendency for growth rates of plants to decline with size is outweighed by a substantial fall in their failure rates. This evidence suggests that our analysis is particularly relevant for multiplant firms. A further testable implication of our model is that downsizing decisions should typically be followed by relatively long periods during which no investment takes place, corresponding to the time it takes for the firm to reach the investment threshold again and resume growing.

Gabaix and Landier (2008) noted that different theoretical explanations have been offered for variations in CEO pay. While some analyses emphasize incentive problems, Gabaix and Landier (2008) proposed to focus on firm size. Empirically, they found that CEO pay increases with firm size. Consistent with these results, our incentive theoretic analysis implies that the size of the firm and the compensation of the agent ought to be positively correlated: after a long stream of good performance, the scale of operations is large and so are the payments to the agent. Conceptually, our analysis suggests that explanations based on size should not be divorced from explanations based on incentives, and that investment and managerial compensation are complementary incentive instruments, in line with the empirical findings of Kaplan and Strömborg (2003).

8. CONCLUSION

This paper analyzes the dynamic moral hazard problem arising when agents with limited liability must exert costly unobservable effort to reduce the likelihood of large but relatively infrequent losses. We characterize the optimal downsizing, investment, and compensation policies, and provide explicit formulae for firm size and its asymptotic growth rate.

Our analysis generates policy and managerial implications for the prevention of large risks. Losses in our model are negative externalities, since they affect society beyond the managers' or the firms' ability to pay for the damages they cause. It is, therefore, natural to think of the optimal dynamic contract as a regulatory tool. For instance, in the context of financial institutions, our analysis suggests that to prevent large losses, downsizing and investment decisions should be made contingent on accumulated performance. This notably provides a rationale for prudential regulations that request that the scale at which financial firms operate be proportionate to their capital. In particular, such regulations imply that banks or insurance companies should be downsized if their capital before large losses is close to the regulatory requirement. This is similar to our optimal contract, provided W is interpreted as a proxy for capital, which is natural since both increase after good performance and decrease after bad performance. Yet our analysis suggests that such capital requirements are not sufficient to induce an optimal level of risk prevention: they should be complemented by an appropriate regulation of managerial compensation. More specifically, the managers' compensation should be based on long-term track records, and it should be reduced after large losses by an amount that increases with the private benefits from shirking and the extent to which shirking is difficult to detect.

Our analysis also generates implications for firm size dynamics. Simon and Bonini (1958) and Ijiri and Simon (1964) analyzed the link between the stochastic process according to which firms grow and the size distribution of firms. While these early works do not rely on the characterization of optimal investment policies, they have been embedded within the neoclassical framework; see, for instance, Lucas (1978) or Luttmer (2007, 2008). In these models, firm growth is limited by technology. In Lucas (1978), managerial skills are assumed to exhibit diminishing returns to scale, while in Luttmer (2008), it is assumed that when ideas are replicated, their quality depreciates. Our modeling framework offers an opportunity to revisit these issues in a context where the endogenous limits to firm growth result from moral hazard. A key issue in models of the size distribution of firms is whether Gibrat's law holds, that is, whether firm growth is independent of firm size. This is not the case in our model, since firm size, and downsizing and investment decisions are correlated in the optimal contract, being all functions of the agent's size-adjusted continuation utility process. It would be interesting, in further research, to analyze the implications of our analysis for the size distribution of firms.

APPENDIX: SKETCHES OF PROOFS

In this appendix, we merely outline the structure of the proofs. The interested reader can find complete proofs in the supplement (Biais, Mariotti, Rochet, and Villeneuve (2010)). All the references hereafter made to sections and auxiliary results correspond to this supplementary document.

PROOF OF LEMMA 1—Sketch: The predictable representation (12) of the martingale $U(\Gamma, \Lambda)$ follows from Brémaud (1981, Chapter III, Theorems T9 and T17). The factor $e^{-\rho s}$ in (12) is just a convenient rescaling. *Q.E.D.*

PROOF OF PROPOSITION 1—Sketch: The proof extends Sannikov's (2008, Proposition 2) arguments to the case where output is modeled as a point process. *Q.E.D.*

PROOF OF PROPOSITION 2—Sketch: It turns out to be more convenient to work with the size-adjusted social value function, defined by $v(w) = f(w) + w$ for all $w \geq 0$. Just as f , the function v is linear over $[0, b]$. From (33) and (35)

$$(57) \quad rv(w) = \mu - \lambda C - (\rho - r)w + \mathcal{L}v(w)$$

for all $w \in (b, w^i]$ and

$$(58) \quad (r - \gamma)v(w) = \mu - \lambda C - \gamma c - (\rho - r)w + \mathcal{L}_\gamma v(w)$$

for all $w \in (w^i, w^p]$. The investment threshold w^i satisfies

$$(59) \quad w^i = \inf\{w > b \mid v(w) - wv'(w) > c\},$$

while the payment threshold w^p satisfies

$$(60) \quad v'(w^p) = 0.$$

Finally, v is constant and equal to $v(w^p)$ over $[w^p, \infty)$. The proof consists of two main parts. In the first part of the proof (Section C.1), we suppose that investment is not feasible, that is, $\gamma = 0$. This allows us to pin down the constant \bar{c} in (40) and provides key insights into the properties of the solution to (41) in the no investment region $(b, w^i]$. In the second part of the proof (Section C.2), we suppose that investment is feasible, that is, $\gamma > 0$, and we use the results of the first part of the proof to solve (41).

Part 1: In the no investment case, we look for the maximal solution to (57) that satisfies (60) at some payment threshold. Note that the only unknown parameter is the slope of that solution over $[0, b]$. To determine that slope, we use the following shooting method. For each $\beta \geq 0$, denote by v_β the function that is linear with slope β over $[0, b]$ and then satisfies (57) over (b, ∞) . It can be shown that v_β can be decomposed over \mathbb{R}_+ as $u_1 + \beta u_2$, where u_2

is a nonnegative function with strictly positive derivative.³¹ This implies that the derivatives of the functions $(v_\beta)_{\beta \geq 0}$ are strictly increasing with respect to β (Proposition C.1.1). We then prove that the ratio $-u'_1/u'_2$ attains a maximum β_0 over (b, ∞) , which implies that v_{β_0} is the maximal function in the family $(v_\beta)_{\beta \geq 0}$ whose derivative has a zero in (b, ∞) (Proposition C.1.2). Thus v_{β_0} is the desired maximal solution. Let $w_{\beta_0}^p$ be the first point at which v'_{β_0} vanishes. The last step of the proof then consists of showing that v_{β_0} is concave over $[0, w_{\beta_0}^p]$, and strictly so over $[b, w_{\beta_0}^p]$ (Proposition C.1.3). As explained in the text, the cost threshold \bar{c} below which investment is strictly profitable is $v_{\beta_0}(w_{\beta_0}^p)$. For $c > \bar{c}$, the size-adjusted social value function is $v_{\beta_0} \wedge v_{\beta_0}(w_{\beta_0}^p)$ (Section D.2, Remark).

Part 2: In the investment case, we look for the maximal solution to (57) and (58) that satisfies (59) and (60) at some investment and payment thresholds. As in part 1, the only unknown parameter is the slope of v over $[0, b]$. To determine that slope, which must clearly be higher than β_0 , we use the following shooting method. For each $\beta \geq \beta_0$, denote by $v_{\beta,\gamma}$ the function that is linear with slope β over $[0, b]$ and then satisfies (57) over $(b, w_\beta^i]$ and (58) over (w_β^i, ∞) , where $w_\beta^i = \inf\{w > b \mid v_{\beta,\gamma}(w) - wv'_{\beta,\gamma}(w) > c\}$. We may have $w_\beta^i = b$, in which case the region $(b, w_\beta^i]$ is empty. We first show that $v_{\beta,\gamma}$ is well defined, and that the threshold w_β^i belongs to $[b, w_{\beta_0}^p)$ and continuously decreases as β increases (Lemma C.2.1). Key to this result is the fact that u_2 is strictly concave over $[b, \infty)$. We then show that, in analogy with the functions $(v_\beta)_{\beta \geq 0}$, the derivatives of the functions $(v_{\beta,\gamma})_{\beta \geq \beta_0}$ are strictly increasing with respect to β (Proposition C.2.1). The next step of the proof, which is crucial, consists of showing that there exists a maximal function $v_{\beta,\gamma,\gamma}$ in the family $(v_{\beta,\gamma})_{\beta \geq \beta_0}$ whose derivative has a zero in (b, ∞) (Proposition C.2.2). To establish this result, we first show that the set of $\beta \geq \beta_0$ such that $v'_{\beta,\gamma}$ has a zero over (b, ∞) is a nonempty interval I that contains β_0 (Lemma C.2.2). Second, we show that I has a finite upper bound β_γ , so that $v'_{\beta,\gamma}$ has no zero in (b, ∞) when $\beta > \beta_\gamma$ (Lemma C.2.3). Third, letting $w_{\beta,\gamma}^p$ be the first point at which $v'_{\beta,\gamma}$ vanishes for any given $\beta \in I$, we show that $w_{\beta,\gamma}^p$ is strictly increasing with respect to β over I and converges to a finite limit when β converges to β_γ from below (Lemma C.2.4). Fourth, we show that the derivatives of the functions $(v_{\beta,\gamma})_{\beta \geq \beta_0}$ vary continuously with β , which in turn implies that I contains its upper bound β_γ (Lemma C.2.5). Thus $v_{\beta,\gamma,\gamma}$ is the desired maximal solution and $w_{\beta_\gamma,\gamma}^p$ is the first point at which $v'_{\beta_\gamma,\gamma}$ vanishes. The last step of the proof then consists of showing that $v_{\beta,\gamma,\gamma}$ is concave over $[0, w_{\beta_\gamma,\gamma}^p]$ and strictly so over $[b, w_{\beta_\gamma,\gamma}^p]$ (Proposition C.2.3). Key to this result is the fact that $\beta_\gamma > \beta_0$ and that the maximal solution v_{β_0} derived in the no investment case is concave over $[0, w_{\beta_0}^p]$ as established in Proposition C.1.3. Finally, let-

³¹The functions u_1 , u_2 , and v_β are continuously differentiable except at b .

ting $f(w) = v_{\beta_\gamma, \gamma}(w) \wedge v_{\beta_\gamma, \gamma}(w_{\beta_\gamma, \gamma}^p) - w$ for all $w \geq 0$ and writing $w^i = w_{\beta_\gamma}^i$ and $w^p = w_{\beta_\gamma, \gamma}^i$ to simplify notation, it is immediate to check that the triple (f, w^i, w^p) satisfies all the properties stated in Proposition 2. Q.E.D.

PROOF OF PROPOSITION 3—Sketch: The argument follows somewhat standard lines in optimal control theory. In the first step of the proof, we establish that F provides an upper bound for the expected payoff that the principal can obtain from any incentive compatible contract that induces maximal risk prevention, that is,

$$(61) \quad F(X_0, W_{0-}) \geq \mathbf{E} \left[\int_0^\tau e^{-rt} \{ X_t [(\mu - g_t c) dt - CdN_t] - dL_t \} \right]$$

for any contract $\Gamma = (X, L, \tau)$ that induces maximal risk prevention. For any such contract, the dynamics of the agent's continuation utility W is given by (13) for a process H that satisfies the incentive compatibility condition (14). Substituting X and L from Γ into the function F , and applying the change of variable formula for processes of locally bounded variation (Dellacherie and Meyer (1982, Chapter VI, Section 92)) yields

$$\begin{aligned} (62) \quad F(X_0, W_{0-}) &= e^{-rT} F(X_{T^+}, W_T) \\ &\quad - \int_0^T e^{-rt} [(\rho W_{t-} + \lambda H_t) F_W(X_t, W_{t-}) \\ &\quad - rF(X_t, W_{t-})] dt \\ &\quad - \int_0^T e^{-rt} F_X(X_t, W_{t-})(dX_t^{d,c} + X_t g_t dt) \\ &\quad + \int_0^T e^{-rt} F_W(X_t, W_{t-}) dL_t^c \\ &\quad - \sum_{t \in [0, T]} e^{-rt} [F(X_{t^+}, W_t) - F(X_t, W_{t-})] \end{aligned}$$

for all $T \in [0, \tau]$, where $X^{d,c}$ and L^c stand for the pure continuous parts of X^d and L . Imposing limited liability and incentive compatibility, along with the homogeneity of F , the concavity of f , and the fact that $f'_+ \geq -1$, we show that, in expectation, the right-hand side of (62) is greater than that of (61).

In the second step of the proof, we establish that the contract described in Proposition 3 yields the principal a value $F(X_0, W_{0-})$. This contract must, therefore, be optimal, since, from the first step of the proof, $F(X_0, W_{0-})$ is an upper bound for the value that the principal can derive from any contract that induces maximal risk prevention. Specifically, we start from (62) and we use the properties of the contract spelled out in Properties 1–6, and more precisely

described in Proposition 3, to show that, in expectation, the right-hand side of (62) is in this case equal to that of (61). $\mathcal{Q.E.D.}$

PROOF OF PROPOSITION 4—Sketch: In the first step of the proof, we establish that the process $\{w_{T_k}\}_{k \geq 1}$ is Markov ergodic and then rely on the strong law of large numbers for Markov ergodic processes (Stout (1974, Theorem 3.6.7)) to show that

$$(63) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^{N_t^-} \ln \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) = \lambda \int_{[b, 2b)} \ln \left(\frac{w - b}{b} \right) \mu^w(dw),$$

\mathbf{P} -almost surely. The main technical difficulty consists of proving that the integral on the right-hand side of (63) is finite.

In the second step of the proof, we establish that

$$(64) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^{T_{N_t^-}} 1_{\{w_s > w^i\}} ds = 1 - \lambda \int_{[b, w^i] \times \mathbb{R}_+} (t_{w, w^i} \wedge s) \mu^{w+} \otimes \lambda(dw, ds),$$

\mathbf{P} -almost surely. The argument goes as follows. Consider for each $k \geq 1$ the integral $I_k = \int_{T_{k-1}}^{T_k} 1_{\{w_s > w^i\}} ds$, where $T_0 = 0$ by convention. If $w_{T_{k-1}^+} \geq w^i$, then $w_s > w^i$ for all $s \in (T_{k-1}, T_k]$, and thus $I_k = T_k - T_{k-1}$. If $w_{T_{k-1}^+} < w^i$ and $T_k - T_{k-1} \leq t_{w_{T_{k-1}^+}, w^i}$, then $w_s \leq w^i$ for all $s \in (T_{k-1}, T_k]$, and thus $I_k = 0$. Last, if $w_{T_{k-1}^+} < w^i$ and $T_k - T_{k-1} > t_{w_{T_{k-1}^+}, w^i}$, then $w_s > w^i$ for all $s \in (T_{k-1} + t_{w_{T_{k-1}^+}, w^i}, T_k]$, and thus $I_k = T_k - T_{k-1} - t_{w_{T_{k-1}^+}, w^i}$. Summing over $k = 1, \dots, n$ and rearranging yields

$$(65) \quad \begin{aligned} \frac{1}{n} \int_0^{T_n} 1_{\{w_s > w^i\}} ds &= \frac{1}{n} \sum_{k=1}^n (T_k - T_{k-1}) \\ &\quad - \frac{1}{n} \sum_{k=1}^n [t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1})] 1_{\{w_{T_{k-1}^+} < w^i\}} \end{aligned}$$

for all $n \geq 1$. Since the random variables $(T_k - T_{k-1})_{k \geq 1}$ are independently and identically distributed according to the exponential distribution λ , it follows from the strong law of large numbers that the sequence $(\frac{1}{n} \sum_{k=1}^n (T_k - T_{k-1}))_{n \geq 1}$ converges \mathbf{P} -almost surely to $1/\lambda$. Furthermore, we show that the process $\{(w_{T_{k-1}^+}, T_k - T_{k-1})\}_{k \geq 1}$ is Markov ergodic, with invariant measure $\mu^{w+} \otimes \lambda$ over $[b, w^p] \times \mathbb{R}_+$. Since the function $(w, s) \mapsto (t_{w, w^i} \wedge s) 1_{\{w < w^i\}}$ is measurable, nonnegative, and bounded above by $(w, s) \mapsto s$, and hence is $\mu^{w+} \otimes \lambda$ -integrable, it follows from the strong law of large numbers for Markov ergodic

processes (Stout (1974, Theorem 3.6.7)) that the sequence $(\frac{1}{n} \sum_{k=1}^n [t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1})] \mathbf{1}_{\{w_{T_{k-1}^+} < w^i\}})_{n \geq 1}$ converges \mathbf{P} -almost surely to

$$\begin{aligned} & \int_{[b, w^p] \times \mathbb{R}_+} (t_{w, w^i} \wedge s) \mathbf{1}_{\{w < w^i\}} \mu^{w+} \otimes \lambda(dw, ds) \\ &= \int_{[b, w^i] \times \mathbb{R}_+} (t_{w, w^i} \wedge s) \mu^{w+} \otimes \lambda(dw, ds). \end{aligned}$$

Using the fact that N_{t^-}/t converges \mathbf{P} -almost surely to λ as t goes to ∞ by the strong law of large numbers for the Poisson process then yields (64).

In the last step of the proof, we establish that

$$(66) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \int_{T_{N_{t^-}}}^t \mathbf{1}_{\{w_s > w^i\}} ds = 0,$$

\mathbf{P} -almost surely. Merging (63), (64), and (66) finally leads to (48). *Q.E.D.*

PROOF OF PROPOSITION 5—Sketch: We first check that (49) holds uniformly in γ whenever c is close enough to 0. This implies that the expression (48) for the long-run growth rate of the firm simplifies to

$$(67) \quad \lim_{t \rightarrow \infty} \frac{\ln(X_t)}{t} = \lambda \int_{[b, 2b]} \ln\left(\frac{w-b}{b}\right) \mu^w(dw) + \gamma.$$

The remainder of the proof consists of constructing upper and lower bounds for the integral on the right-hand side of (67).

To construct the upper bound, we first define $\bar{w}^p = (\mu - \lambda C)/(\rho - r)$ and show that $w^p < \bar{w}^p$ uniformly in γ . We then define auxiliary processes $\{\bar{w}_t\}_{t \geq 0}$ and $\{\bar{l}_t\}_{t \geq 0}$ by

$$\begin{aligned} \bar{w}_t &= w_0 + \int_0^{t^-} \left\{ (\rho \bar{w}_s + \lambda b) ds - b \left(\frac{\bar{w}_s - b}{b} \wedge 1 \right) dN_s - d\bar{l}_s \right\}, \\ \bar{l}_t &= \max\{\bar{w}_0 - \bar{w}^p, 0\} + \int_0^t (\rho \bar{w}^p + \lambda b) \mathbf{1}_{\{\bar{w}_{s+} = \bar{w}^p\}} ds \end{aligned}$$

for all $t \geq 0$, that are independent of γ . It is easy to check that $w_t \leq \bar{w}_t$ for all $t \geq 0$ and that $\{\bar{w}_{T_k}\}_{k \geq 1}$ has a unique stationary initial distribution $\mu^{\bar{w}}$. Furthermore,

$$\int_{[b, 2b]} \ln\left(\frac{w-b}{b}\right) \mu^w(dw) \leq \int_{[b, 2b]} \ln\left(\frac{w-b}{b}\right) \mu^{\bar{w}}(dw) < 0,$$

uniformly in γ , which yields the desired upper bound. The strict inequality follows from the fact that for each $k \geq 1$ and $w \in (b, \bar{w}^p]$, there is for each $\varepsilon > 0$

close enough to 0 a strictly positive probability that $\bar{w}_{T_{k+1}} < w$ given that $\bar{w}_{T_k} = w + \varepsilon$, which implies that the lower bound of the support of the stationary initial distribution $\mu^{\bar{w}}$ of $\{\bar{w}_{T_k}\}_{k \geq 1}$ is b . Therefore, for γ close enough to 0,

$$\lambda \int_{[b, 2b)} \ln\left(\frac{w-b}{b}\right) \mu^w(dw) + \gamma < 0,$$

which establishes (50).

The lower bound is provided by the fact that $\int_{[b, 2b)} \ln((w-b)/b) \mu^w(dw)$ is finite (Section E, Proof of Proposition 4, Claim 1, Step 2). Specifically, one can show that

$$\int_{[b, 2b)} \ln\left(\frac{w-b}{b}\right) \mu^w(dw) \geq -\frac{\lambda}{\rho - \gamma + \lambda}$$

uniformly in γ . Therefore, if $\gamma > \lambda^2/(\rho - \gamma + \lambda)$,

$$\lambda \int_{[b, 2b)} \ln\left(\frac{w-b}{b}\right) \mu^w(dw) + \gamma > 0,$$

which establishes (51). *Q.E.D.*

PROOF OF PROPOSITION 6—Sketch: Consider for each $k \geq 1$ the σ -fields

$$\begin{aligned} \mathcal{F}_1^k &= \sigma((w_0, T_1 - T_0), (w_{T_1}, T_2 - T_1), \dots, (w_{T_{k-1}}, T_k - T_{k-1})), \\ \mathcal{F}_k^\infty &= \sigma((w_{T_{k-1}}, T_k - T_{k-1}), (w_{T_k}, T_{k+1} - T_k), \dots), \end{aligned}$$

and denote by

$$\mathcal{T} = \bigcap_{k=1}^{\infty} \mathcal{F}_k^\infty$$

the corresponding tail σ -field. The first step of the proof consists of showing that for each $E \in \mathcal{T}$, either $\mathbf{P}[E] = 0$ or $\mathbf{P}[\bar{E}] = 1$. To establish this zero–one law, we first show that for each $\varepsilon > 0$, there exists $n_0 \geq 1$ such that

$$\begin{aligned} (68) \quad \Delta(k, n, w, t, A) &= \mathbf{P}[(w_{T_{k+n-1}}, T_{k+n} - T_{k+n-1}) \in A \mid (w_{T_{k-1}}, T_k - T_{k-1}) = (w, t)] \\ &\quad - \mathbf{P}[(w_{T_{k+n-1}}, T_{k+n} - T_{k+n-1} \in A)] \\ &\leq \varepsilon \end{aligned}$$

for all $k \geq 1, n \geq n_0, (w, t) \in [b, w^p] \times \mathbb{R}_+$, and $A \in \mathcal{B}([b, w^p] \times \mathbb{R}_+)$. Following Bártfai and Révész (1967, Example 2), we can then show that a consequence

of condition (68) is that for each $\varepsilon > 0$, there exists $n_0 \geq 1$ such that the mixing property

$$(69) \quad \mathbf{P}[E | \mathcal{F}_1^k] - \mathbf{P}[E] \leq \varepsilon$$

holds for all $k \geq 1$, $n \geq n_0$, and $E \in \mathcal{F}_{k+n}^\infty$, \mathbf{P} -almost surely. Fix some $E \in \mathcal{T}$, so that in particular $E \in \mathcal{F}_{k+n}^\infty$ for all $n \geq n_0$. Since ε is arbitrary, the mixing property (69) then implies that $\mathbf{P}[E | \mathcal{F}_1^k] \leq \mathbf{P}[E]$ for all $k \geq 1$, \mathbf{P} -almost surely. From Doob (1953, Chapter VII, Theorem 4.3), it follows that $\mathbf{P}[E | \bigvee_{k=1}^\infty \mathcal{F}_1^k] \leq \mathbf{P}[E]$, \mathbf{P} -almost surely. Since $E \in \mathcal{T} \subset \bigvee_{k=1}^\infty \mathcal{F}_1^k$, we finally have $\mathbf{P}[E] = \int_E \mathbf{P}[E | \bigvee_{k=1}^\infty \mathcal{F}_1^k] d\mathbf{P} \leq \int_E \mathbf{P}[E] d\mathbf{P} = \mathbf{P}[E]^2$. Thus either $\mathbf{P}[E] = 0$ or $\mathbf{P}[E] = 1$, as claimed.

The second step of the proof consists of showing that each of the events $\{\lim_{n \rightarrow \infty} X_{T_n} = 0\}$ and $\{\lim_{n \rightarrow \infty} X_{T_n^+} = \infty\}$ belongs to \mathcal{T} . First consider $\{\lim_{n \rightarrow \infty} X_{T_n} = 0\}$. Fix some $k_0 \geq 1$. For each $n \geq k_0 + 1$, we have

$$\begin{aligned} X_{T_n} &= X_0 \prod_{k=1}^{N_{T_n^-}} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) \exp \left(\int_0^{T_n} \gamma 1_{\{w_s > w^i\}} ds \right) \\ &= X_0 \prod_{k=1}^{n-1} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) \\ &\quad \times \exp \left(\gamma \left\{ \sum_{k=1}^n (T_k - T_{k-1}) \right. \right. \\ &\quad \left. \left. - \sum_{k=1}^n [t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1})] 1_{\{w_{T_{k-1}^+} < w^i\}} \right\} \right) \\ &= X_{T_{k_0}} \prod_{k=k_0}^{n-1} \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) \\ &\quad \times \exp \left(\gamma \left\{ \sum_{k=k_0+1}^n (T_k - T_{k-1}) \right. \right. \\ &\quad \left. \left. - \sum_{k=k_0+1}^n [t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1})] 1_{\{w_{T_{k-1}^+} < w^i\}} \right\} \right) \end{aligned}$$

with $\prod_{\emptyset} = 1$ by convention, where the second equality follows from (65) and from the fact that $N_{T_n^-} = n - 1$. Since $X_{T_{k_0}}$ is a strictly positive random variable, it follows that $\{\lim_{n \rightarrow \infty} X_{T_n} = 0\} \in \mathcal{F}_{k_0+1}^\infty$. Since k_0 is arbitrary, $\{\lim_{n \rightarrow \infty} X_{T_n} = 0\} \in \mathcal{T}$.

$0\} \in \mathcal{T}$. The proof for $\{\lim_{n \rightarrow \infty} X_{T_n^+} = \infty\}$ is similar, observing that

$$\begin{aligned} X_{T_n^+} &= X_{T_{k_0}^+} \prod_{k=k_0+1}^n \left(\frac{w_{T_k} - b}{b} \wedge 1 \right) \\ &\times \exp \left(\gamma \left\{ \sum_{k=k_0+1}^n (T_k - T_{k-1}) \right. \right. \\ &\quad \left. \left. - \sum_{k=k_0+1}^n [t_{w_{T_{k-1}^+}, w^i} \wedge (T_k - T_{k-1})] 1_{\{w_{T_{k-1}^+} < w^i\}} \right\} \right) \end{aligned}$$

and that $X_{T_{k_0}^+}$ is a finite random variable.

Finally, to conclude the proof, we verify that $\{\lim_{t \rightarrow \infty} X_t = 0\} = \{\lim_{n \rightarrow \infty} X_{T_n} = 0\}$ and $\{\lim_{t \rightarrow \infty} X_t = \infty\} = \{\lim_{n \rightarrow \infty} X_{T_n^+} = \infty\}$. *Q.E.D.*

PROOF OF PROPOSITION 7: Define $w_{\beta_0}^p$ as in the proof of Proposition 2. It can be shown that

$$(70) \quad rf(w_t) \geq \mu - \lambda C + (\rho w_t + \lambda b)f'(w_t) - \lambda[f(w_t) - f(w_t - b)]$$

for any value of $w_t > b$, with equality if $w_t \in (b, w_{\beta_0}^p]$ (Section D.2, Remark). Hence a sufficient condition for (54) to hold is that the right-hand side of (70) be larger than the right-hand side of (54), which is the case if

$$(71) \quad \Delta\lambda[C + bf'(w_t)] \geq \lambda[f(w_t) - f(w_t - b) - bf'(w_t)]$$

since $b = B/\Delta\lambda$. The right-hand side of (71) is nonnegative by concavity of f and it is bounded because f is affine over $(w_{\beta_0}^p, \infty)$. Consider next the left-hand side of (71). By (2), we have $C > b$, reflecting that maximal risk prevention is socially optimal in the first-best case.³² Since $f' \geq -1$, this implies that the mapping $C + bf'$ is strictly positive and bounded away from 0. Since f depends on B and $\Delta\lambda$ only through their ratio b , it follows that (71) is satisfied for any value of $w_t > b$ when $\Delta\lambda$ is high enough, while B is proportionally adjusted so as to keep b constant. The result follows. *Q.E.D.*

REFERENCES

ABREU, D., P. MILGROM, AND D. PEARCE (1991): “Information and Timing in Repeated Partnerships,” *Econometrica*, 59, 1713–1733. [78]

³²Note that in the limit first-best case, the principal’s continuation value is linear in the agent’s continuation utility, with a slope equal to -1 . Condition (71) then reduces to $C > b$, as postulated in (2).

- AKERLOF, G. A., AND L. F. KATZ (1989): "Workers' Trust Funds and the Logic of Wage Profiles," *Quarterly Journal of Economics*, 104, 525–536. [75]
- BÁRTFAI, P., AND P. RÉVÉSZ (1967): "On a Zero-One Law," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 7, 43–47. [113]
- BECK, T., A. DEMIRGÜC, AND V. MAKSIMOVIC (2005): "Financial and Legal Constraints to Growth: Does Firm Size Matter?" *Journal of Finance*, 60, 137–177. [106]
- BIAIS, B., T. MARIOTTI, G. PLANTIN, AND J.-C. ROCHE (2007): "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications," *Review of Economic Studies*, 74, 345–390. [77,81,82,96]
- BIAIS, B., T. MARIOTTI, J.-C. ROCHE, AND S. VILLENEUVE (2010): "Supplement to 'Large Risks, Limited Liability and Dynamic Moral Hazard': Proofs," *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/7261_Proofs.pdf. [79,96,105,108]
- BRÉMAUD, P. (1981): *Point Processes and Queues: Martingale Dynamics*. New York, Heidelberg, Berlin: Springer-Verlag. [84,108]
- CLEMENTI, G. L., AND H. HOPENHAYN (2006): "A Theory of Financing Constraints and Firm Dynamics," *Quarterly Journal of Economics*, 121, 229–265. [76,81,96,102]
- DELLACHERIE, C., AND P.-A. MEYER (1978): *Probabilities and Potential*, Vol. A. Amsterdam: North-Holland. [83]
- (1982): *Probabilities and Potential*, Vol. B. Amsterdam: North-Holland. [88,110]
- DEMARZO, P. M., AND M. J. FISHMAN (2007a): "Agency and Optimal Investment Dynamics," *Review of Financial Studies*, 20, 151–188. [76,77,81,82]
- (2007b): "Optimal Long-Term Financial Contracting," *Review of Financial Studies*, 20, 2079–2128. [76,81,82,96]
- DEMARZO, P. M., AND Y. SANNIKOV (2006): "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model," *Journal of Finance*, 61, 2681–2724. [77,81,82,96]
- DEMARZO, P. M., M. J. FISHMAN, Z. HE, AND N. WANG (2008): "Dynamic Agency and the q Theory of Investment," Unpublished Manuscript, Stanford University. [77]
- DENIS, D. K., AND D. K. SHOME (2005): "An Empirical Investigation of Corporate Asset Downsizing," *Journal of Corporate Finance*, 11, 427–448. [106]
- DOOB, J. L. (1953): *Stochastic Processes*. New York: Wiley. [114]
- DUNNE, T., M. J. ROBERTS, AND L. SAMUELSON (1989): "The Growth and Failure of U.S. Manufacturing Plants," *Quarterly Journal of Economics*, 104, 671–698. [106]
- EVANS, D. S. (1987a): "The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," *Journal of Industrial Economics*, 35, 567–581. [106]
- (1987b): "Tests of Alternative Theories of Firm Growth," *Journal of Political Economy*, 95, 657–674. [106]
- GABAIX, X., AND A. LANDIER (2008): "Why Has CEO Pay Increased so Much?" *Quarterly Journal of Economics*, 123, 49–100. [106]
- GORDON, R. P. E., R. H. FLIN, K. MEANS, AND M. T. FLEMING (1996): "Assessing the Human Factors Causes of Accidents in the Offshore Oil Industry," in *Proceedings of the 3rd International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production*, ed. by M. Ognedal, R. J. Edwardes, and J. P. Visser. Richardson: Society of Petroleum Engineers, 635–644. [73]
- GREEN, E. (1987): "Lending and the Smoothing of Uninsurable Income," in *Contractual Arrangements for Intertemporal Trade*, ed. by E. C. Prescott and N. Wallace. Minneapolis: University of Minnesota Press, 3–25. [76]
- HAYASHI, F. (1982): "Tobin's Marginal q and Average q : A Neoclassical Interpretation," *Econometrica*, 50, 213–224. [75,79]
- HOLLNAGEL, E. (2002): "Understanding Accidents—From Root Causes to Performance Variability," in *New Century, New Trends: Proceedings of the 7th IEEE Conference on Human Factors and Power Plants*, ed. by J. J. Persensky, B. P. Hallbert, and H. S. Blackman. New York: Intitute of Electrical and Electronics Engineers, 1–6. [73]

- HOLMSTRÖM, B., AND J. TIROLE (1997): "Financial Intermediation, Loanable Funds, and the Real Sector," *Quarterly Journal of Economics*, 112, 663–691. [80]
- IJIRI, Y., AND H. A. SIMON (1964): "Business Firm Growth and Size," *American Economic Review*, 54, 77–89. [107]
- JOST, P.-J. (1996): "Limited Liability and the Requirement to Purchase Insurance," *International Review of Law and Economics*, 16, 259–276. [78]
- KALESNIK, V. (2005): "Continuous Time Partnerships With Discrete Events," Unpublished Manuscript, University of California, Los Angeles. [78]
- KAPLAN, S. N., AND P. STRÖMBERG (2003): "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts," *Review of Economic Studies*, 70, 281–315. [106]
- _____, (2004): "Characteristics, Contracts, and Actions: Evidence From Venture Capitalist Analyses," *Journal of Finance*, 59, 2177–2210. [105]
- KATZMAN, M. T. (1988): "Pollution Liability Insurance and Catastrophic Environmental Risk," *The Journal of Risk and Insurance*, 55, 75–100. [74]
- KYDLAND, F. E., AND E. C. PRESCOTT (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370. [75,79]
- LAFFONT, J.-J., AND D. MARTIMORT (2002): *The Theory of Incentives: The Principal-Agent Model*. Princeton: Princeton University Press. [86]
- LEPLAT, J., AND J. RASMUSSEN (1984): "Analysis of Human Errors in Industrial Incidents and Accidents for Improvements of Work Safety," *Accident Analysis and Prevention*, 16, 77–88. [73]
- LUCAS, R. E., JR. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508–523. [107]
- LUTTMER, E. G. J. (2007): "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, 122, 1103–1144. [107]
- _____, (2008): "On the Mechanics of Firm Growth," Working Paper 657, Federal Reserve Bank of Minneapolis. [107]
- MYERSON, R. B. (2008): "Leadership, Trust, and Power: Dynamic Moral Hazard in High Office," Unpublished Manuscript, University of Chicago. [78]
- PHELAN, C., AND R. M. TOWNSEND (1991): "Private Information and Aggregate Behaviour: Computing Multi-Period, Information-Constrained Optima," *Review of Economic Studies*, 58, 853–881. [76,81]
- POLBORN, M. K. (1998): "Mandatory Insurance and the Judgment-Proof Problem," *International Review of Law and Economics*, 18, 141–146. [78]
- QUADRINI, V. (2004): "Investment and Liquidation in Renegotiation-Proof Contracts With Moral Hazard," *Journal of Monetary Economics*, 51, 713–751. [82]
- ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53, 69–76. [75,81]
- SANNIKOV, Y. (2005): "Agency Problems, Screening and Increasing Credit Lines," Unpublished Manuscript, University of California, Berkeley. [78]
- _____, (2008): "A Continuous-Time Version of the Principal-Agent Problem," *Review of Economic Studies*, 75, 957–984. [77,81,83,85,96,108]
- SANNIKOV, Y., AND A. SKRZYPACZ (2010): "The Role of Information in Repeated Games with Frequent Actions," *Econometrica* (forthcoming). [78]
- SHAPIRO, C., AND J. E. STIGLITZ (1984): "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74, 433–444. [75]
- SHAVELL, S. (1984): "Liability for Harm versus Regulation of Safety," *Journal of Legal Studies*, 13, 357–374. [74]
- _____, (1986): "The Judgment Proof Problem," *International Review of Law and Economics*, 6, 45–58. [74,78]
- _____, (2000): "On the Social Function and the Regulation of Liability Insurance," *Geneva Papers on Risk and Insurance*, 25, 166–179. [78]
- SIMON, H. A., AND C. P. BONINI (1958): "The Size Distribution of Business Firms," *American Economic Review*, 48, 607–617. [107]

- SPEAR, S. E., AND S. SRIVASTAVA (1987): “On Repeated Moral Hazard With Discounting,” *Review of Economic Studies*, 54, 599–617. [76,81]
- SRAER, D., AND D. THESMAR (2007): “Performance and Behavior of Family Firms: Evidence From the French Stock Market,” *Journal of the European Economic Association*, 5, 709–751. [79]
- STOUT, W. F. (1974): *Almost Sure Convergence*. New York, San Francisco, London: Academic Press. [97,111,112]
- THOMAS, J., AND T. WORRALL (1990): “Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem,” *Journal of Economic Theory*, 51, 367–390. [76,77,101]

Toulouse School of Economics (CNRS, GREMAQ, IDEI), Université Toulouse 1, 21 Allée de Brienne, 31000 Toulouse, France; biais@cict.fr;

Toulouse School of Economics (CNRS, GREMAQ, IDEI), Université Toulouse 1, 21 Allée de Brienne, 31000 Toulouse, France; mariotti@cict.fr;

Toulouse School of Economics (GREMAQ, IDEI), Université Toulouse 1, 21 Allée de Brienne, 31000 Toulouse, France; rochet@cict.fr,
and

Toulouse School of Economics (CRM, IDEI), Université Toulouse 1, 21 Allée de Brienne, 31000 Toulouse, France; stephane.villeneuve@univ-tlse1.fr.

Manuscript received July, 2007; final revision received September, 2009.

INFERENCE FOR PARAMETERS DEFINED BY MOMENT INEQUALITIES USING GENERALIZED MOMENT SELECTION

BY DONALD W. K. ANDREWS AND GUSTAVO SOARES¹

The topic of this paper is inference in models in which parameters are defined by moment inequalities and/or equalities. The parameters may or may not be identified. This paper introduces a new class of confidence sets and tests based on generalized moment selection (GMS). GMS procedures are shown to have correct asymptotic size in a uniform sense and are shown not to be asymptotically conservative.

The power of GMS tests is compared to that of subsampling, m out of n bootstrap, and “plug-in asymptotic” (PA) tests. The latter three procedures are the only general procedures in the literature that have been shown to have correct asymptotic size (in a uniform sense) for the moment inequality/equality model. GMS tests are shown to have asymptotic power that dominates that of subsampling, m out of n bootstrap, and PA tests. Subsampling and m out of n bootstrap tests are shown to have asymptotic power that dominates that of PA tests.

KEYWORDS: Asymptotic size, asymptotic power, confidence set, exact size, generalized moment selection, m out of n bootstrap, subsampling, moment inequalities, moment selection, test.

1. INTRODUCTION

THIS PAPER CONSIDERS INFERENCE in models in which parameters are defined by moment inequalities and/or equalities. The parameters need not be identified. Numerous examples of such models are now available in the literature, for example, see Manski and Tamer (2002), Imbens and Manski (2004), Andrews, Berry, and Jia (2004), Pakes, Porter, Ishii, and Ho (2004), Moon and Schorfheide (2006), Chernozhukov, Hong, and Tamer (2007) (CHT), and Ciliberto and Tamer (2009).

The paper introduces confidence sets (CS’s) based on a method called *generalized moment selection* (GMS). The CS’s considered in the paper are obtained by inverting tests that are of an Anderson–Rubin type. This method was first considered in the moment inequality context by CHT.

In this paper, we analyze GMS critical values. We note that the choice of critical value is much more important in moment inequality/equality models than in most models. In most models, the choice of critical value does not affect the first-order asymptotic properties of a test or CS. In the moment inequality/equality model, however, it does, and the effect can be large.

The results of the paper hold for a broad class of test statistics including modified method of moments (MMM) statistics, Gaussian quasilelihood ratio (QLR) statistics, generalized empirical likelihood ratio (GEL) statistics,

¹Andrews gratefully acknowledges the research support of the National Science Foundation via Grants SES-0417911 and SES-0751517. The authors thank three referees, a co-editor, Ivan Canay, Victor Chernozhukov, Jörg Stoye, and especially Patrik Guggenberger for comments.

and a variety of others. The results apply to CS's for the true parameter, as in Imbens and Manski (2004), rather than for the identified set (i.e., the set of points that are consistent with the population moment inequalities/equalities), as in CHT. We focus on CS's for the true parameter because answers to policy questions typically depend on the true parameter rather than on the identified set.

Subsampling CS's for the moment inequality/equality model are considered in CHT, Andrews and Guggenberger (2009b) (hereafter AG4), and Romano and Shaikh (2008, 2010). “Plug-in asymptotic” (PA) CS's are widely used in the literature on multivariate one-sided tests and CS's. They are considered in the moment inequality/equality model in AG4 and a variant of them is considered in Rosen (2008).

Here we introduce GMS critical values. Briefly, the idea behind GMS critical values is as follows. The $1 - \alpha$ quantile of the finite-sample null distribution of a typical test statistic depends heavily on the extent to which the moment inequalities are binding (i.e., are close to being equalities). In consequence, the asymptotic null distribution of the test statistic under a suitable drifting sequence of parameters depends heavily on a nuisance parameter $h = (h_1, \dots, h_p)'$, whose j th element $h_j \in [0, \infty]$ indexes the extent to which the j th moment inequality is binding. For a suitable class of test statistics, the larger is h , the smaller is the asymptotic null distribution in a stochastic sense. This is key for obtaining procedures that are uniformly asymptotically valid.

The parameter h cannot be estimated consistently in a uniform sense, but one can use the sample moment inequalities to estimate or test how close h is to 0_p . A computationally simple procedure is to use inequality-by-inequality t -tests to test whether $h_j = 0$ for $j = 1, \dots, p$. If a test rejects $h_j = 0$, then that inequality is removed from the asymptotic null distribution that is used to calculate the critical value. The t -tests have to be designed so that the probability of incorrectly omitting a moment inequality from the asymptotic distribution is asymptotically negligible. Continuous/smooth versions of such procedures can be employed in which moment inequalities are not “in or out,” but are “more in or more out” depending on the magnitude of the t statistics.

Another type of GMS procedure is based on a modified moment selection criterion (MMSC), which is an information-type criterion analogous to the Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan–Quinn information criterion (HQIC) model selection criteria; see Hannan and Quinn (1979) regarding HQIC. Andrews (1999a) used an information-type moment selection criterion to determine which moment equalities are invalid in a standard moment equality model. Here we employ one-sided versions of such procedures to determine which moment inequalities are not binding. In contrast to inequality-by-inequality t -tests, the MMSC jointly determines which moment inequalities to select and takes account of correlations between sample moment inequalities.

The results of the paper cover a broad class of GMS procedures that includes all of those discussed above. Section 4.2 gives a step-by-step description of how to calculate GMS procedures.

In this paper, we show that GMS critical values yield uniformly asymptotically valid CS's and tests. These results hold for both independent and identically distributed (i.i.d.) and dependent observations. We also show that GMS procedures are not asymptotically conservative. They are asymptotically non-similar, but are less so than subsampling and PA procedures.

The volume of a CS that is based on inverting a test depends on the power of the test; see Pratt (1961). Thus, power is important for both tests and CS's. We determine and compare the power of GMS, subsampling, and PA tests. CHT and Beresteanu and Molinari (2008) also provided some asymptotic local power results for testing procedures in models with partially identified parameters. Otsu (2006), Bugni (2007a, 2007b), and Canay (2007) considered asymptotic power against fixed alternatives. Tests typically have asymptotic power equal to 1 against such alternatives.

We investigate the asymptotic power of GMS, subsampling, and PA tests for local and nonlocal alternatives. Such alternatives are more complicated in the moment inequality/equality model than in most models. The reason is that some inequalities may be violated while others may be satisfied as equalities, as inequalities that are relatively close to being equalities, and/or as inequalities that are far from being equalities. Furthermore, depending upon the particular alternative hypothesis scenario considered, the data-dependent critical values behave differently asymptotically. We derive the asymptotic power of the tests under the complete range of alternatives from $n^{-1/2}$ local, to more distant local, through to fixed alternatives for each of the different moment inequalities and equalities that appear in the model.

We show that (under reasonable assumptions) GMS tests are as powerful asymptotically as subsampling and PA tests with strictly greater power in certain scenarios. The asymptotic power differences can be substantial. Furthermore, we show that subsampling tests are as powerful asymptotically as PA tests with greater power in certain scenarios. m out of n bootstrap tests have the same asymptotic properties as subsampling tests (at least in i.i.d. scenarios when $m = o(n^{1/2})$; see Politis, Romano, and Wolf (1999, p. 48)).

GMS tests are shown to be strictly more powerful asymptotically than subsampling tests whenever (i) at least one population moment inequality is satisfied under the alternative and differs from an equality by an amount that is $O(b^{-1/2})$ and is larger than $O(\kappa_n n^{-1/2})$, where b is the subsample size and κ_n is a GMS constant such as $\kappa_n = (\ln n)^{1/2}$, and (ii) the GMS and subsampling critical values do not have the degenerate probability limit of 0. Note that good choices of b and κ_n in terms of size and power satisfy $b \approx n^\eta$ for some $\eta \in (0, 1)$ and $\kappa_n = o(n^\varepsilon) \forall \varepsilon > 0$, so that condition (i) holds.

GMS and subsampling tests are shown to be strictly more powerful asymptotically than PA tests whenever at least one population moment inequality is

satisfied under the alternative, and differs from an equality by an amount that is larger than $O(\kappa_n n^{-1/2})$ for GMS tests and is larger than $o(b^{-1/2})$ for subsampling tests.

The paper shows that (pure) generalized empirical likelihood (GEL) tests, which are based on fixed critical values, are dominated in terms of asymptotic power by GMS and subsampling tests based on a QLR or GEL test statistic.

The paper reports some finite-sample size and power results obtained via simulation for tests based on GMS, subsampling, and PA critical values. The GMS critical values are found to deliver very good null rejection probabilities and power in the scenarios considered. They are found to outperform the subsampling and PA critical values by a substantial margin, especially for larger values of p , the number of moment inequalities. Additional simulation results are reported in the Supplemental Material (Andrews and Soares (2010)) and in Andrews and Jia (2008).

The determination of a best test statistic/GMS procedure is difficult because uniformly best choices do not exist. Nevertheless, it is possible to make comparisons based on all-around performance. Doing so is beyond the scope of the present paper and is the subject of research reported in Andrews and Jia (2008). In the latter paper, the QLR test statistic combined with the GMS procedure based on t -tests is found to work very well in practice and hence is recommended.

Bootstrap versions of GMS critical values are obtained by replacing the multivariate normal random vector that appears in the asymptotic distribution by a bootstrap distribution based on the recentered sample moments. The block bootstrap can be employed in time series contexts. GMS bootstrap critical values, however, do not yield higher-order improvements, because the asymptotic null distribution is not asymptotically pivotal. Bugni (2007a, 2007b) and Canay (2007) considered particular types of bootstrap GMS critical values. Andrews and Jia (2008) found that the bootstrap version of the GMS critical values outperforms the asymptotic normal version in i.i.d. scenarios and hence is recommended.

The paper introduces GMS model specification tests based on the GMS tests discussed above. These tests are shown to be uniformly asymptotically valid. They can be asymptotically conservative.

We now discuss related literature. Bugni (2007a, 2007b) showed that a particular type of GMS test (based on $\varphi^{(1)}$ defined below) has more accurate pointwise asymptotic size than a (recentered) subsampling test. Such results should extend to all GMS tests and to asymptotic size defined in a uniform sense. Given that they do, GMS tests have both asymptotic power and size advantages over subsampling tests. The relatively low accuracy of the size of subsampling tests and CS's in many models is well known in the literature. We are not aware of any other papers or scenarios where the asymptotic power of subsampling tests has been shown to be dominated by other procedures.

GMS critical values based on $\varphi^{(1)}$, defined below, are a variant of the Wald test procedure in Andrews (1999b, Sec. 6.4; 2000, Sec. 4) for the problem of inference when a parameter is on or near a boundary. The 2003 working paper version of CHT discusses a bootstrap version of the GMS method based on $\varphi^{(1)}$ in the context of the interval outcome model. Soares (2005) analyzed the properties of GMS critical values based on $\varphi^{(1)}$ and introduced GMS critical values based on the function $\varphi^{(5)}$. The present paper supplants Soares (2005). CHT mentioned critical values of GMS type based on $\varphi^{(1)}$; see their Remark 4.5. GMS critical values of types $\varphi^{(2)}-\varphi^{(4)}$ were considered by the authors in January 2007. Galichon and Henry (2009) consider a set selection method that is analogous to GMS based on $\varphi^{(1)}$. Bugni (2007a, 2007b) considered GMS critical values based on $\varphi^{(1)}$. His work was done independently of, but subsequently to, Soares (2005). Canay (2007) independently considered GMS critical values based on $\varphi^{(3)}$. Bugni (2007a, 2007b) and Canay (2007) focused on bootstrap versions of the GMS critical values.

Other papers in the literature that consider inference with moment inequalities include Andrews, Berry, and Jia (2004), Pakes et al. (2004), Romano and Shaikh (2008, 2010), Moon and Schorfheide (2006), Otsu (2006), Woutersen (2006), Bontemps, Magnac, and Maurin (2007), Bugni (2007a, 2007b), Canay (2007), CHT, Fan and Park (2007), Beresteanu, Molchanov, and Molinari (2008), Beresteanu and Molinari (2008), Guggenberger, Hahn, and Kim (2008), Rosen (2008), AG4, Andrews and Han (2009), and Stoye (2009).

The remainder of the paper is organized as follows. Section 2 describes the moment inequality/equality model. Section 3 introduces the class of test statistics that is considered and states assumptions. Section 4 introduces the class of GMS CS's. Section 5 introduces GMS model specification tests. Sections 6 and 7 define subsampling CS's and PA CS's, respectively. Section 8 determines and compares the $n^{-1/2}$ -local alternative power of GMS, subsampling, and PA tests. Section 9 considers the power of these tests against more distant alternatives. Section 10 discusses extensions to GEL test statistics and preliminary estimation of identified parameters. Section 11 provides the simulation results. The Appendix contains some assumptions concerning the test statistics considered, an alternative parametrization of the moment inequality/equality model, and the treatment of dependent observations. The proofs of all results are given in the Supplemental Material (Andrews and Soares (2010)).

For notational simplicity, throughout the paper we write partitioned column vectors as $h = (h_1, h_2)$, rather than $h = (h'_1, h'_2)'$. Let $R_+ = \{x \in R : x \geq 0\}$, $R_{+\infty} = R_+ \cup \{+\infty\}$, $R_{[+\infty]} = R \cup \{+\infty\}$, $R_{[\pm\infty]} = R \cup \{\pm\infty\}$, $K^p = K \times \cdots \times K$ (with p copies) for any set K , and $\infty^p = (+\infty, \dots, +\infty)'$ (with p copies). All limits are as $n \rightarrow \infty$ unless specified otherwise. Let pd abbreviate positive definite. Let $\text{cl}(\Psi)$ denote the closure of a set Ψ . We let AG1 abbreviate Andrews and Guggenberger (2010b).

2. MOMENT INEQUALITY MODEL

We now introduce the moment inequality/equality model. The true value θ_0 ($\in \Theta \subset R^d$) is assumed to satisfy the moment conditions:

$$(2.1) \quad E_{F_0} m_j(W_i, \theta_0) \begin{cases} \geq 0 & \text{for } j = 1, \dots, p, \\ = 0 & \text{for } j = p + 1, \dots, p + v, \end{cases}$$

where $\{m_j(\cdot, \theta) : j = 1, \dots, k\}$ are known real-valued moment functions, $k = p + v$, and $\{W_i : i \geq 1\}$ are i.i.d. or stationary random vectors with joint distribution F_0 . The observed sample is $\{W_i : i \leq n\}$. A key feature of the model is that the true value θ_0 is not necessarily identified. That is, knowledge of $E_{F_0} m_j(W_i, \theta)$ for $j = 1, \dots, k$ for all $\theta \in \Theta$ does not necessarily imply knowledge of θ_0 . In fact, even knowledge of F_0 does not necessarily imply knowledge of the true value θ_0 . More information than is available in $\{W_i : i \leq n\}$ may be needed to identify the true parameter θ_0 .

Note that both moment inequalities and moment equalities arise in the entry game models considered in Ciliberto and Tamer (2009) and Andrews, Berry, and Jia (2004), and in the macroeconomic model in Moon and Schorfheide (2006). There are numerous models where only moment inequalities arise; for example, see Manski and Tamer (2002) and Imbens and Manski (2004). There are also unidentified models in which only moment equalities arise; see CHT for references.

We are interested in CS's for the true value θ_0 .

Generic values of the parameters are denoted (θ, F) . For the case of i.i.d. observations, the parameter space \mathcal{F} for (θ, F) is the set of all (θ, F) that satisfy

- $$(2.2) \quad \begin{aligned} \text{(i)} \quad & \theta \in \Theta, \\ \text{(ii)} \quad & E_F m_j(W_i, \theta) \geq 0 \quad \text{for } j = 1, \dots, p, \\ \text{(iii)} \quad & E_F m_j(W_i, \theta) = 0 \quad \text{for } j = p + 1, \dots, k, \\ \text{(iv)} \quad & \{W_i : i \geq 1\} \text{ are i.i.d. under } F, \\ \text{(v)} \quad & \sigma_{F,j}^2(\theta) = \text{Var}_F(m_j(W_i, \theta)) \in (0, \infty) \quad \text{for } j = 1, \dots, k, \\ \text{(vi)} \quad & \text{Corr}_F(m(W_i, \theta)) \in \Psi, \\ \text{(vii)} \quad & E_F |m_j(W_i, \theta)/\sigma_{F,j}(\theta)|^{2+\delta} \leq M \quad \text{for } j = 1, \dots, k, \end{aligned}$$

where Ψ is a set of $k \times k$ correlation matrices specified below, and $M < \infty$ and $\delta > 0$ are constants. For expositional convenience, we specify \mathcal{F} for dependent observations in the Appendix, Section A.2.

We consider a confidence set obtained by inverting a test. The test is based on a test statistic $T_n(\theta_0)$ for testing $H_0: \theta = \theta_0$. The nominal level $1 - \alpha$ CS for θ is

$$(2.3) \quad \text{CS}_n = \{\theta \in \Theta : T_n(\theta) \leq c_{1-\alpha}(\theta)\},$$

where $c_{1-\alpha}(\theta)$ is a critical value.² We consider GMS, subsampling, and plug-in asymptotic critical values. These are data-dependent critical values and their probability limits, when they exist, typically depend on the true distribution generating the data.

The exact and asymptotic confidence sizes of CS_n are

$$(2.4) \quad \text{ExCS}_n = \inf_{(\theta, F) \in \mathcal{F}} P_F(T_n(\theta) \leq c_{1-\alpha}(\theta)) \quad \text{and} \quad \text{AsyCS} = \liminf_{n \rightarrow \infty} \text{ExCS}_n,$$

respectively. The definition of AsyCS takes the $\inf_{(\theta, F) \in \mathcal{F}}$ before the $\lim_{n \rightarrow \infty}$. This builds uniformity over (θ, F) into the definition of AsyCS. Uniformity is required for the asymptotic size to give a good approximation to the finite-sample size of CS's. Andrews and Guggenberger (2009a, 2010a, 2010b) and Mikusheva (2007) showed that when a test statistic has a discontinuity in its limit distribution, as occurs in the moment inequality/equality model, pointwise asymptotics (in which one takes the lim before the inf) can be very misleading in some models. See AG4 for further discussion.

The exact and asymptotic maximum coverage probabilities are

$$(2.5) \quad \text{ExMaxCP}_n = \sup_{(\theta, F) \in \mathcal{F}} P_F(T_n(\theta) \leq c_{1-\alpha}(\theta)),$$

$$\text{AsyMaxCP} = \limsup_{n \rightarrow \infty} \text{ExMaxCP}_n,$$

respectively. The magnitude of asymptotic nonsimilarity of the CS is measured by the difference $\text{AsyMaxCP} - \text{AsyCS}$.

If interest is in a subvector, say β , of θ , then confidence sets for β can be constructed via projection. That is, one takes the CS to be $\{\beta \in R^{d_\beta} : \text{for some } \lambda \in R^{d-d_\beta}, (\beta', \lambda')' \in \text{CS}_n\}$. By a standard argument, if CS_n is a CS for θ with asymptotic size greater than or equal to $1 - \alpha$, then this CS for β has the same property. Typically, however, a CS for β constructed in this way has an asymptotic size that is strictly greater than $1 - \alpha$, which implies that it is asymptotically conservative.

3. TEST STATISTICS

In this section, we define the main class of test statistics $T_n(\theta)$ that we consider. GEL statistics are discussed in Section 10 below.

²It is important that the inequality in the definition of CS_n is less than or equal to, not less than. When θ is in the interior of the identified set, it is often the case that $T_n(\theta) = 0$ and $c_{1-\alpha}(\theta) = 0$.

3.1. Form of the Test Statistics

The sample moment functions are

$$(3.1) \quad \bar{m}_n(\theta) = (\bar{m}_{n,1}(\theta), \dots, \bar{m}_{n,k}(\theta))', \quad \text{where}$$

$$\bar{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^n m_j(W_i, \theta) \quad \text{for } j = 1, \dots, k.$$

Let $\widehat{\Sigma}_n(\theta)$ be an estimator of the asymptotic variance, $\Sigma(\theta)$, of $n^{1/2}\bar{m}_n(\theta)$. When the observations are i.i.d., we take

$$(3.2) \quad \widehat{\Sigma}_n(\theta) = n^{-1} \sum_{i=1}^n (m(W_i, \theta) - \bar{m}_n(\theta))(m(W_i, \theta) - \bar{m}_n(\theta))', \quad \text{where}$$

$$m(W_i, \theta) = (m_1(W_i, \theta), \dots, m_k(W_i, \theta))'.$$

With temporally dependent observations, a different definition of $\widehat{\Sigma}_n(\theta)$ often is required. For example, a heteroskedasticity and autocorrelation consistent (HAC) estimator may be required.

The statistic $T_n(\theta)$ is defined to be of the form

$$(3.3) \quad T_n(\theta) = S(n^{1/2}\bar{m}_n(\theta), \widehat{\Sigma}_n(\theta)),$$

where S is a real function on $R_{[+\infty]}^p \times R^v \times \mathcal{V}_{k \times k}$, where $\mathcal{V}_{k \times k}$ is the space of $k \times k$ variance matrices. (The set $R_{[+\infty]}^p \times R^v$ contains k -vectors whose first p elements are either real or $+\infty$ and whose last v elements are real.) The function S is required to satisfy Assumptions 1–6 stated below. We now give several examples of functions that do so.

First, consider the MMM test function $S = S_1$ defined by

$$(3.4) \quad S_1(m, \Sigma) = \sum_{j=1}^p [m_j/\sigma_j]_-^2 + \sum_{j=p+1}^{p+v} (m_j/\sigma_j)^2, \quad \text{where}$$

$$[x]_- = \begin{cases} x, & \text{if } x < 0, \\ 0, & \text{if } x \geq 0, \end{cases} \quad m = (m_1, \dots, m_k)',$$

and σ_j^2 is the j th diagonal element of Σ . With the function S_1 , the parameter space Ψ for the correlation matrices in condition (vi) of (2.2) is $\Psi = \Psi_1$, where Ψ_1 contains all $k \times k$ correlation matrices.³ The function S_1 yields a test statistic that gives positive weight to moment inequalities only when they are violated.

³Note that with temporally dependent observations, Ψ is the parameter space for the limiting correlation matrix, $\lim_{n \rightarrow \infty} \text{Corr}_F(n^{1/2}\bar{m}_n(\theta))$.

This type of statistic has been considered in Romano and Shaikh (2008, 2010), Soares (2005), CHT, and AG4. Note that S_1 normalizes the moment functions by dividing by σ_j in each summand. One could consider a function without this normalization, as in Pakes et al. (2004) and Romano and Shaikh (2008, 2010), but the resulting statistic is not invariant to rescaling of the moment conditions and, hence, is not likely to have good properties in terms of the volume of its CS. We use the function S_1 in the simulations reported in Section 11 below.

Second, we consider a QLR test function defined by

$$(3.5) \quad S_2(m, \Sigma) = \inf_{t=(t_1, 0_v) : t_1 \in R_{+, \infty}^p} (m - t)' \Sigma^{-1} (m - t).$$

With this function, the parameter space Ψ in (2.2) is $\Psi = \Psi_2$, where Ψ_2 contains all $k \times k$ correlation matrices whose determinant is greater than or equal to ε for some $\varepsilon > 0$.^{4,5} This type of statistic has been considered in many papers on tests of inequality constraints (e.g., see Kudo (1963) and Silvapulle and Sen (2005, Sec. 3.8)), as well as papers in the moment inequality literature (see Rosen (2008)). We note that GEL test statistics behave asymptotically (to the first order) under the null and alternative hypotheses like the statistic $T_n(\theta)$ based on S_2 ; see Section 10 below and AG4.

The requirement that $\Psi = \Psi_2$ for S_2 is restrictive in some cases, such as when two moment inequalities have correlation equal to 1 in absolute value. In such cases, one can alter the definition of S_2 in (3.5) by replacing Σ by $\Sigma + \varepsilon \text{Diag}(\Sigma)$ for some $\varepsilon > 0$, where $\text{Diag}(\Sigma)$ denotes the $k \times k$ diagonal matrix whose diagonal elements equal those of Σ . With this alteration, one can take $\Psi = \Psi_1$.

For a test with power directed against alternatives with $p_1 (< p)$ moment inequalities violated, the following function is suitable:

$$(3.6) \quad S_3(m, \Sigma) = \sum_{j=1}^{p_1} [m_{(j)} / \sigma_{(j)}]_-^2 + \sum_{j=p+1}^{p+v} (m_j / \sigma_j)^2,$$

where $[m_{(j)} / \sigma_{(j)}]_-^2$ denotes the j th largest value among $\{[m_\ell / \sigma_\ell]^2 : \ell = 1, \dots, p\}$ and $p_1 < p$ is some specified integer. The function S_3 satisfies (2.2) with $\Psi = \Psi_1$. The function S_3 is considered in Andrews and Jia (2008).

⁴The condition that $\Psi = \Psi_2$ for the function S_2 is used in the proofs of various asymptotic results. This condition may be just a consequence of the method of proof. It may not actually be needed.

⁵The definition of $S_2(m, \Sigma)$ takes the infimum over $t_1 \in R_{+, \infty}^p$, rather than over $t_1 \in R_+^p$. For calculation of the test statistic based on S_2 , using the latter gives an equivalent value. To obtain the correct asymptotic distribution, however, the former definition is required because it leads to continuity at infinity of S_2 when some elements of m may equal infinity. For example, suppose $k = p = 1$. In this case, when $m \in R_+$, $\inf_{t_1 \in R_{+, \infty}} (m - t_1)^2 = \inf_{t_1 \in R_+} (m - t_1)^2 = 0$. However, when $m = \infty$, $\inf_{t_1 \in R_{+, \infty}} (m - t_1)^2 = 0$, but $\inf_{t_1 \in R_+} (m - t_1)^2 = \infty$.

Other examples of test functions S that satisfy Assumptions 1–6 are variations of S_1 and S_3 with the step function $[x]_-$ replaced by a smooth function, with the square replaced by the absolute value to a different positive power (such as 1), or with weights added.

It is difficult to compare the performance of one test function S with another function without specifying the critical values to be used. Most critical values, such as the GMS, subsampling, and PA critical values considered here, are data dependent and have limits as $n \rightarrow \infty$ that depend on the distribution of the observations. For a given test function S , a different test is obtained for each type of critical value employed and the differences do not vanish asymptotically. The relative performances of different functions S are considered elsewhere; see Andrews and Jia (2008).

3.2. Test Statistic Assumptions

Next, we state the most important assumptions concerning the function S , namely, Assumptions 1, 3, and 6. For ease of reading, technical assumptions (mostly continuity and strictly-increasing assumptions on asymptotic distribution functions (df's)), namely, Assumptions 2, 4, 5, and 7, are stated in the Appendix. We show below that the functions S_1 – S_3 automatically satisfy Assumptions 1–6. Assumption 7 is not restrictive.

Let $B \subset R^w$. We say that a real function G on $R_{[+\infty]}^p \times B$ is continuous at $x \in R_{[+\infty]}^p \times B$ if $y \rightarrow x$ for $y \in R_{[+\infty]}^p \times B$ implies that $G(y) \rightarrow G(x)$. In the assumptions below, the set Ψ is as in condition (vi) of (2.2).⁶ For p -vectors m_1 and m_1^* , $m_1 < m_1^*$ means that $m_1 \leq m_1^*$ and at least one inequality in the p -vector of inequalities holds strictly.

- ASSUMPTION 1: (a) $S((m_1, m_2), \Sigma)$ is nonincreasing in m_1 for all $m_1 \in R^p$, $m_2 \in R^v$, and variance matrices $\Sigma \in R^{k \times k}$.
- (b) $S(m, \Sigma) = S(Dm, D\Sigma D)$ for all $m \in R^k$, $\Sigma \in R^{k \times k}$, and pd diagonal $D \in R^{k \times k}$.
- (c) $S(m, \Omega) \geq 0$ for all $m \in R^k$ and $\Omega \in \Psi$.
- (d) $S(m, \Omega)$ is continuous at all $m \in R_{[+\infty]}^p \times R^v$ and $\Omega \in \Psi$.⁷

ASSUMPTION 3: $S(m, \Omega) > 0$ if and only if $m_j < 0$ for some $j = 1, \dots, p$ or $m_j \neq 0$ for some $j = p+1, \dots, k$, where $m = (m_1, \dots, m_k)'$ and $\Omega \in \Psi$.

ASSUMPTION 6: For some $\chi > 0$, $S(am, \Omega) = a^\chi S(m, \Omega)$ for all scalars $a > 0$, $m \in R^k$, and $\Omega \in \Psi$.

⁶For dependent observations, Ψ is as in condition (v) of (A.2) in the Appendix.

⁷In Assumption 1(d) (and in Assumption 4(b) in the Appendix), $S(m, \Omega)$ and $c(\Omega, 1 - \alpha)$ are viewed as functions defined on the space of all correlation matrices Ψ_1 . By definition, $c(\Omega, 1 - \alpha)$ is continuous in Ω uniformly for $\Omega \in \Psi$ if for all $\eta > 0$, there exists $\delta > 0$ such that whenever $\|\Omega^* - \Omega\| < \delta$ for $\Omega^* \in \Psi_1$ and $\Omega \in \Psi$, we have $|c_{\Omega^*}(1 - \alpha) - c_\Omega(1 - \alpha)| < \eta$.

Assumptions 1–6 are shown in Lemma 1 below not to be restrictive. Assumption 1(a) is the key assumption that is needed to ensure that GMS and subsampling CS’s have correct asymptotic size. Assumption 1(b) is a natural assumption that specifies that the test statistic is invariant to the scale of each sample moment. Assumption 1(b) and 1(d) are conditions that enable one to determine the asymptotic properties of $T_n(\theta)$. Assumption 1(c) normalizes the test statistic to be nonnegative.

Assumption 3 implies that a positive value of $S(m, \Omega)$ only occurs if some inequality or equality is violated. Assumption 3 implies that $S(\infty^p, \Sigma) = 0$ when $v = 0$. Assumption 6 requires S to be homogeneous of degree $\chi > 0$ in m . This is used to show that the test based on S has asymptotic power equal to 1 against fixed alternatives.

LEMMA 1: *The functions $S_1(m, \Sigma)$ – $S_3(m, \Sigma)$ satisfy Assumptions 1–6 with $\Psi = \Psi_1$ for $S_1(m, \Sigma)$ and $S_3(m, \Sigma)$, and with $\Psi = \Psi_2$ for $S_2(m, \Sigma)$.*

4. GENERALIZED MOMENT SELECTION

This section is concerned with critical values for use with the test statistics introduced in Section 3.

4.1. Description of the GMS Method

We start by motivating the GMS method. Consider the null hypothesis $H_0: \theta = \theta_0$. The finite-sample null distribution of $T_n(\theta_0)$ depends continuously on the degree of *slackness* of the moment inequalities. That is, it depends on how much greater than zero is $E_F m_j(W_i, \theta_0)$ for $j = 1, \dots, p$. Under Assumption 1(a), the least favorable case (at least asymptotically) can be shown to be the case where there is no slackness—each of the moments is zero. That is, the distribution of $T_n(\theta_0)$ is stochastically largest over distributions in the null hypothesis when the inequality moments equal zero. One way to construct a critical value for $T_n(\theta_0)$, then, is to take the $1 - \alpha$ quantile of the distribution (or asymptotic distribution) of $T_n(\theta_0)$ when the inequality moments all equal zero. This yields a test with correct (asymptotic) size, but its power properties are poor against many alternatives of interest.

The reason for its poor power is that the least favorable critical value is relatively large. This is especially true if the number of moment inequalities, p , is large. For example, consider power against an alternative for which only the first moment inequality is violated (i.e., $E_F m_1(W_i, \theta_0) < 0$) and the last $p - 1$ moment inequalities are satisfied by a wide margin (i.e., $E_F m_j(W_i, \theta_0) \gg 0$ for $j = 2, \dots, p$). Then the last $p - 1$ moment inequalities have little or no effect on the value of the test statistic $T_n(\theta_0)$. (This holds for typical test statistics and is implied by Assumption 3.) Yet, the critical value *does* depend on the existence of the last $p - 1$ moment inequalities and is much larger than it would

be if these moment inequalities were absent. In consequence, the test has significantly lower power than if the last $p - 1$ moment inequalities were absent.

The idea behind *generalized moment selection* is to use the data to determine whether a given moment inequality is satisfied and is far from being an equality, and if so to take the critical value to be smaller than otherwise—both under the null and under the alternative. Of course, in doing so, one has to make sure that the (asymptotic) size of the resulting test is correct. We use the sample moment functions to estimate or test whether the population moment inequalities are close to or far from being equalities.

Using Assumption 1(b), we can write

$$(4.1) \quad \begin{aligned} T_n(\theta) &= S(n^{1/2}\bar{m}_n(\theta), \hat{\Sigma}_n(\theta)) \\ &= S(\hat{D}_n^{-1/2}(\theta)n^{1/2}\bar{m}_n(\theta), \hat{\Omega}_n(\theta)), \quad \text{where} \\ \hat{D}_n(\theta) &= \text{Diag}(\hat{\Sigma}_n(\theta)) \quad \text{and} \quad \hat{\Omega}_n(\theta) = \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta). \end{aligned}$$

Thus, the test statistic $T_n(\theta)$ depends only on the normalized sample moments and the sample correlation matrix. Under an appropriate sequence of null distributions $\{F_n : n \geq 1\}$, the asymptotic null distribution of $T_n(\theta_0)$ is that of

$$(4.2) \quad S(\Omega_0^{1/2}Z^* + (h_1, 0_v), \Omega_0), \quad \text{where} \quad Z^* \sim N(0_k, I_k),$$

$h_1 \in R_{+, \infty}^p$, and Ω_0 is a $k \times k$ correlation matrix. This result holds by (4.1), the central limit theorem, and the convergence in probability of the sample correlation matrix; see the proof of Theorem 1 of AG4. The p -vector h_1 is the limit of $(n^{1/2}E_{F_n}m_1(W_i, \theta_0)/\sigma_{F_n, 1}(\theta_0), \dots, n^{1/2}E_{F_n}m_p(W_i, \theta_0)/\sigma_{F_n, p}(\theta_0))'$ under the null distributions $\{F_n : n \geq 1\}$. By considering suitable sequences of distributions F_n that depend on n , rather than a fixed distribution F , we obtain an asymptotic distribution that depends continuously on the degree of slackness of the population moment inequalities via the parameter h_1 ($\geq 0_p$). This reflects the finite-sample situation.

Note that the correlation matrix Ω_0 can be consistently estimated, but the $n^{-1/2}$ -local asymptotic mean parameter h_1 cannot be (uniformly) consistently estimated. It is the latter property that makes it challenging to determine a critical value that yields a test with correct asymptotic size and good power properties.

The GMS critical value is defined to be the $1 - \alpha$ quantile of a data-dependent version of the asymptotic null distribution, $S(\Omega_0^{1/2}Z^* + (h_1, 0_v), \Omega_0)$, that replaces Ω_0 by a consistent estimator and replaces h_1 with a p -vector in $R_{+, \infty}^p$ whose value depends on a measure of the slackness of the moment inequalities. We measure the degree of slackness of the moment inequalities via

$$(4.3) \quad \xi_n(\theta) = \kappa_n^{-1}n^{1/2}\hat{D}_n^{-1/2}(\theta)\bar{m}_n(\theta) \in R^k$$

evaluated at $\theta = \theta_0$, where $\{\kappa_n : n \geq 1\}$ is a sequence of constants that diverges to infinity as $n \rightarrow \infty$. A suitable choice of κ_n is the BIC choice

$$(4.4) \quad \kappa_n = (\ln n)^{1/2}.$$

The law of the iterated logarithm choice, $\kappa_n = (2 \ln \ln n)^{1/2}$, also is possible, but the simulations reported in Section 11 below indicate that the BIC choice is preferable. CHT also suggested using the BIC value.

Let $\xi_{n,j}(\theta)$, $h_{1,j}$, and $[\Omega_0^{1/2} Z^*]_j$ denote the j th elements of $\xi_n(\theta)$, h_1 , and $\Omega_0^{1/2} Z^*$, respectively, for $j = 1, \dots, p$. When $\xi_{n,j}(\theta_0)$ is zero or close to zero, this indicates that $h_{1,j}$ is zero or fairly close to zero and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is 0. On the other hand, when $\xi_{n,j}(\theta_0)$ is large, this indicates $h_{1,j}$ is quite large (where the adjective “quite” is due to the κ_n factor) and the desired replacement of $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is ∞ .

We replace $h_{1,j}$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ by $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ for $j = 1, \dots, p$, where $\varphi_j : (R_{[+\infty]}^p \times R_{[\pm\infty]}^v) \times \Psi \rightarrow R_{[\pm\infty]}$ is a function that is chosen to deliver the properties described above. Suppose φ_j satisfies (i) $\varphi_j(\xi, \Omega) = 0$ for all $\xi = (\xi_1, \dots, \xi_k)' \in R_{[+\infty]}^p \times R_{[\pm\infty]}^v$ with $\xi_j = 0$ and all $\Omega \in \Psi$, and (ii) $\varphi_j(\xi, \Omega) \rightarrow \infty$ as $(\xi, \Omega) \rightarrow (\xi_*, \Omega_*)$ for all $\xi_* = (\xi_{*,1}, \dots, \xi_{*,k})' \in R_{[+\infty]}^p \times R_{[\pm\infty]}^v$ with $\xi_{*,j} = \infty$ and all $\Omega_* \in \Psi$, where $\xi \in R^k$ and $\Omega \in \Psi$. In this case, if $\xi_{n,j}(\theta_0) = 0$, then $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) = 0$ and $h_{1,j}$ is replaced by 0, as desired. On the other hand, if $\xi_{n,j}(\theta_0)$ is large, condition (ii) implies that $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ is large and $h_{1,j}$ is replaced by a large value, as desired, for $j = 1, \dots, p$. For $j = p+1, \dots, k$, we define $\varphi_j(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) = 0$ because no $h_{1,j}$ term appears in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$.

Examples of functions φ_j include

$$(4.5) \quad \begin{aligned} \varphi_j^{(1)}(\xi, \Omega) &= \begin{cases} 0, & \text{if } \xi_j \leq 1 \\ \infty, & \text{if } \xi_j > 1, \end{cases} & \varphi_j^{(2)}(\xi, \Omega) &= \psi(\xi_j), \\ \varphi_j^{(3)}(\xi, \Omega) &= [\xi_j]_+, & \text{and} & \varphi_j^{(4)}(\xi, \Omega) = \xi_j \end{aligned}$$

for $j = 1, \dots, p$, where ψ is defined below. Let $\varphi^{(r)}(\xi, \Omega) = (\varphi_1^{(r)}(\xi, \Omega), \dots, \varphi_p^{(r)}(\xi, \Omega), 0, \dots, 0)' \in R_{[\pm\infty]}^p \times \{0\}^v$ for $r = 1, \dots, 4$.

The function $\varphi^{(1)}$ generates a “moment selection t -test” procedure. Using $\varphi^{(1)}$, $h_{1,j}$ is replaced in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ by ∞ if $\xi_{n,j}(\theta_0) > 1$ and by 0 otherwise. Note that $\xi_{n,j}(\theta_0) > 1$ is equivalent to

$$(4.6) \quad \frac{n^{1/2} \bar{m}_{n,j}(\theta_0)}{\widehat{\sigma}_{n,j}(\theta_0)} > \kappa_n,$$

where $\widehat{\sigma}_{n,j}^2(\theta_0)$ is the (j, j) element of $\widehat{\Sigma}_n(\theta_0)$ for $j = 1, \dots, p$. The GMS procedure based on $\varphi^{(1)}$ is the same as the Wald test procedure in Andrews (1999b,

Sec. 6.4; 2000, Sec. 4) for the related problem of inference when a parameter is on or near a boundary.

The function $\varphi^{(2)}$ in (4.5) depends on a nondecreasing function $\psi(x)$ that satisfies $\psi(x) = 0$ if $x \leq a_L$, $\psi(x) \in [0, \infty]$ if $a_L < x < a_U$, and $\psi(x) = \infty$ if $x > a_U$ for some $0 < a_L \leq a_U \leq \infty$. A key condition is that $a_L > 0$; see Assumption GMS1(a) below. The function $\varphi^{(2)}$ is a continuous version of $\varphi^{(1)}$ when ψ is taken to be continuous on R (where continuity at a_U means that $\lim_{x \rightarrow a_U} \psi(x) = \infty$).

The functions $\varphi^{(3)}$ and $\varphi^{(4)}$ exhibit a less steep rate of increase than $\varphi^{(1)}$ as a function of ξ_j for $j = 1, \dots, p$.

The functions $\varphi^{(r)}$ for $r = 1, \dots, 4$ all exhibit “element-by-element” determination of $\varphi_j^{(r)}(\xi, \Omega)$ because the latter depends only on ξ_j . This has significant computational advantages because $\varphi_j^{(r)}(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ is very easy to compute. On the other hand, when $\widehat{\Omega}_n(\theta_0)$ is nondiagonal, the whole vector $\xi_n(\theta_0)$ contains information about the magnitude of $h_{1,j}$. We now introduce a function $\varphi^{(5)}$ that exploits this information (at least for certain choices of function S such as S_2). It is related to the information criterion-based moment selection criteria (MSC) considered in Andrews (1999a) for a different moment selection problem. We refer to $\varphi^{(5)}$ as the modified MSC (MMSC) φ function. It is computationally more expensive than the $\varphi^{(r)}$ functions considered above.

Define $c = (c_1, \dots, c_k)'$ to be a selection k -vector of 0's and 1's. If $c_j = 1$, the j th moment condition is selected; if $c_j = 0$, it is not selected. The moment equality functions are always selected, that is, $c_j = 1$ for $j = p + 1, \dots, k$. Let $|c| = \sum_{j=1}^k c_j$. For $\xi \in R_{[+\infty]}^p \times R_{[\pm\infty]}^v$, define $c \cdot \xi = (c_1 \xi_1, \dots, c_k \xi_k)' \in R_{[+\infty]}^p \times R_{[\pm\infty]}^v$, where $c_j \xi_j = 0$ if $c_j = 0$ and $\xi_j = \infty$. Let \mathcal{C} denote the parameter space for the selection vectors. In many cases, $\mathcal{C} = \{0, 1\}^p \times \{1\}^v$. However, if there is a priori information that one moment inequality cannot hold as an equality if some other does and the sum of the degrees of slackness of the two moment inequalities is bounded away from zero over all admissible distributions, then this can be built into the definition of \mathcal{C} ; see Rosen (2008) for a discussion of examples of this sort. Let $\zeta(\cdot)$ be a strictly increasing real function on R_+ . Given $(\xi, \Omega) \in (R_{[+\infty]}^p \times R_{[\pm\infty]}^v) \times \Psi$, the selected moment vector $c(\xi, \Omega) \in \mathcal{C}$ is the vector in \mathcal{C} that minimizes the MMSC defined by

$$(4.7) \quad S(-c \cdot \xi, \Omega) - \zeta(|c|).$$

Note the minus sign that appears in the first argument of the S function. This ensures that a large *positive* value of ξ_j yields a large value of $S(-c \cdot \xi, \Omega)$ when $c_j = 1$, as desired. Since $\zeta(\cdot)$ is increasing, $-\zeta(|c|)$ is a bonus term that rewards inclusion of more moments. Hence, the minimizing selection vector $c(\xi, \Omega)$ trades off the minimization of $S(-c \cdot \xi, \Omega)$, which is achieved by selecting few

moment functions, with the maximization of the bonus term, which is decreasing in the number of selected moments. For $j = 1, \dots, p$, define

$$(4.8) \quad \varphi_j^{(5)}(\xi, \Omega) = \begin{cases} 0 & \text{if } c_j(\xi, \Omega) = 1, \\ \infty & \text{if } c_j(\xi, \Omega) = 0. \end{cases}$$

Using Assumptions 1(b) and 6,

$$(4.9) \quad \kappa_n^\chi(S(-c \cdot \xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)) - \zeta(|c|)) \\ = S(-c \cdot n^{1/2} \overline{m}_n(\theta_0), \widehat{\Sigma}_n(\theta_0)) - \zeta(|c|) \kappa_n^\chi,$$

where χ is as in Assumption 6. In consequence, the MMSC selection vector $c(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$ minimizes both the left-hand and right-hand sides (r.h.s.) of (4.9) over \mathcal{C} . The r.h.s. of (4.9) is analogous to the BIC and HQIC criteria considered in the model selection literature in which case $\zeta(x) = x$, $\kappa_n = (\ln n)^{1/2}$ for BIC, $\kappa_n = (Q \ln \ln n)^{1/2}$ for some $Q \geq 2$ for HQIC, and $\chi = 2$ (which holds for the functions S_1 – S_3). Note that some calculations show that when $\widehat{\Omega}_n(\theta_0)$ is diagonal, $S = S_1$ or S_2 , and $\zeta(x) = x$, the function $\varphi^{(5)}$ reduces to $\varphi^{(1)}$.

Returning now to the general case, given a choice of function $\varphi(\xi, \Omega) = (\varphi_1(\xi, \Omega), \dots, \varphi_p(\xi, \Omega), 0, \dots, 0)' \in R_{[+\infty]}^p \times \{0\}^v$, the replacement for the k -vector $(h_1, 0_v)$ in $S(\Omega_0^{1/2} Z^* + (h_1, 0_v), \Omega_0)$ is $\varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$. Thus, the GMS critical value, $\widehat{c}_n(\theta_0, 1 - \alpha)$, is the $1 - \alpha$ quantile of

$$(4.10) \quad L_n(\theta_0, Z^*) = S(\widehat{\Omega}_n^{1/2}(\theta_0) Z^* + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_n(\theta_0)),$$

where $Z^* \sim N(0_k, I_k)$ and Z^* is independent of $\{W_i : i \geq 1\}$. That is,

$$(4.11) \quad \widehat{c}_n(\theta_0, 1 - \alpha) = \inf\{x \in R : P(L_n(\theta_0, Z^*) \leq x) \geq 1 - \alpha\},$$

where $P(L_n(\theta_0, Z^*) \leq x)$ denotes the conditional df at x of $L_n(\theta_0, Z^*)$ given $(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$. One can compute $\widehat{c}_n(\theta_0, 1 - \alpha)$ by simulating R i.i.d. random vectors $\{Z_r^* : r = 1, \dots, R\}$ with $Z_r^* \sim N(0_k, I_k)$ and taking $\widehat{c}_n(\theta_0, 1 - \alpha)$ to be the $1 - \alpha$ sample quantile of $\{L_n(\theta_0, Z_r^*) : r = 1, \dots, R\}$, where R is large.

A bootstrap version of the GMS critical value is obtained by replacing $L_n(\theta_0, Z^*)$ in (4.11) by

$$(4.12) \quad S(M_n^*(\theta_0) + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_n^*(\theta_0)),$$

where $M_n^*(\theta)$ is a recentered bootstrapped version of $n^{1/2} \widehat{D}_n^{-1/2}(\theta) \overline{m}_n(\theta)$ and $\widehat{\Omega}_n^*(\theta)$ is a bootstrapped version of $\widehat{\Omega}_n(\theta)$ (defined as follows). Let $\{W_i^* : i \leq n\}$ be a bootstrap sample, such as a nonparametric i.i.d. bootstrap sample in an i.i.d. scenario or a block bootstrap sample in a time series scenario. By definition,

$$(4.13) \quad M_n^*(\theta) = n^{1/2} (\widehat{D}_n^*(\theta))^{-1/2} (\overline{m}_n^*(\theta) - \overline{m}_n(\theta)),$$

$$\widehat{\Omega}_n^*(\theta) = (\widehat{D}_n^*(\theta))^{-1/2} \widehat{\Sigma}_n^*(\theta) (\widehat{D}_n^*(\theta))^{-1/2}, \quad \text{where}$$

$$\overline{m}_n^*(\theta) = n^{-1} \sum_{i=1}^n m(W_i^*, \theta), \quad \widehat{D}_n^*(\theta) = \text{Diag}(\widehat{\Sigma}_n^*(\theta)),$$

and $\widehat{\Sigma}_n^*(\theta)$ is defined in the same manner as $\widehat{\Sigma}_n(\theta)$ is defined (e.g., as in (3.2) in the i.i.d. case) with W_i^* in place of W_i . One can compute the bootstrap critical value by simulating R bootstrap samples $\{(W_{i,r}^* : i \leq n) : r = 1, \dots, R\}$ (i.i.d. across samples), computing $\{(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0)) : r = 1, \dots, R\}$ (defined as in (4.13)), and taking the bootstrap critical value to be the $1 - \alpha$ sample quantile of $\{S(M_{n,r}^*(\theta_0) + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_{n,r}^*(\theta_0)) : r = 1, \dots, R\}$, where R is large.

For the asymptotic results given below to hold with a bootstrap GMS critical value, one needs that $M_n^*(\theta_{n,h}) \rightarrow_d \Omega_0^{1/2} Z^*$ under certain triangular arrays of true distributions and true parameters $\theta_{n,h}$, where Ω_0 is a $k \times k$ correlation matrix and Z^* is as in (4.10).⁸ This can be established for the nonparametric i.i.d. and block bootstraps using fairly standard arguments. For brevity, we do not do so here.

The 2003 working paper version of CHT discusses a bootstrap version of the GMS critical value based on $\varphi^{(1)}$ in the context of the interval outcome regression model. CHT mentions critical values of GMS type based on $\varphi^{(1)}$; see their Remark 4.5. Bugni (2007a, 2007b) and Canay (2007) provided results regarding the pointwise asymptotic null properties of nonparametric i.i.d. bootstrap procedures applied with $\varphi^{(1)}$ and $\varphi^{(3)}$, respectively. Note that GMS bootstrap critical values do not generate higher-order improvements in the present context because the asymptotic null distribution of the test statistic $T_n(\theta)$ is not asymptotically pivotal. Fan and Park (2007) considered a critical value based on a function that is analogous to $\varphi^{(1)}$ except with ∞ replaced by $\kappa_n \xi_j$ and with ξ_j replaced by $\xi_j \widehat{\sigma}_{n,j}(\theta)$. The latter makes their procedure lack invariance to the scaling of the moment functions, which is not desirable. If Fan and Park's (2007) φ function is altered to be scale invariant, then the test based on it has the same asymptotic properties under the null and local alternatives as the test based on $\varphi^{(1)}$ because $\kappa_n \rightarrow \infty$.

4.2. Step-by-Step Calculation of GMS Tests and CIs

Here we describe the steps in the calculation of a nominal level α GMS test of $H_0: \theta = \theta_0$. First, we describe the bootstrap version of the GMS procedure based on $(S_2, \varphi^{(1)})$, which is the recommended procedure for i.i.d. observa-

⁸More specifically, this convergence must hold under any sequence of distributions $\{\gamma_n : n \geq 1\}$ defined just above (A.3) in the Appendix (in which case $\Omega_0 = \Omega_{h_{2,2}}$), the convergence needs to be joint with that in (A.3) of the Appendix, and the convergence must hold with $\{n\}$ replaced by any subsequence $\{w_n\}$ of sample sizes.

tions. (For non-i.i.d. observations, we recommend using $(S_2, \varphi^{(1)})$, but the asymptotic version may perform as well as a bootstrap version.)

Compute (i) the sample moments $\bar{m}_n(\theta_0)$, (ii) the sample variance estimator $\widehat{\Sigma}_n(\theta_0)$, and (iii) the test statistic $T_n(\theta_0) = S_2(n^{1/2}\bar{m}_n(\theta_0), \widehat{\Sigma}_n(\theta_0))$. Next, to determine the GMS critical value $\widehat{c}_n(\theta_0, 1 - \alpha)$, use the following steps: (iv) simulate R bootstrap samples each of size n , i.i.d. across bootstrap samples, denoted $\{\{W_{i,r}^*: i \leq n\}: r = 1, \dots, R\}$ (according to a bootstrap procedure that is suitable for the observations under consideration, such as a nonparametric i.i.d. bootstrap for i.i.d. observations and a block bootstrap for time series), where the number of bootstrap simulations R is large, say 1000 or more; (v) compute $\{(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0)): r = 1, \dots, R\}$ (defined in (4.13) with $W_{i,r}^*$ in place of W_i^*); (vi) determine whether $n^{1/2}\bar{m}_{n,j}(\theta_0)/\widehat{\sigma}_{n,j}(\theta_0) > \kappa_n = (\ln n)^{1/2}$ for $j = 1, \dots, p$, where $\widehat{\sigma}_{n,j}^2(\theta_0)$ is the (j, j) element of $\widehat{\Sigma}_n(\theta_0)$; (vii) eliminate the elements in $(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0))$ for all $r = 1, \dots, R$ that correspond to the moment conditions that satisfy the condition in (vi), with the resulting quantities denoted by $(M_{n,r}^{**}(\theta_0), \widehat{\Omega}_{n,r}^{**}(\theta_0))$ for $r = 1, \dots, R$; (viii) take the critical value $\widehat{c}_n(\theta_0, 1 - \alpha)$ to be the $1 - \alpha$ sample quantile of $\{S_2(M_{n,r}^{**}(\theta_0), \widehat{\Omega}_{n,r}^{**}(\theta_0)): r = 1, \dots, R\}$. The GMS test rejects $H_0: \theta = \theta_0$ if $T_n(\theta_0) > \widehat{c}_n(\theta_0, 1 - \alpha)$.

A GMS CS is obtained by inverting tests of $H_0: \theta = \theta_0$ for $\theta_0 \in \Theta$. A GMS CS can be calculated by employing a grid search (or some more sophisticated) algorithm using the method described above to calculate whether $T_n(\theta_0) \leq \widehat{c}_n(\theta_0, 1 - \alpha)$, which implies that θ_0 should be included in the CS.

For a general choice of (S, φ) , the asymptotic version of the GMS test is computed as follows. Compute (i) the sample moments $\bar{m}_n(\theta_0)$, (ii) the sample variance estimator $\widehat{\Sigma}_n(\theta_0)$, and (iii) the test statistic $T_n(\theta_0) = S(n^{1/2} \times \bar{m}_n(\theta_0), \widehat{\Sigma}_n(\theta_0))$. Next, to determine the critical value $\widehat{c}_n(\theta_0, 1 - \alpha)$, compute (iv) $\widehat{\Omega}_n(\theta_0) = \text{Diag}^{-1/2}(\widehat{\Sigma}_n(\theta_0))\widehat{\Sigma}_n(\theta_0)\text{Diag}^{-1/2}(\widehat{\Sigma}_n(\theta_0))$, (v) $\xi_n(\theta_0) = \kappa_n^{-1}n^{1/2}\text{Diag}^{-1/2}(\widehat{\Sigma}_n(\theta_0))\bar{m}_n(\theta_0)$, where $\kappa_n = (\ln n)^{1/2}$, and (vi) $\varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0))$; (vii) simulate R i.i.d. random vectors $\{Z_r^*: r = 1, \dots, R\}$ with $Z_r^* \sim N(0_k, I_k)$, where R is large; and (viii) take $\widehat{c}_n(\theta_0, 1 - \alpha)$ to be the $1 - \alpha$ sample quantile of $\{S(\widehat{\Omega}_n^{1/2}(\theta_0)Z_r^* + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_n(\theta_0)): r = 1, \dots, R\}$. The GMS test rejects $H_0: \theta = \theta_0$ if $T_n(\theta_0) > \widehat{c}_n(\theta_0, 1 - \alpha)$.

The bootstrap version of the GMS test with general choice of (S, φ) replaces steps (vii) and (viii) with the following: (vii*) simulate R bootstrap samples each of size n , i.i.d. across bootstrap samples, denoted $\{\{W_{i,r}^*: i \leq n\}: r = 1, \dots, R\}$, (viii*) compute $\{(M_{n,r}^*(\theta_0), \widehat{\Omega}_{n,r}^*(\theta_0)): r = 1, \dots, R\}$ (defined in (4.13) with $W_{i,r}^*$ in place of W_i^*), and (ix*) take the critical value $\widehat{c}_n(\theta_0, 1 - \alpha)$ to be the $1 - \alpha$ sample quantile of $\{S(M_{n,r}^*(\theta_0) + \varphi(\xi_n(\theta_0), \widehat{\Omega}_n(\theta_0)), \widehat{\Omega}_{n,r}^*(\theta_0)): r = 1, \dots, R\}$. When $S = S_2$ and $\varphi = \varphi^{(1)}$, this procedure is equivalent to that for the GMS test described three paragraphs above.

4.3. Assumptions

Next we state assumptions on the function φ and the constants $\{\kappa_n : n \geq 1\}$ that define a GMS procedure. The first two assumptions are used to show that GMS CS's and tests have correct asymptotic size.

ASSUMPTION GMS1: (a) $\varphi_j(\xi, \Omega)$ is continuous at all $(\xi, \Omega) \in (R_{[+\infty]}^p \times R_{[\pm\infty]}^v) \times \Psi$ with $\xi_j = 0$, where $\xi = (\xi_1, \dots, \xi_k)'$, for $j = 1, \dots, p$.

(b) $\varphi_j(\xi, \Omega) = 0$ for all $(\xi, \Omega) \in (R_{[+\infty]}^p \times R_{[\pm\infty]}^v) \times \Psi$ with $\xi_j = 0$, where $\xi = (\xi_1, \dots, \xi_k)'$, for $j = 1, \dots, p$.

(c) $\varphi_j(\xi, \Omega) = 0$ for all $j = p + 1, \dots, k$ for all $(\xi, \Omega) \in (R_{[+\infty]}^p \times R_{[\pm\infty]}^v) \times \Psi$.

ASSUMPTION GMS2: $\kappa_n \rightarrow \infty$.

Assumptions **GMS1** and **GMS2** are not restrictive. For example, the functions $\varphi^{(1)} - \varphi^{(4)}$ satisfy Assumption **GMS1** and $\kappa_n = (\ln n)^{1/2}$ satisfies Assumption **GMS2**. Assumption **GMS1** also holds for $\varphi^{(5)}$ for all functions S that satisfy Assumption 1(d), which includes $S_1 - S_3$; see the Supplemental Material (Andrews and Soares (2010)) for a proof.

The next two assumptions are used in conjunction with Assumptions **GMS1** and **GMS2** to show that GMS CS's and tests are not asymptotically conservative. They also are used to determine the formula for the asymptotic power of GMS tests against $n^{-1/2}$ -local alternatives.

ASSUMPTION GMS3: $\varphi_j(\xi, \Omega) \rightarrow \infty$ as $(\xi, \Omega) \rightarrow (\xi_*, \Omega_*)$ for all $(\xi_*, \Omega_*) \in R_{[+\infty]}^p \times R_{[\pm\infty]}^v \times \text{cl}(\Psi)$ with $\xi_{*,j} = \infty$, where $\xi_* = (\xi_{*,1}, \dots, \xi_{*,k})'$, for $j = 1, \dots, p$.

ASSUMPTION GMS4: $\kappa_n^{-1} n^{1/2} \rightarrow \infty$.

Assumptions **GMS3** and **GMS4** are not restrictive and are satisfied by $\varphi^{(1)} - \varphi^{(4)}$ and $\kappa_n = (\ln n)^{1/2}$. Assumption GMS3 also holds for $\varphi^{(5)}$ for all functions S that satisfy Assumption 1(d) and for which $S(-c \cdot \xi, \Omega) \rightarrow \infty$ as $(\xi, \Omega) \rightarrow (\xi_*, \Omega_*)$ whenever $c_j = 1$, see the Supplemental Material for a proof. The latter holds for the test functions $S_1 - S_3$.

The next two assumptions are used in conjunction with Assumptions **GMS2** and **GMS3** to show that GMS tests dominate subsampling tests (based on a subsample size b) in terms of $n^{-1/2}$ -local asymptotic power.

ASSUMPTION GMS5: $\kappa_n^{-1} (n/b)^{1/2} \rightarrow \infty$, where $b = b_n$ is the subsample size.

ASSUMPTION GMS6: $\varphi_j(\xi, \Omega) \geq 0$ for all $(\xi, \Omega) \in (R_{[+\infty]}^p \times R_{[\pm\infty]}^v) \times \Psi$ for $j = 1, \dots, p$.

Assumption **GMS5** holds for all reasonable choices of κ_n and b . For example, for $\kappa_n = (\ln n)^{1/2}$, Assumption **GMS5** holds for $b = n^\eta$ for any $\eta \in (0, 1)$. Any reasonable choice of b satisfies the latter condition. Note that for recentered subsampling tests, the optimal value of η in terms of size is $2/3$; see Bugni (2007a, 2007b). When $\eta = 2/3$, Assumption **GMS5** holds easily for κ_n as above. Assumption **GMSS** fails when $\kappa_n = (\ln n)^{1/2}$ (or some other logarithmic function) only if b is larger than $O(n^\eta)$ for all $\eta \in (0, 1)$ and the latter yields a recentered subsampling test whose error in the null rejection probability is very large—of order larger than $O(n^{-\varepsilon})$ for all $\varepsilon > 0$. For a non-recentered subsampling test, it yields a test whose power against $n^{1/2}$ -local alternatives converges from below and very slowly to its asymptotic local power. The reason is that if b is larger than $O(n^\eta)$ for all $\eta \in (0, 1)$, then under the alternative, the subsampling critical value mimics the $1 - \alpha$ quantile of the alternative distribution of the test statistic unless n is very, very large, because the subsample size is almost equal to that of the full-sample statistic unless n is very, very large. Hence, the finite-sample power of the subsampling test is poor in sample sizes that are of interest in practice.

Assumption **GMS6** is satisfied by the functions $\varphi^{(1)} - \varphi^{(5)}$ except for $\varphi^{(4)}$. Hence, it is slightly restrictive.

The last assumption is used to show that GMS tests are consistent against alternatives that are more distant from the null than $n^{-1/2}$ -local alternatives.

ASSUMPTION GMS7: $\varphi_j(\xi, \Omega) \geq \min\{\xi_j, 0\}$ for all $(\xi, \Omega) \in (R_{[+\infty]}^p \times R_{[\pm\infty]}^v) \times \Psi$ for $j = 1, \dots, p$.

Assumption **GMS7** is not restrictive. For example, it is satisfied by $\varphi^{(1)} - \varphi^{(5)}$.

Next we introduce a condition that depends on the model, not on the GMS method, and is only used when showing that GMS CS's have AsyMaxCP = 1 when $v = 0$.

ASSUMPTION M: For some $(\theta, F) \in \mathcal{F}$, $E_F m_j(W_i, \theta) > 0$ for all $j = 1, \dots, p$.

Assumption **M** typically holds if the identified set (i.e., the set of parameter values θ that satisfy the population moment inequalities and equalities under F) has a nonempty interior for some data-generating process included in the model.

4.4. Asymptotic Size Results

The following theorem applies to i.i.d. observations, in which case \mathcal{F} is as defined in (2.2), and to dependent observations, in which case for brevity \mathcal{F} is as defined in (A.2) and (A.3) in the Appendix.

THEOREM 1: Suppose Assumptions 1–3, **GMS1**, and **GMS2** hold and $0 < \alpha < 1/2$. Then the nominal level $1 - \alpha$ GMS CS based on $T_n(\theta)$ satisfies the following statements:

- (a) $\text{AsyCS} \geq 1 - \alpha$.
- (b) $\text{AsyCS} = 1 - \alpha$ if Assumptions GMS3, GMS4, and 7 also hold.
- (c) $\text{AsyMaxCP} = 1$ if $v = 0$ (i.e., no moment equalities appear) and Assumption M also holds.

COMMENTS: (i) Theorem 1(a) shows that a GMS CS is asymptotically valid in a uniform sense. Theorem 1(b) shows it is not asymptotically conservative. Theorem 1(c) shows it is not asymptotically similar.

(ii) Theorem 1 places no assumptions on the moment functions $m(W_i, \theta)$ beyond the existence of mild moment conditions that appear in the definition of \mathcal{F} . Thus, the results apply to moment conditions based on instruments that are weak. (The reason is that the test statistics considered are of the Anderson–Rubin type.)

(iii) Theorem 1 holds even when there are restrictions on the moment inequalities such that when one moment inequality holds as an equality, then another moment inequality cannot. Restrictions of this sort arise in some models, such as models with interval outcomes (e.g., see Rosen (2008)).

(iv) The proof of Theorem 1 and all other results below are given in the Supplemental Material (Andrews and Soares (2010)).

5. GMS MODEL SPECIFICATION TESTS

Tests of model specification can be constructed using the GMS CS introduced above. The null hypothesis of interest is that (2.1) holds for some parameter $\theta_0 \in \Theta$ (with additional conditions imposed by the parameter space for (θ, F)). By definition, the GMS test rejects the model specification if $T_n(\theta)$ exceeds the GMS critical value $\widehat{c}_n(\theta, 1 - \alpha)$ for all $\theta \in \Theta$. Equivalently, it rejects if the GMS CS is empty. The idea behind such a test is the same as for the J test of overidentifying restrictions in GMM; see Hansen (1982).

When the model of (2.1) is correctly specified, the GMS CS includes the true value with asymptotic probability $1 - \alpha$ (or greater) uniformly over the parameter space. Thus, under the null hypothesis of correct model specification, the limit as $n \rightarrow \infty$ of the finite-sample size of the GMS model specification test is less than or equal to α under the assumptions of Theorem 1(a). In other words, the asymptotic size of this specification test is valid uniformly over the parameter space.

Note that the asymptotic size of the GMS model specification test is not necessarily equal to α under the assumptions of Theorem 1(b).⁹ That is, the GMS model specification test may be asymptotically conservative.

⁹The reason is that when the null of correct model specification holds and (θ_0, F_0) is the truth, the GMS test may fail to reject the null even when $T_n(\theta_0) > \widehat{c}_n(\theta_0, 1 - \alpha)$ because $T_n(\theta) \leq \widehat{c}_n(\theta, 1 - \alpha)$ for some $\theta \neq \theta_0$.

6. SUBSAMPLING CONFIDENCE SETS

The volume of a CS is directly related to the power of the tests used in its construction. Below we compare the power of GMS tests to that of subsampling and PA tests. In this section and the following one we define subsampling and PA CS's.

We now define subsampling critical values and CS's. Let $b = b_n$ denote the subsample size when the full-sample size is n . We assume $b \rightarrow \infty$ and $b/n \rightarrow 0$ as $n \rightarrow \infty$ (here and below). The number of subsamples of size b considered is q_n . With i.i.d. observations, there are $q_n = n!/(n-b)!b!$ subsamples of size b . With time series observations, there are $q_n = n-b+1$ subsamples, each based on b consecutive observations.

Let $T_{n,b,j}(\theta)$ be a subsample statistic defined exactly as $T_n(\theta)$ is defined but based on the j th subsample of size b rather than the full sample for $j = 1, \dots, q_n$. The empirical df and the $1 - \alpha$ sample quantile of $\{T_{n,b,j}(\theta) : j = 1, \dots, q_n\}$ are

$$(6.1) \quad U_{n,b}(\theta, x) = q_n^{-1} \sum_{j=1}^{q_n} 1(T_{n,b,j}(\theta) \leq x) \quad \text{for } x \in R,$$

$$c_{n,b}(\theta, 1 - \alpha) = \inf\{x \in R : U_{n,b}(\theta, x) \geq 1 - \alpha\}.$$

The subsampling test rejects $H_0 : \theta = \theta_0$ if $T_n(\theta_0) > c_{n,b}(\theta_0, 1 - \alpha)$. The nominal level $1 - \alpha$ subsampling CS is given by (2.3) with $c_{1-\alpha}(\theta) = c_{n,b}(\theta, 1 - \alpha)$.

One also can define “recentered” subsample statistics by defining $T_{n,b,j}(\theta)$ using $b^{1/2}(\bar{m}_{n,b,j}(\theta) - \bar{m}_n(\theta))$, rather than $b^{1/2}\bar{m}_{n,b,j}(\theta)$, in place of $n^{1/2}\bar{m}_n(\theta)$ in (3.3), where $\bar{m}_{n,b,j}(\theta)$ is the average of the moment conditions over the observations in the j th subsample; see AG4.

It is shown in AG4 that under Assumptions 1–3 and $0 < \alpha < 1/2$, the nominal level $1 - \alpha$ subsampling CS based on $T_n(\theta)$ satisfies (a) AsyCS $\geq 1 - \alpha$, (b) AsyCS $= 1 - \alpha$ if Assumption 7 also holds, and (c) AsyMaxCP $= 1$ if $v = 0$ (i.e., no moment equalities appear) and Assumption M also holds.

7. PLUG-IN ASYMPTOTIC CONFIDENCE SETS

Now we discuss CS's based on a PA critical value. The least favorable asymptotic null distributions of the statistic $T_n(\theta)$ are shown in AG4 to be those for which the moment inequalities hold as equalities. These distributions depend on the correlation matrix Ω of the moment functions. We analyze plug-in asymptotic (PA) critical values that are determined by the least favorable asymptotic null distribution for given Ω evaluated at a consistent estimator of Ω . Such critical values have been considered for many years in the literature on multivariate one-sided tests; see Silvapulle and Sen (2005) for references. AG4 considered them in the context of the moment inequality literature. Rosen (2008) considered variations of PA critical values that make adjustments in

the case where it is known that if one moment inequality holds as an equality, then another cannot.

Let $c(\Omega, 1 - \alpha)$ denote the $1 - \alpha$ quantile of $S(Z, \Omega)$, where $Z \sim N(0_k, \Omega)$. This is the $1 - \alpha$ quantile of the asymptotic null distribution of $T_n(\theta)$ when the moment inequalities hold as equalities.

The nominal $1 - \alpha$ PA CS is given by (2.3) with critical value $c_{1-\alpha}(\theta)$ equal to

$$(7.1) \quad c(\widehat{\Omega}_n(\theta), 1 - \alpha).$$

AG4 showed that if Assumptions 1 and 4 hold and $0 < \alpha < 1/2$, then the nominal level $1 - \alpha$ PA CS based on $T_n(\theta)$ satisfies AsyCS $\geq 1 - \alpha$.

8. LOCAL ALTERNATIVE POWER COMPARISONS

In this section and the next, we compare the power of GMS, subsampling, and PA tests. These results have immediate implications for the volume of CS's based on these tests because the power of a test for a point that is not the true value is the probability that the CS does not include that point. Here we analyze the power of tests against $n^{-1/2}$ -local alternatives. In the next section we consider “distant alternatives,” which differ from the null by more than $O(n^{-1/2})$ and may be fixed or local.

We show that a GMS test has asymptotic power that is greater than or equal to that of a subsampling or PA test (based on the same test statistic) under all alternatives. We show that a GMS test's power is *strictly greater* than that of a subsampling test in the scenario stated in the [Introduction](#). In addition, we show that GMS and subsampling tests have asymptotic power that is greater than or equal to that of a PA test with strictly greater power in the scenarios stated in the [Introduction](#).

For given $\theta_{n,*}$, we consider tests of

$$(8.1) \quad H_0 : E_{F_n} m_j(W_i, \theta_{n,*}) \begin{cases} \geq 0 & \text{for } j = 1, \dots, p, \\ = 0 & \text{for } j = p + 1, \dots, k, \end{cases}$$

where F_n denotes the true distribution of the data, versus $H_1 : H_0$ does not hold. For brevity, we only give results for the case of i.i.d. observations. (The results can be extended to dependent observations, and the advantage of GMS tests over subsampling and PA tests also holds with dependent observations.) The parameter space \mathcal{F} for (θ, F) is assumed to satisfy (2.2).

With i.i.d. observations, F denotes the distribution of W_i . We consider the Kolmogorov–Smirnov metric on the space of distributions F . Let

$$(8.2) \quad D(\theta, F) = \text{Diag}\{\sigma_{F,1}^2(\theta), \dots, \sigma_{F,k}^2(\theta)\}, \quad \Omega(\theta, F) = \text{Corr}_F(m(W_i, \theta)).$$

We now introduce the $n^{-1/2}$ -local alternatives that are considered.

ASSUMPTION LA1: *The true parameters $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ satisfy the following statements:*

- (a) $\theta_n = \theta_{n,*} - \lambda n^{-1/2}(1 + o(1))$ for some $\lambda \in R^d$, $\theta_{n,*} \rightarrow \theta_0$, and $F_n \rightarrow F_0$ for some $(\theta_0, F_0) \in \mathcal{F}$.
- (b) $n^{1/2} E_{F_n} m_j(W_i, \theta_n) / \sigma_{F_n, j}(\theta_n) \rightarrow h_{1,j}$ for some $h_{1,j} \in R_{+, \infty}$ for $j = 1, \dots, p$.
- (c) $\sup_{n \geq 1} E_{F_n} |m_j(W_i, \theta_{n,*}) / \sigma_{F_n, j}(\theta_{n,*})|^{2+\delta} < \infty$ for $j = 1, \dots, k$ for some $\delta > 0$.

ASSUMPTION LA2: *The $k \times d$ matrix $\Pi(\theta, F) = (\partial/\partial\theta')[D^{-1/2}(\theta, F)E_F m(W_i, \theta)]$ exists and is continuous in (θ, F) for all (θ, F) in a neighborhood of (θ_0, F_0) .*

Assumption **LA1(a)** specifies that the true values $\{\theta_n : n \geq 1\}$ are local to the null values $\{\theta_{n,*} : n \geq 1\}$. Assumption **LA1(b)** specifies the asymptotic behavior of the (normalized) moment inequality functions when evaluated at the true parameter values $\{\theta_n : n \geq 1\}$. Under the true values, these (normalized) moment inequalities are nonnegative. Assumption **LA1(a)** and (c) imply that $\Omega(\theta_{n,*}, F_n)$ exists and $\Omega(\theta_{n,*}, F_n) \rightarrow \Omega_0 = \Omega(\theta_0, F_0)$.

The asymptotic distribution of the test statistic $T_n(\theta_{n,*})$ under $n^{-1/2}$ -local alternatives depends on the limit of the (normalized) moment inequality functions when evaluated at the null value $\theta_{n,*}$ because $T_n(\theta_{n,*})$ is evaluated at $\theta_{n,*}$. Under Assumptions **LA1** and **LA2**, we show that

$$(8.3) \quad \lim_{n \rightarrow \infty} n^{1/2} D^{-1/2}(\theta_{n,*}, F_n) E_{F_n} m(W_i, \theta_{n,*}) = (h_1, 0_v) + \Pi_0 \lambda \in R^k, \quad \text{where} \\ h_1 = (h_{1,1}, \dots, h_{1,p})' \quad \text{and} \quad \Pi_0 = \Pi(\theta_0, F_0).$$

By definition, if $h_{1,j} = \infty$, then $h_{1,j} + y = \infty$ for any $y \in R$. Let $\Pi_{0,j}$ denote the j th row of Π_0 written as a column d -vector for $j = 1, \dots, k$. Note that $(h_1, 0_v) + \Pi_0 \lambda \in R_{[+\infty]}^p \times R^v$.

The following assumption states that the true distribution of the data F_n is in the alternative, not the null (for n large).

ASSUMPTION LA3: *$h_{1,j} + \Pi_{0,j}' \lambda < 0$ for some $j = 1, \dots, p$ or $\Pi_{0,j}' \lambda \neq 0$ for some $j = p+1, \dots, k$.*

The following is a simple example to illustrate Assumptions **LA1–LA3**. Suppose $m(W_i, \theta) = W_i - \theta$, $E_F m(W_i, \theta) \geq 0$, and $\text{Var}_F(m(W_i, \theta)) = 1$ for all $(\theta, F) \in \mathcal{F}$. Then $p = 1$, $v = 0$, and $D(\theta, F) = 1$. Consider a sequence of true parameters/distributions $\{(\theta_n, F_n) \in \mathcal{F} : n \geq 1\}$ that satisfy $\theta_n = \theta_{n,*} - \lambda n^{-1/2}$, $E_{F_n} W_i = \theta_n + h_1 n^{-1/2}$ for some $\theta_{n,*}, \lambda \in R$, and $h_1 \geq 0$, and $\theta_{n,*} \rightarrow \theta_0$. Then Assumption **LA1(a)** holds and in Assumption **LA1(b)**, we have $n^{1/2} E_{F_n} m_j(W_i, \theta_n) / \sigma_{F_n, j}(\theta_n) = n^{1/2} (E_{F_n} W_i - \theta_n) = h_1 \geq 0$ for all n (using $\sigma_{F_n, j}(\theta_n) = 1$). So, Assumption **LA1(b)** also holds. We have $\Pi(\theta, F) = -1$ for all (θ, F) . Hence, in Assumption **LA3**, $h_1 + \Pi_0 \lambda = h_1 - \lambda$, which is negative whenever $\lambda > h_1$. Hence, if the null value $\theta_{n,*}$ deviates from the true value θ_n

by enough (i.e., if $\theta_{n,*} - \theta_n = \lambda n^{-1/2}$ is large enough) relative to the magnitude of the slackness of the moment condition (i.e., $E_{F_n} W_i - \theta_n = h_1 n^{-1/2}$), then the null hypothesis is violated for all n and Assumption **LA3** holds.

The asymptotic distribution of $T_n(\theta_{n,*})$ under $n^{-1/2}$ -local alternatives is shown to be $J_{h_1,\lambda}$, where $J_{h_1,\lambda}$ is defined by

$$(8.4) \quad S(\Omega_0^{1/2} Z^* + (h_1, 0_v) + \Pi_0 \lambda, \Omega_0) \sim J_{h_1,\lambda}$$

for $Z^* \sim N(0_k, I_k)$. For notational simplicity, the dependence of $J_{h_1,\lambda}$ on Ω_0 and Π_0 is suppressed. Let $c_{h_1,\lambda}(1 - \alpha)$ denote the $1 - \alpha$ quantile of $J_{h_1,\lambda}$.

We now introduce two assumptions that are used for GMS tests only.

ASSUMPTION LA4: $\kappa_n^{-1} n^{1/2} E_{F_n} m_j(W_i, \theta_n) / \sigma_{F_n,j}(\theta_n) \rightarrow \pi_{1,j}$ for some $\pi_{1,j} \in R_{+, \infty}$ for $j = 1, \dots, p$.

Note that in Assumption **LA4**, the functions are evaluated at the true value θ_n , not at the null value $\theta_{n,*}$, and $(\theta_n, F_n) \in \mathcal{F}$. In consequence, the moment functions in Assumption **LA4** satisfy the inequalities and $\pi_{1,j} \geq 0$ for all $j = 1, \dots, p$.

Let $\pi_1 = (\pi_{1,1}, \dots, \pi_{1,p})'$. Let $c_{\pi_1}(\varphi, 1 - \alpha)$ denote the $1 - \alpha$ quantile of

$$(8.5) \quad S(\Omega_0^{1/2} Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0), \quad \text{where } Z^* \sim N(0_k, I_k).$$

Below the probability limit of the GMS critical value, $\widehat{c}_n(\theta_{n,*}, 1 - \alpha)$ is shown to be $c_{\pi_1}(\varphi, 1 - \alpha)$.

The following assumption is used to obtain the $n^{-1/2}$ -local alternative power function of the GMS test. Let $C(\varphi) = \{\tilde{\pi}_1 = (\tilde{\pi}_{1,1}, \dots, \tilde{\pi}_{1,p})' \in R_{[+\infty]}^p : \text{for } j = 1, \dots, p, \tilde{\pi}_{1,j} = \infty \text{ or } \varphi_j(\xi, \Omega) \rightarrow \varphi_j((\tilde{\pi}_1, 0_v), \Omega_0) \text{ as } (\xi, \Omega) \rightarrow ((\tilde{\pi}_1, 0_v), \Omega_0)\}$. Roughly speaking, $C(\varphi)$ is the set of $\tilde{\pi}_1$ vectors for which φ is continuous at $((\tilde{\pi}_1, 0_v), \Omega_0)$. For example, $C(\varphi^{(1)}) = \{\tilde{\pi}_1 \in R_{[+\infty]}^p : \tilde{\pi}_{1,j} \neq 1 \text{ for } j = 1, \dots, p\}$, $C(\varphi^{(2)}) = R_{[+\infty]}^p$ provided ψ is continuous on $[a_L, a_U]$ (where continuity at a_U means that $\lim_{x \rightarrow a_U} \psi(x) = \infty$), $C(\varphi^{(3)}) = R_{[+\infty]}^p$, $C(\varphi^{(4)}) = R_{[+\infty]}^p$, and $C(\varphi^{(5)}) = \{\pi_1 \in R_{[+\infty]}^p : S(-c \cdot (\tilde{\pi}_1, 0_v), \Omega_0) - \zeta(|c|) \text{ has a unique minimum over } c \in \mathcal{C}\}$.

ASSUMPTION LA5: (a) $\pi_1 \in C(\varphi)$.

(b) The df of $S(\Omega_0^{1/2} Z^* + \varphi((\pi_1, 0_v), \Omega_0), \Omega_0)$ is continuous and strictly increasing at $x = c_{\pi_1}(\varphi, 1 - \alpha)$.

Assumption **LA5(a)** implies that the $n^{-1/2}$ -local power formulae given below do not apply to certain “discontinuity vectors” $\pi_1 = (\pi_{1,1}, \dots, \pi_{1,p})'$. However, this does not affect the power comparisons between GMS, subsampling, and PA tests, because Assumption **LA5** is not needed for those results. The power comparisons hold for all π_1 vectors.

We now introduce an assumption that is used for subsampling tests only.

ASSUMPTION LA6: $b^{1/2}E_{F_n}m_j(W_i, \theta_n)/\sigma_{F_n,j}(\theta_n) \rightarrow g_{1,j}$ for some $g_{1,j} \in R_{+, \infty}$ for $j = 1, \dots, p$.

Assumption **LA6** is not restrictive. It specifies the limit of the (normalized) moment inequality functions when evaluated at the true parameter values $\{\theta_n : n \geq 1\}$ and when scaled by the square root of the subsample size $b^{1/2}$.

Define $g_1 = (g_{1,1}, \dots, g_{1,p})'$. Note that $0_p \leq g_1 \leq \pi_1 \leq h_1$.¹⁰ The probability limit of the subsampling critical value is shown to depend on

$$(8.6) \quad \lim_{n \rightarrow \infty} b^{1/2}D^{-1/2}(\theta_{n,*}, F_n)E_{F_n}m(W_i, \theta_{n,*}) = (g_1, 0_v) \in R_{+, \infty}^k.$$

Note that $(g_1, 0_v) \in R_{+, \infty}^p \times \{0_v\}$. Thus, elements of $(g_1, 0_v)$ are necessarily non-negative. The probability limit of the subsampling critical value is shown to be $c_{g_1, 0_d}(1 - \alpha)$, which denotes the $1 - \alpha$ quantile of $J_{g_1, 0_d}$ (where $J_{g_1, 0_d}$ equals $J_{h_1, \lambda}$ with $h_1 = g_1$ and $\lambda = 0$). The probability limit of the PA critical value is shown to be $c_{0_p, 0_d}(1 - \alpha)$, which is the $1 - \alpha$ quantile of $J_{0_p, 0_d}$ (and also can be written as $c(\Omega_0, 1 - \alpha)$ using the notation introduced just above (7.1)).

THEOREM 2: Under Assumptions 1–5, **LA1**, and **LA2**, the following statements hold:

- (a) $\lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1 - \alpha)) = 1 - J_{h_1, \lambda}(c_{\pi_1}(\varphi, 1 - \alpha))$ provided Assumptions **GMS2**, **GMS3**, **LA4**, and **LA5** hold.
- (b) $\lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1 - \alpha)) = 1 - J_{h_1, \lambda}(c_{g_1, 0_d}(1 - \alpha))$ provided Assumption **LA6** holds.
- (c) $\lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1 - \alpha)) = 1 - J_{h_1, \lambda}(c_{0_p, 0_d}(1 - \alpha))$.

COMMENTS: (i) Theorem 2(a) provides the $n^{-1/2}$ -local alternative power function of the GMS test. The probability limit of the GMS critical value $\widehat{c}_n(\theta_{n,*}, 1 - \alpha)$ under $n^{-1/2}$ -local alternatives is $c_{\pi_1}(\varphi, 1 - \alpha)$. Theorem 2(b) and (c) provide the $n^{-1/2}$ -local alternative power function of the subsampling and PA tests.

(ii) The results of Theorem 2 hold under the null hypothesis as well as under the alternative. The results under the null quantify the degree of asymptotic nonsimilarity of the GMS, subsampling, and PA tests.

The next result provides power comparisons of GMS, subsampling, and PA tests.

THEOREM 3: Under Assumptions 1–5, **LA1–LA4**, **LA6**, **GMS2**, **GMS3**, **GMS5**, and **GMS6**, the following statements hold:

- (a) $\liminf_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1 - \alpha)) \geq \lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1 - \alpha))$ with strict inequality whenever $g_{1,j} < \infty$ and $\pi_{1,j} = \infty$ for some $j = 1, \dots, p$ and $c_{g_1, 0_d}(1 - \alpha) > 0$.

¹⁰This holds by condition (ii) of (2.2) (since $(\theta_n, F_n) \in \mathcal{F}$), Assumptions **LA1(b)**, **LA6**, and **GMS5**, and $b/n \rightarrow 0$.

- (b) $\liminf_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1 - \alpha)) \geq \lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1 - \alpha))$ with strict inequality whenever $\pi_{1,j} = \infty$ for some $j = 1, \dots, p$.
- (c) $\lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1 - \alpha)) \geq \lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1 - \alpha))$ with strict inequality whenever $g_1 > 0_p$, where Assumptions **GMS2**, **GMS3**, **GMS5**, **GMS6**, and **LA4** are not needed for this result.

COMMENTS: (i) Theorem 3(a) and (b) show that a GMS test based on a given test statistic has asymptotic power greater than or equal to that of subsampling and PA tests based on the same test statistic. For GMS versus subsampling tests, the inequality is strict whenever one or more moment inequality is satisfied and has a magnitude that is $o(b^{-1/2})$, and is larger than $O(\kappa_n n^{-1/2})$ and $c_{g_1, 0_d}(1 - \alpha) > 0$.¹¹ For GMS versus PA tests, the inequality is strict whenever one or more moment inequality is satisfied and has a magnitude that is larger than $O(\kappa_n n^{-1/2})$.

The reason the GMS test has higher power in these cases is that its (data-dependent) critical value is smaller asymptotically than the subsampling and PA critical values. It is smaller because when some moment inequality is satisfied under the alternative and is sufficiently far from being an equality (specifically, is larger than $O(\kappa_n n^{-1/2})$), then the GMS critical value takes this into account and delivers a critical value that is suitable for the case where this moment inequality is omitted. On the other hand, in the scenarios specified, the subsampling critical value does not take this into account, and in all scenarios the PA critical value is based on the least favorable distribution (for given Ω_0) which occurs when all moment inequalities hold as equalities.

(ii) Theorem 3(c) shows that the subsampling test has asymptotic power greater than or equal to that of the PA test for all local alternatives and is more powerful asymptotically than the PA test for many local alternatives. The reason is that when some moment inequality is satisfied under the alternative and is sufficiently far from being an equality (specifically, is larger than $o(b^{-1/2})$), then the subsampling critical value automatically takes this (at least partially) into account and delivers a smaller critical value than the PA critical value.

(iii) The comparison of the power of GMS tests and subsampling tests given in Theorem 3(a) does not impose Assumption **LA5**. Hence, the comparison holds for all $n^{-1/2}$ -local alternatives.

(iv) We now show that the difference in power between the GMS test and the subsampling and PA tests can be quite large. Suppose there are no equality constraints (i.e., $v = 0$) and the distribution considered is such that the first inequality constraint may or may not be violated, but the other $j = 2, \dots, p$ inequality constraints are not violated and differ from being equalities by mag-

¹¹For most test functions S , $c_{g_1, 0_d}(1 - \alpha) > 0$ whenever one or more of the moment inequalities is violated asymptotically, so the latter condition holds under local alternatives.

nitudes that are $o(b^{-1/2})$ and are larger than $O(\kappa_n n^{-1/2})$. In this case, $g_{1,j} = 0$, $h_{1,j} = \pi_{1,j} = \infty$, and $h_{1,j} + \Pi'_{0,j}\lambda = \infty$ for $j = 2, \dots, p$. Let $\mu_1 = h_{1,1} + \Pi'_{0,1}\lambda$. If $\mu_1 \in (-\infty, 0)$, the first inequality constraint is violated asymptotically. If the null hypothesis is true (for all n large), then $\theta_n = \theta_{n,*}$, $\lambda = 0$, and $\mu_1 = h_{1,1} \geq 0$. Since $|\mu_1| < \infty$, we have $|h_{1,1}| < \infty$ and $g_{1,1} = 0$. Thus, $g_1 = 0_p$. For simplicity, suppose $\Omega_0 = I_p$. In this case, the asymptotic powers of the tests based on the functions S_1 and S_2 are the same, so we consider the S_1 test statistic. The asymptotic distribution $J_{h_1, \lambda}$ in this case is the distribution of

$$(8.7) \quad \sum_{j=1}^p [Z_j^* + h_{1,j} + \Pi'_{0,j}\lambda]_-^2 = [Z_1^* + \mu_1]_-^2,$$

where $Z^* = (Z_1^*, \dots, Z_p^*)' \sim N(0_p, I_p)$, because $Z_j^* + \infty = \infty$ for $j = 2, \dots, p$.

The probability limit of the GMS critical value, $c_{\pi_1}(\varphi, 1 - \alpha)$, is the $1 - \alpha$ quantile of $[Z_1^*]_-^2$ which equals $z_{1-\alpha}^2$, where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of a standard normal distribution. This holds using (8.5) because $\pi_{1,1} = 0$ and Assumption **GMS1(b)** imply that $\varphi_1((\pi_1, 0_v), \Omega_0) = 0$, and for $j = 2, \dots, p$, $\pi_{1,j} = \infty$ and Assumption **GMS3** imply that $\varphi_j((\pi_1, 0_v), \Omega_0) = \infty$. On the other hand, $J_{g_{1,0_d}} = J_{0_p, 0_d}$ is the distribution of $\sum_{j=1}^p [Z_j^*]_-^2$. Hence, the probability limit of the subsampling and PA critical values, $c_{0_p, 0_d}(1 - \alpha)$, is the $1 - \alpha$ quantile of $\sum_{j=1}^p [Z_j^*]_-^2$, call it $z_\alpha(p)$. Clearly, $z_\alpha(1) = z_{1-\alpha}^2$, $z_\alpha(p) > z_{1-\alpha}^2$ for $p \geq 2$, and the difference is strictly increasing in p .

Table I provides the value of $z_\alpha(p)$ for $\alpha = .05$ and several values of p . One sees that the critical value of the subsampling and PA tests increases substantially as the number of nonviolated moment inequalities, $p - 1$, increases. Just one nonviolated moment inequality (i.e., $p = 2$) increases the critical value from 2.71 to 4.25.

TABLE I
ASYMPTOTIC CRITICAL VALUES AND POWER OF THE NOMINAL .05
GMS TEST COMPARED TO SUBSAMPLING AND PA TESTS

		μ_1			
		Critical Values	Asy. Null		
		$z_\alpha(p)$	Rej. Prob.	.00	-1.645
					-2.170
					-2.930
GMS test	All p	2.71	.050	.50	.70
Sub & PA tests	2	4.25	.020	.34	.54
	3	5.43	.010	.25	.44
	4	6.34	.005	.18	.35
	5	7.49	.003	.14	.29
	10	11.83	.000	.04	.10
	20	19.28	.000	.00	.01
					.07

By Theorem 2, the asymptotic powers of the GMS, subsampling, and PA tests in the present scenario are

$$(8.8) \quad \begin{aligned} \text{AsyPow}_{\text{GMS}}(\mu_1) &= P([Z_1^* + \mu_1]_-^2 > z_{1-\alpha}^2) = \Phi(-\mu_1 - z_{1-\alpha}), \\ \text{AsyPow}_{\text{Sub}}(\mu_1) &= \text{AsyPow}_{\text{PA}}(\mu_1) \\ &= P([Z_1^* + \mu_1]_-^2 > z_\alpha(p)) = \Phi(-\mu_1 - z_\alpha^{1/2}(p)), \end{aligned}$$

respectively. Table I reports the asymptotic power of the GMS test and the subsampling and PA tests, where the power of the latter depends on p , for four values of μ_1 . (Note that the asymptotic size of each test is α , so that no asymptotic size correction is needed before comparing the asymptotic powers of the tests.) The first value of μ_1 is zero, which corresponds to a distribution in the null hypothesis. In this case, the asymptotic rejection rate of the GMS test is precisely .05, while that of the subsampling and PA tests is much less than .05 due to the asymptotic nonsimilarity of these tests. The last three values of μ_1 are negative, which correspond to distributions in the alternative. Table I shows that the power of the GMS test is substantially higher than that of the subsampling and PA tests even when $p = 2$ and the difference increases with p .

(v) The difference in powers of the subsampling and PA tests can be as large as the differences illustrated in Table I between GMS and PA tests. Consider the same scenario as in comment (iv) except that the $j = 2, \dots, p$ inequality constraints differ from being equalities by a magnitude that is greater than $O(b^{-1/2})$. In this case, $g_{1,j} = \infty$ for $j = 2, \dots, p$ and $J_{g_{1,0_d}}$ is the distribution of $[Z_1^*]_-^2$ because $g_1 = (0, \infty, \dots, \infty)'$. Hence, the probability limit of the subsampling critical value, $c_{g_{1,0_d}}(1 - \alpha)$, equals that of the GMS critical value and $\text{AsyPow}_{\text{Sub}}(\mu_1) = \text{AsyPow}_{\text{GMS}}(\mu_1)$. Everything else is the same as in comment (iv). Hence, in the present scenario, Table I applies but with the results for the subsampling test given by those of the GMS test.

(vi) The GMS, subsampling, and PA tests are not asymptotically unbiased. That is, there exist local alternatives for which the asymptotic rejection probabilities of the tests, namely $1 - J_{h_1,\lambda}(c_{\pi_1}(\varphi, 1 - \alpha))$, $1 - J_{h_1,\lambda}(c_{g_{1,0_d}}(1 - \alpha))$, and $1 - J_{h_1,\lambda}(c_{0_p,0_d}(1 - \alpha))$, respectively, are less than α (e.g., see Table I with $p = 10$ or 20). This occurs because these tests are not asymptotically similar on the boundary of the null hypothesis. Lack of asymptotic unbiasedness is a common feature of tests of multivariate one-sided hypotheses, so this property of GMS, subsampling, and PA tests in the moment inequality example is not surprising.

(vii) Rosen (2008) introduced a critical value method that is a variant of the PA critical value. His method has the advantage of being simple computationally. However, it sacrifices power relative to GMS critical values in two respects. First, an upper bound on the $1 - \alpha$ quantile of the asymptotic null distribution is employed. Second, in models in which some moment inequality can be slack without another being binding, his procedure yields larger critical

values than GMS critical values because it does not use the data to detect slack inequalities. His procedure only adjusts for slack moment inequalities when it is known that if some inequality is binding, then some other necessarily cannot be.

(viii) For moment conditions based on weak instruments, the results of Theorem 2 still hold. But, the power comparisons of Theorem 3 do not because $\Pi'_{0,j}\lambda = 0$ for all $j = 1, \dots, k$ in this case and so Assumption LA3 does not hold. With weak instruments, all of the tests have power less than or equal to α against $n^{-1/2}$ -local alternatives, as is expected.

9. POWER AGAINST DISTANT ALTERNATIVES

Next we consider power against alternatives that are more distant from the null than $n^{-1/2}$ -local alternatives. For all such alternatives, the powers of GMS, subsampling, and PA tests are shown to converge to 1 as $n \rightarrow \infty$. Thus, all three tests are consistent tests.

The following assumption specifies the properties of “distant alternatives” (DA), which include fixed alternatives and local alternatives that deviate from the null hypothesis by more than $O(n^{-1/2})$. Define

$$(9.1) \quad m_{n,j}^* = E_{F_n} m_j(W_i, \theta_{n,*}) / \sigma_{F_n,j}(\theta_{n,*}), \\ \beta_n = \max\{-m_{n,1}^*, \dots, -m_{n,p}^*, |m_{n,p+1}^*|, \dots, |m_{n,k}^*|\}.$$

ASSUMPTION DA: (a) $n^{1/2}\beta_n \rightarrow \infty$.

(b) $\Omega(\theta_{n,*}, F_n) \rightarrow \Omega_1$ for some $k \times k$ correlation matrix $\Omega_1 \in \Psi$.

Assumption DA(a) requires that some moment inequality term $m_{n,j}^*$ violates the nonnegativity condition and is not $o(n^{-1/2})$ for $j = 1, \dots, p$ or some moment equality term $m_{n,j}^*$ violates the zero condition and is not $o(n^{-1/2})$ for $j = p + 1, \dots, k$. In contrast to Assumption DA, under Assumptions LA1–LA3 above, $n^{1/2}\beta_n \rightarrow \max\{-h_{1,1} - \Pi'_{0,1}\lambda, \dots, -h_{1,p} - \Pi'_{0,p}\lambda, |\Pi'_{0,p+1}\lambda|, \dots, |\Pi'_{0,k}\lambda|\} < \infty$.

As in Section 8, we consider i.i.d. observations and \mathcal{F} satisfies (2.2).

THEOREM 4: *Under Assumptions 1, 3, 6, and DA, we can make the following assertions:*

- (a) $\lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > \widehat{c}_n(\theta_{n,*}, 1 - \alpha)) = 1$ provided Assumption GMS7 holds.
- (b) $\lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c_{n,b}(\theta_{n,*}, 1 - \alpha)) = 1$.
- (c) $\lim_{n \rightarrow \infty} P_{F_n}(T_n(\theta_{n,*}) > c(\widehat{\Omega}_n(\theta_{n,*}), 1 - \alpha)) = 1$.

COMMENT: Theorem 4 shows that GMS, subsampling, and PA tests are consistent against all fixed alternatives and all non- $n^{-1/2}$ -local alternatives.

10. EXTENSIONS

10.1. Generalized Empirical Likelihood Statistics

We now discuss CS's based on generalized empirical likelihood (GEL) test statistics. For definitions and regularity conditions concerning GEL test statistics, see AG4. The asymptotic distribution of a GEL test statistic (under any drifting sequence of parameters) is the same as that of the QLR test statistic; see AG4 for a proof. Given the structure of the proofs below, this implies that all of the asymptotic results stated above for QLR tests also hold for GEL tests.

Specifically, under the assumptions of Theorems 1–4, we have (i) GEL CS's based on GMS critical values have correct size asymptotically, (ii) GEL tests based on GMS critical values have asymptotic power greater than or equal to that of GEL tests based on subsampling or PA critical values with strictly greater power in certain scenarios, and (iii) the “pure” GEL test that uses a constant critical value (equal to $c_{\text{GEL}}(1 - \alpha) = \sup_{\Omega \in \Psi_2} c(\Omega, 1 - \alpha)$, where $c(\Omega, 1 - \alpha)$ is as defined above using the function S_2) is dominated asymptotically by various alternative tests. Such tests include tests constructed from a GEL or QLR test statistic combined with GMS, subsampling, or PA critical values. The results of (iii) indicate that there are notable drawbacks to the asymptotic optimality criteria based on large deviation probabilities considered by Otsu (2006) and Canay (2007).

10.2. Preliminary Estimation of Identified Parameters

Here we consider the case where the moment functions in (2.2) depend on a parameter τ (i.e., are of the form $\{m_j(W_i, \theta, \tau) : j \leq k\}$), and a preliminary consistent and asymptotically normal estimator $\widehat{\tau}_n(\theta_0)$ of τ exists when θ_0 is the true value of θ . This requires that τ is identified. The sample moment functions in this case are of the form $\overline{m}_{n,j}(\theta) = \overline{m}_{n,j}(\theta, \widehat{\tau}_n(\theta))$. The asymptotic variance of $n^{1/2}\overline{m}_{n,j}(\theta)$ is different when τ is replaced by the estimator $\widehat{\tau}_n(\theta)$ and so $\widehat{\Sigma}_n(\theta)$ needs to be defined accordingly, but otherwise the theoretical treatment of this model is the same as that given above. In fact, Theorem 1 holds in this case using the conditions given in (A.3) of the [Appendix](#). These are high-level conditions that essentially just require that $n^{-1} \sum_{i=1}^n m_{n,j}(W_i, \theta, \widehat{\tau}_n(\theta))$ is asymptotically normal (after suitable normalization).

Furthermore, the power comparisons in Section 8, which are stated for i.i.d. observations and no preliminary estimated parameters, can be extended to the case of preliminary estimated parameters. Thus, in this case too, GMS tests have power advantages over subsampling, and PA tests and subsampling tests have power advantages over PA tests.

11. MONTE CARLO EXPERIMENT

11.1. Experimental Design

In this section, we provide finite-sample comparisons of the maximum null rejection probability (MNRP) over different null mean vectors and MNRP-corrected power of GMS, recentered subsampling, and PA tests.¹² (MNRP is defined precisely below.) For each test, we consider the QLR test statistic. For the GMS tests, we use the $\varphi^{(1)}$ critical value function, which yields the t -test selection method. We use two values of κ_n , the BIC value $\kappa_n = (\ln n)^{1/2}$ and the law of the iterated logarithm (LIL) value $\kappa_n = (2 \ln \ln n)^{1/2}$, which yield $\kappa_n = 2.35$ and $\kappa_n = 1.85$, respectively, when $n = 250$, which is the sample size considered here. We provide results for the bootstrap and asymptotic versions of the GMS test. The GMS tests are denoted by GMS/Boot1 and GMS/Asy1, which use the BIC value of κ_n , and GMS/Boot2 and GMS/Asy2, which use the LIL value of κ_n . (The focus on $(S_2, \varphi^{(1)})$ is based on results in Andrews and Jia (2008) that compare different choices of (S, φ) in terms of asymptotic size and power.)

The subsampling and PA tests considered here also employ the QLR test statistic. Hence, the tests differ only in the way in which the critical value is calculated. Bugni (2007a, 2007b) showed that taking b of order $n^{2/3}$ minimizes the error in the null rejection probability for the recentered subsampling test. In consequence, we use subsample sizes $b = .75n^{2/3}$, $n^{2/3}$, and $1.25n^{2/3}$, which for $n = 250$ yields $b = 30$, 40, and 50, respectively. These subsampling tests are denoted Sub1, Sub2, and Sub3, respectively.

We consider the case in which no equalities arise (i.e., $v = 0$) and the number of inequalities, p , is 2, 4, or 10. For given θ , the null hypothesis is $H_0: Em(W_i, \theta) \geq 0_p$ for some given moment functions $m(W_i, \theta)$ and the alternative hypothesis is that H_0 does not hold. We consider a general formulation of the testing problem of interest which does not require the specification of a particular form for $m(W_i, \theta)$, as in Andrews and Jia (2008). The finite-sample properties of tests of H_0 depend on $m(W_i, \theta)$ only through (i) $\mu = Em(W_i, \theta)$, (ii) $\Omega = \text{Corr}(m(W_i, \theta))$, and (iii) the distribution of the mean zero, variance I_p random vector $Z^\dagger = \text{Var}^{-1/2}(m(W_i, \theta))(m(W_i, \theta) - Em(W_i, \theta))$. We consider the case in which $Z^\dagger \sim N(0_p, I_p)$. We consider three representative correlation matrices Ω_{Neg} , Ω_{Zero} , and Ω_{Pos} , which exhibit negative, zero, and positive correlations, respectively. By definition, MNRP denotes the maximum null rejection probability over mean vectors in H_0 given the correlation matrix Ω_{Neg} , Ω_{Zero} , or Ω_{Pos} and under the assumption of normally-distributed moment inequalities.¹³

¹²We consider recentered subsampling tests because non-recentered subsampling tests are found to underreject the null hypothesis substantially for sample sizes of 250 and 1000. Even for a sample size of 5000, there is some underrejection. For details, see the Supplemental Material.

¹³The MNRP of a test is the same as the size of the test except that the MNRP is for a fixed correlation matrix and distribution—in the present case a normal distribution—whereas the size is given by the maximum over all allowable correlation matrices and distributions.

Specifically, Ω_{Zero} equals I_p for $p = 2, 4$, and 10 . The matrices Ω_{Neg} and Ω_{Pos} are Toeplitz matrices with correlations on the diagonals given by a $p - 1$ vector ρ . For $p = 2$: $\rho = -.9$ for Ω_{Neg} and $\rho = .5$ for Ω_{Pos} . For $p = 4$, $\rho = (-.9, .7, -.5)$ for Ω_{Neg} and $\rho = (.9, .7, .5)$ for Ω_{Pos} . For $p = 10$, $\rho = (-.9, .8, -.7, .6, -.5, .4, -.3, .2, -.1)$ for Ω_{Neg} and $\rho = (.9, .8, .7, .6, .5, \dots, .5)$ for Ω_{Pos} .

Note that the finite-sample testing problem for *any* moment inequality model fits into the framework above for some correlation matrix Ω and some distribution of Z^\dagger . In large samples, the impact of the distribution of Z^\dagger vanishes because of the central limit theorem (CLT).

For all of the tests considered below, calculations for a subset of the cases considered show without exception that the maximum null rejection probabilities occur for mean vectors μ whose elements are 0's and ∞ 's. In consequence, the MNRP results are obtained by computing the maximum null rejection probabilities over μ vectors with this form.

For the power comparisons given here, we compare MNRP-corrected power. By this we mean that the critical values are adjusted by a constant so that the MNRP equals the nominal level .05.

The power comparisons are made based on average power over certain sets, $\mathcal{M}_p(\Omega)$, of vectors μ in the alternative (i.e., $\mu \not\geq 0_p$). For $p = 2$, the set of μ vectors $\mathcal{M}_2(\Omega)$ includes seven elements and is of the form $\mathcal{M}_2(\Omega) = \{(-\mu_1, 0), (-\mu_2, 1), (-\mu_3, 2), (-\mu_4, 3), (-\mu_5, 4), (-\mu_6, 7), (-\mu_7, -\mu_7)\}$, where $\mu_j > 0$ depends on Ω for $j = 1, \dots, 7$ and is such that the finite-sample power envelope (for known Ω) is .73 at each $\mu \in \mathcal{M}_2(\Omega)$ (see Andrews and Jia (2008) for more details). For brevity, the values of μ_j are given in the Supplemental Material. For $p = 4$ and 10 , $\mathcal{M}_4(\Omega)$ includes 24 and 40 elements, respectively. For brevity, the complete specifications of $\mathcal{M}_4(\Omega)$ and $\mathcal{M}_{10}(\Omega)$ are given in the Supplemental Material. The sets $\mathcal{M}_4(\Omega)$ and $\mathcal{M}_{10}(\Omega)$ are defined such that the finite-sample power envelope (for known Ω) is (approximately) .79 and .84, respectively, at each $\mu \in \mathcal{M}_p(\Omega)$ for $p = 4$ and 10 .

The simulation results are based on 2500 repetitions for the calculation of the GMS, subsampling, and PA critical values, 2500 simulation repetitions for the finite-sample MNRP results, and 1000 simulation repetitions for the finite-sample MNRP-corrected power results.

11.2. Simulation Results

Table II provides the MNRP and power results. The table shows that the GMS/Boot1 and GMS/Asy1 tests have better MNRP and power properties than the GMS/Boot2 and GMS/Asy2 tests. The GMS1 tests have good MNRPs in all cases. For example, the GMS/Boot1 test has MNRP in the interval [.048, .065] for all cases considered. The GMS2 tests tend to overreject the null somewhat with Ω_{Neg} . The GMS/Boot1 test has slightly higher power than the GMS/Asy1 test. Hence, the GMS/Boot1 test performs the best of the GMS tests in terms of MNRP and power by a slight margin over the GMS/Asy1 test.

TABLE II

FINITE-SAMPLE MNRPs AND MNP-CORRECTED POWER OF NOMINAL .05 TESTS BASED ON THE QLR TEST STATISTIC COMBINED WITH GMS, RECENTERED SUBSAMPLING (SUB), AND PA CRITICAL VALUES FOR SAMPLE SIZE $n = 250$

Number of Moment Inequalities	Critical Value	Ω_{Neg}		Ω_{Zero}		Ω_{Pos}	
		MNRP	Avg. Power	MNRP	Avg. Power	MNRP	Avg. Power
2	GMS/Boot1	.054	.66	.048	.71	.048	.73
	GMS/Asy1	.047	.66	.041	.71	.046	.72
	GMS/Boot2	.073	.61	.052	.71	.048	.73
	GMS/Asy2	.066	.61	.043	.70	.046	.73
	PA	.040	.56	.039	.61	.046	.66
	Sub1	.050	.60	.050	.65	.051	.68
	Sub2	.061	.60	.061	.65	.060	.67
	Sub3	.071	.60	.068	.65	.068	.68
	P. Envelope	—	.73	—	.73	—	.73
4	GMS/Boot1	.065	.58	.051	.68	.051	.76
	GMS/Asy1	.064	.57	.052	.66	.052	.76
	GMS/Boot2	.080	.53	.053	.69	.051	.76
	GMS/Asy2	.079	.53	.055	.67	.044	.76
	PA	.050	.43	.045	.52	.045	.69
	Sub1	.046	.43	.047	.53	.051	.71
	Sub2	.060	.43	.062	.53	.062	.71
	Sub3	.077	.43	.075	.53	.068	.71
	P. Envelope	—	.79	—	.79	—	.77
10	GMS/Boot1	.059	.58	.050	.65	.051	.78
	GMS/Asy1	.064	.56	.054	.63	.048	.78
	GMS/Boot2	.075	.54	.051	.67	.051	.79
	GMS/Asy2	.083	.52	.058	.64	.048	.79
	PA	.055	.29	.055	.37	.043	.66
	Sub1	.010	.27	.017	.36	.049	.68
	Sub2	.030	.27	.040	.35	.058	.69
	Sub3	.052	.28	.060	.35	.067	.69
	P. Envelope	—	.85	—	.84	—	.83

The power of the GMS/Boot1 test is substantially greater than that of the PA test. The relative advantage is increasing in p . The average power differences for $p = 2, 4$, and 10 are .09, .13, and .23, respectively, where the average is over Ω_{Neg} , Ω_{Zero} , and Ω_{Pos} .

The best subsampling test in terms of MNRP is Sub1. The MNRP's of Sub1 are quite good except for $p = 10$ with Ω_{Neg} and Ω_{Zero} , in which case Sub1 dramatically underrejects with MNRP's of .010 and .017. The MNRP-corrected power of Sub1, Sub2, and Sub3 is the same and hence does not depend on b .

The power gains of the GMS/Boot1 test over the subsampling tests are quite similar to those of the GMS/Boot1 test over the PA test, although they are

a bit smaller for $p = 2$. The power gains are substantial, especially for $p = 4, 10$. For example, in the most extreme case, where $p = 10$ and $\Omega = \Omega_{\text{Neg}}$, the GMS/Boot1 and Sub1 tests have power .58 and .27, respectively.

The differences between the power of the GMS/Boot1 test and the power envelope increase quickly in p . This is because the GMS/Boot1 test is a p -directional test, whereas the power envelope is attained by a unidirectional test. The differences are small for $p = 2$, but quite large for $p = 10$ when $\Omega = \Omega_{\text{Neg}}$ and Ω_{Pos} . The differences in power are decreasing as one moves from Ω_{Neg} to Ω_{Zero} to Ω_{Pos} . Even for $p = 10$, the difference for Ω_{Pos} is only .05, which is remarkably small.

In conclusion, the finite-sample simulations reported here indicate that the GMS/Boot1 test has good MNRP for the cases considered and good power relative to the PA and subsampling tests that are considered. In consequence, the GMS/Boot1 test is the recommended test. It is the bootstrap version of the GMS test based on the QLR test statistic, the t -test moment selection critical value, and the tuning parameter $\kappa_n = (\ln n)^{1/2}$.

APPENDIX

In this [Appendix](#), we start by stating some assumptions on the test statistic function S . Next, we give an alternative parametrization of the moment inequality/equality model to that of Section 2. The new parametrization is conducive to the calculation of the asymptotic properties of CS's and tests. It was first used in AG4. We also specify the parameter space for the case of dependent observations. Proofs of the results of the paper are given in the Supplemental Material (Andrews and Soares (2010)).

A.1. Test Statistic Assumptions

The following assumptions concern the test statistic function S .

ASSUMPTION 2: For all $h_1 \in R_{+, \infty}^p$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the df of $S(Z + (h_1, 0_v), \Omega)$ at $x \in R$ is (a) continuous for $x > 0$, (b) strictly increasing for $x > 0$ unless $v = 0$ and $h_1 = \infty^p$, and (c) less than or equal to 1/2 at $x = 0$ whenever $v \geq 1$ or $h_1 = 0_p$.

ASSUMPTION 4: (a) The df of $S(Z, \Omega)$ is continuous at its $1 - \alpha$ quantile, $c(\Omega, 1 - \alpha)$, for all $\Omega \in \Psi$, where $Z \sim N(0_k, \Omega)$ and $\alpha \in (0, 1/2)$.

(b) $c(\Omega, 1 - \alpha)$ is continuous in Ω uniformly for $\Omega \in \Psi$.

ASSUMPTION 5: (a) For all $\ell \in R_{[+\infty]}^p \times R^v$, all $\Omega \in \Psi$, and $Z \sim N(0_k, \Omega)$, the df of $S(Z + \ell, \Omega)$ at x is (i) continuous for $x > 0$ and (ii) unless $v = 0$ and $\ell = \infty^p$, strictly increasing for $x > 0$.

(b) $P(S(Z + (m_1, 0_v), \Omega) \leq x) < P(S(Z + (m_1^*, 0_v), \Omega) \leq x)$ for all $x > 0$ for all $m_1, m_1^* \in R_{+, \infty}^p$ with $m_1 < m_1^*$.

For $(\theta, F) \in \mathcal{F}$, define $h_{1,j}(\theta, F) = \infty$ if $E_F m_j(W_i, \theta) > 0$ and $h_{1,j}(\theta, F) = 0$ if $E_F m_j(W_i, \theta) = 0$ for $j = 1, \dots, p$. Let $h_1(\theta, F) = (h_{1,1}(\theta, F), \dots, h_{1,p}(\theta, F))'$ and $\Omega(\theta, F) = \lim_{n \rightarrow \infty} \text{Corr}_F(n^{1/2} \bar{m}_n(\theta))$.

ASSUMPTION 7: For some $(\theta, F) \in \mathcal{F}$, the df of $S(Z + (h_1(\theta, F), 0_v), \Omega(\theta, F))$ is continuous at its $1 - \alpha$ quantile, where $Z \sim N(0_k, \Omega(\theta, F))$.

In Assumption 2, if an element of h_1 equals $+\infty$, then by definition the corresponding element of $Z + (h_1, 0_v)$ equals $+\infty$.

Assumption 2 is used to show that certain asymptotic df's satisfy suitable continuity/strictly increasing properties. These properties ensure that the GMS critical value converges in probability to a constant and the CS has asymptotic size that is not affected by a jump in a df. Assumption 4 is a mild continuity assumption. Assumption 5 is used for the $n^{-1/2}$ -local power results. Assumption 5(a) is a continuity/strictly increasing df condition that is the same as Assumption 2(a) except that ℓ can take negative values. Assumption 5(b) is a stochastically strictly increasing condition. With a nonstrict inequality, it is implied by Assumption 1(a). Assumption 7 is used to show that GMS CS's are not asymptotically conservative (i.e., $\text{AsyCS} \not> 1 - \alpha$). It is a very weak continuity condition. If the $1 - \alpha$ quantile of $S(Z + (h_1(\theta, F), 0_v), \Omega(\theta, F))$ is positive for some $(\theta, F) \in \mathcal{F}$, which holds quite generally, Assumption 7 is implied by Assumption 2(a). For example, Assumption 7 holds for $S = S_1$ or S_2 whenever (i) $E_F m_j(W_i, \theta) = 0$ for some $j \leq p$ for some $(\theta, F) \in \mathcal{F}$ or (ii) $v \geq 1$ (which holds if an equality is present). It is hard to envision cases of interest where condition (i) fails.

A.2. Alternative Parametrization and Dependent Observations

In this section we specify a one-to-one mapping between the parameters (θ, F) with parameter space \mathcal{F} and a new parameter $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ with corresponding parameter space Γ . We define $\gamma_1 = (\gamma_{1,1}, \dots, \gamma_{1,p})' \in R_+^p$ by writing the moment inequalities in (2.1) as moment equalities:

$$(A.1) \quad \sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \quad \text{for } j = 1, \dots, p,$$

where $\sigma_{F,j}^2(\theta) = \text{AsyVar}_F(n^{1/2} \bar{m}_{n,j}(\theta))$ denotes the variance of the asymptotic distribution of $n^{1/2} \bar{m}_{n,j}(\theta)$ when the true parameter is θ and the true distribution of the data is F . Let $\Omega = \Omega(\theta, F) = \text{AsyCorr}_F(n^{1/2} \bar{m}_n(\theta))$, where $\text{AsyCorr}_F(n^{1/2} \bar{m}_n(\theta))$ denotes the correlation matrix of the asymptotic distribution of $n^{1/2} \bar{m}_n(\theta)$ when the true parameter is θ and the true distribution of the data is F . (We only consider (θ, F) for which these asymptotic variances and correlation matrices exist, see conditions (iv) and (v) of (A.2) below.) When no preliminary parameter τ is estimated $\sigma_{F,j}^2(\theta) = \lim_{n \rightarrow \infty} \text{Var}_F(n^{1/2} \bar{m}_{n,j}(\theta))$

and $\Omega(\theta, F) = \lim_{n \rightarrow \infty} \text{Corr}_F(n^{1/2}\bar{m}_n(\theta))$, where $\text{Var}_F(\cdot)$ and $\text{Corr}_F(\cdot)$ denote finite-sample variance and correlation under (θ, F) , respectively. Let $\gamma_2 = (\gamma_{2,1}, \gamma_{2,2}) = (\theta, \text{vech}_*(\Omega(\theta, F))) \in R^q$, where $\text{vech}_*(\Omega)$ denotes the vector of elements of Ω that lie below the main diagonal, $q = d + k(k-1)/2$, and $\gamma_3 = F$.

For the case described in Section 10.2 (where the sample moment functions depend on a preliminary estimator $\hat{\tau}_n(\theta)$ of an identified parameter vector τ_0), we define $m_j(W_i, \theta) = m_j(W_i, \theta, \tau_0)$, $m(W_i, \theta) = (m_1(W_i, \theta, \tau_0), \dots, m_k(W_i, \theta, \tau_0))'$, $\bar{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^n m_j(W_i, \theta, \hat{\tau}_n(\theta))$, and $\bar{m}_n(\theta) = (\bar{m}_{n,1}(\theta), \dots, \bar{m}_{n,k}(\theta))'$. (Hence, in this case, $\bar{m}_n(\theta) \neq n^{-1} \sum_{i=1}^n m(W_i, \theta)$.)

For i.i.d. observations (and no preliminary estimator $\hat{\tau}_n(\theta)$), the parameter space for γ is defined by $\Gamma = \{\gamma = (\gamma_1, \gamma_2, \gamma_3) : \text{for some } (\theta, F) \in \mathcal{F}, \text{ where } \mathcal{F} \text{ is defined in (2.2), } \gamma_1 \text{ satisfies (A.1), } \gamma_2 = (\theta, \text{vech}_*(\Omega(\theta, F))), \text{ and } \gamma_3 = F\}$.

For dependent observations and for sample moment functions that depend on preliminary estimators of identified parameters, we specify the parameter space Γ for the moment inequality model using a set of high-level conditions. To verify the high-level conditions using primitive conditions, one has to specify an estimator $\hat{\Sigma}_n(\theta)$ of the asymptotic variance matrix $\Sigma(\theta)$ of $n^{1/2}\bar{m}_n(\theta)$. For brevity, we do not do so here. Since there is a one-to-one mapping from γ to (θ, F) , Γ also defines the parameter space \mathcal{F} of (θ, F) . Let Ψ be a specified set of $k \times k$ correlation matrices. The parameter space Γ is defined to include parameters $\gamma = (\gamma_1, \gamma_2, \gamma_3) = (\gamma_1, (\theta, \gamma_{2,2}), F)$ that satisfy

- (A.2) (i) $\theta \in \Theta$,
- (ii) $\sigma_{F,j}^{-1}(\theta) E_F m_j(W_i, \theta) - \gamma_{1,j} = 0 \quad \text{for } j = 1, \dots, p$,
- (iii) $E_F m_j(W_i, \theta) = 0 \quad \text{for } j = p+1, \dots, k$,
- (iv) $\sigma_{F,j}^2(\theta) = \text{AsyVar}_F(n^{1/2}\bar{m}_{n,j}(\theta))$ exists and lies in $(0, \infty)$
for $j = 1, \dots, k$,
- (v) $\text{AsyCorr}_F(n^{1/2}\bar{m}_n(\theta))$ exists and equals $\Omega_{\gamma_{2,2}} \in \Psi$,
- (vi) $\{W_i : i \geq 1\}$ are stationary under F ,

where $\gamma_1 = (\gamma_{1,1}, \dots, \gamma_{1,p})'$ and $\Omega_{\gamma_{2,2}}$ is the $k \times k$ correlation matrix determined by $\gamma_{2,2}$.¹⁴ Furthermore, Γ must be restricted by enough additional conditions such that, under any sequence $\{\gamma_{n,h} = (\gamma_{n,h,1}, (\theta_{n,h}, \text{vech}_*(\Omega_{n,h})), F_{n,h}) : n \geq 1\}$ of parameters in Γ that satisfies $n^{1/2}\gamma_{n,h,1} \rightarrow h_1$ and $(\theta_{n,h}, \text{vech}_*(\Omega_{n,h})) \rightarrow h_2 =$

¹⁴In AG4, a strong mixing condition is imposed in condition (vi) of (A.2). This condition is used to verify Assumption E0 in that paper and is not needed with GMS critical values. To extend the subsampling power results of the paper to dependent observations, this assumption needs to be imposed.

$(h_{2,1}, h_{2,2})$ for some $h = (h_1, h_2) \in R_{+, \infty}^p \times R_{[\pm \infty]}^q$, we have

- (A.3) (vii) $A_n = (A_{n,1}, \dots, A_{n,k})' \rightarrow_d Z_{h_{2,2}} \sim N(0_k, \Omega_{h_{2,2}})$ as
 $n \rightarrow \infty$, where

$$A_{n,j} = n^{1/2} \left(\bar{m}_{n,j}(\theta_{n,h}) - n^{-1} \sum_{i=1}^n E_{F_{n,h}} m_j(W_i, \theta_{n,h}) \right) \\ / \sigma_{F_{n,h},j}(\theta_{n,h}),$$

- (viii) $\widehat{\sigma}_{n,j}(\theta_{n,h}) / \sigma_{F_{n,h},j}(\theta_{n,h}) \rightarrow_p 1$ as $n \rightarrow \infty$ for $j = 1, \dots, k$,
- (ix) $\widehat{D}_n^{-1/2}(\theta_{n,h}) \widehat{\Sigma}_n(\theta_{n,h}) \widehat{D}_n^{-1/2}(\theta_{n,h}) \rightarrow_p \Omega_{h_{2,2}}$ as $n \rightarrow \infty$,
- (x) conditions (vii)–(ix) hold for all subsequences $\{w_n\}$ in place of $\{n\}$,

where $\Omega_{h_{2,2}}$ is the $k \times k$ correlation matrix for which $\text{vech}_*(\Omega_{h_{2,2}}) = h_{2,2}$, $\widehat{\sigma}_{n,j}^2(\theta) = [\widehat{\Sigma}_n(\theta)]_{jj}$ for $1 \leq j \leq k$, and $\widehat{D}_n(\theta) = \text{Diag}\{\widehat{\sigma}_{n,1}^2(\theta), \dots, \widehat{\sigma}_{n,k}^2(\theta)\}$ ($= \text{Diag}(\widehat{\Sigma}_n(\theta))$).^{15,16}

For example, for i.i.d. observations, conditions (i)–(vi) of (2.2) imply conditions (i)–(vi) of (A.2). Furthermore, conditions (i)–(vi) of (2.2) plus the definition of $\widehat{\Sigma}_n(\theta)$ in (3.2) and the additional condition (vii) of (2.2) imply conditions (vii)–(x) of (A.3). For a proof, see Lemma 2 of AG4.

For dependent observations, one needs to specify a particular variance estimator $\widehat{\Sigma}_n(\theta)$ before one can specify primitive “additional conditions” beyond conditions (i)–(vi) in (A.2) that ensure that Γ is such that any sequences $\{\gamma_{n,h} : n \geq 1\}$ in Γ satisfy (A.3). For brevity, we do not do so here.

REFERENCES

- ANDREWS, D. W. K. (1999a): “Consistent Moment Selection Procedures for Generalized Method of Moments Estimation,” *Econometrica*, 67, 543–564. [120,132]
— (1999b): “Estimation When a Parameter Is on a Boundary,” *Econometrica*, 67, 1341–1383. [123,131]
— (2000): “Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space,” *Econometrica*, 68, 399–405. [123,132]

¹⁵When a preliminary estimator $\widehat{\tau}_n(\theta)$ appears, $A_{n,j}$ can be written equivalently as $n^{1/2}(n^{-1} \sum_{i=1}^n m_j(W_i, \theta_{n,h}, \widehat{\tau}_n(\theta_{n,h})) - n^{-1} \sum_{i=1}^n E_{F_{n,h}} m_j(W_i, \theta_{n,h}, \tau_0)) / \sigma_{F_{n,h},j}(\theta_{n,h})$, which typically is asymptotically normal with an asymptotic variance matrix $\Omega_{h_{2,2}}$ that reflects the fact that τ_0 has been estimated. When a preliminary estimator $\widehat{\tau}_n(\theta)$ appears, $\widehat{\Sigma}_n(\theta)$ needs to be defined to take account of the fact that τ_0 has been estimated. When no preliminary estimator $\widehat{\tau}_n(\theta)$ appears, $A_{n,j}$ can be written equivalently as $n^{1/2}(\bar{m}_{n,j}(\theta_{n,h}) - E_{F_{n,h}} \bar{m}_{n,j}(\theta_{n,h})) / \sigma_{F_{n,h},j}(\theta_{n,h})$.

¹⁶Condition (x) of (A.3) requires that conditions (vii)–(ix) must hold under any sequence of parameters $\{\gamma_{w_n,h} : n \geq 1\}$ that satisfies the conditions preceding (A.3) with n replaced by w_n .

- ANDREWS, D. W. K., AND P. GUGGENBERGER (2009a): "Hybrid and Size-Corrected Subsampling Methods," *Econometrica*, 77, 721–762. [125]
- (2009b): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709. [120]
- (2010a): "Applications of Subsampling, Hybrid, and Size-Correction Methods," *Journal of Econometrics* (forthcoming). [125]
- (2010b): "Asymptotic Size and a Problem With Subsampling and With the m Out of n Bootstrap," *Econometric Theory*, 26 (forthcoming). [123,125]
- ANDREWS, D. W. K., AND S. HAN (2009): "Invalidity of the Bootstrap and m Out of n Bootstrap for Interval Endpoints," *Econometrics Journal*, 12, S172–S199. [123]
- ANDREWS, D. W. K., AND P. JIA (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," Discussion Paper 1676, Cowles Foundation, Yale University. [122,127,128,149,150]
- ANDREWS, D. W. K., AND G. SOARES (2010): "Supplement to 'Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,'" *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/7502_Proofs.pdf; http://www.econometricsociety.org/ecta/Supmat/7502_data_and_programs.zip. [122,123,136,138,152]
- ANDREWS, D. W. K., S. BERRY, AND P. JIA (2004): "Confidence Regions for Parameters in Discrete Games With Multiple Equilibria, With an Application to Discount Chain Store Location," Unpublished Manuscript, Cowles Foundation, Yale University. [119,123,124]
- BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814. [121]
- BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2008): "Sharp Identification Regions in Games," CEMMAP Working Paper CWP15/08, Institute for Fiscal Studies, UCL. [123]
- BONTEMPS, C., T. MAGNAC, AND E. MAURIN (2007): "Set Identified Linear Models," Unpublished Manuscript, Toulouse School of Economics. [123]
- BUGNI, F. (2007a): "Bootstrap Inference in Partially Identified Models," Unpublished Manuscript, Department of Economics, Northwestern University. [121-123,134,137,149]
- (2007b): "Bootstrap Inference in Partially Identified Models: Pointwise Construction," Unpublished Manuscript, Department of Economics, Northwestern University. [121-123,134,137,149]
- CANAY, I. A. (2007): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," Unpublished Manuscript, Department of Economics, University of Wisconsin. [121-123,134,148]
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [119]
- CILIBERTO, F., AND E. TAMER (2009): "Market Structure and Multiple Equilibrium in Airline Markets," *Econometrica*, 77, 1791–1828. [119,124]
- FAN, Y., AND S. PARK (2007): "Confidence Sets for Some Partially Identified Parameters," Unpublished Manuscript, Department of Economics, Vanderbilt University. [123,134]
- GALICHON, A., AND M. HENRY (2009): "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics* (forthcoming). [123]
- GUGGENBERGER, P., J. HAHN, AND K. KIM (2008): "Specification Testing Under Moment Inequalities," *Economics Letters*, 99, 375–378. [123]
- HANNAN, E. J., AND B. G. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, Ser. B*, 41, 190–195. [120]
- HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [138]
- IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [119,120,124]
- KUDO, A. (1963): "A Multivariate Analog of a One-Sided Test," *Biometrika*, 50, 403–418. [127]

- MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions With Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546. [119,124]
- MIKUSHEVA, A. (2007): "Uniform Inference in Autoregressive Models," *Econometrica*, 75, 1411–1452. [125]
- MOON, H. R., AND F. SCHORFHEIDE (2006): "Boosting Your Instruments: Estimation With Overidentifying Inequality Moment Conditions," Unpublished Working Paper, Department of Economics, University of Southern California. [119,123,124]
- OTSU, T. (2006): "Large Deviation Optimal Inference for Set Identified Moment Inequality Models," Unpublished Manuscript, Cowles Foundation, Yale University. [121,123,148]
- PAKES, A., J. PORTER, K. HO, AND J. ISHII (2004): "Applications of Moment Inequalities," Unpublished Working Paper, Department of Economics, Harvard University. [119,123,127]
- POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer. [121]
- PRATT, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 541–567. [121]
- ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Inference and Planning*, 138, 2786–2807. [120,123,127]
- (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211. [120,123,127]
- ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters That Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117. [120,123,127,132,138,139,146]
- SILVAPULLE, M. J., AND P. K. SEN (2005): *Constrained Statistical Inference*. New York: Wiley. [127,139]
- SOARES, G. (2005): "Inference With Inequality Moment Constraints," Unpublished Working Paper, Department of Economics, Yale University. [123,127]
- STOYE, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315. [123]
- WOUTERSEN, T. (2006): "A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters," Unpublished Manuscript, Department of Economics, Johns Hopkins University. [123]

Cowles Foundation for Research in Economics, Dept. of Economics, Yale University, P.O. Box 208281, Yale Station, New Haven, CT 06520-8281, U.S.A.; Donald.andrews@yale.edu

and

Dept. of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268, U.S.A.; gsoares@aya.yale.edu.

Manuscript received October, 2007; final revision received February, 2009.

BINARY RESPONSE MODELS FOR PANEL DATA: IDENTIFICATION AND INFORMATION

BY GARY CHAMBERLAIN¹

This paper considers a panel data model for predicting a binary outcome. The conditional probability of a positive response is obtained by evaluating a given distribution function (F) at a linear combination of the predictor variables. One of the predictor variables is unobserved. It is a random effect that varies across individuals but is constant over time. The semiparametric aspect is that the conditional distribution of the random effect, given the predictor variables, is unrestricted.

This paper has two results. If the support of the observed predictor variables is bounded, then identification is possible only in the logistic case. Even if the support is unbounded, so that (from Manski (1987)) identification holds quite generally, the information bound is zero unless F is logistic. Hence consistent estimation at the standard \sqrt{n} rate is possible only in the logistic case.

KEYWORDS: Panel data, binary response, correlated random effects, identification, information bound.

1. INTRODUCTION

THIS PAPER CONSIDERS a panel data model for predicting a binary outcome. The conditional probability of a positive response is obtained by evaluating a given distribution function (F) at a linear combination of the predictor variables. One of the predictor variables is unobserved. It is a random effect that varies across individuals but is constant over time. The semiparametric aspect is that the conditional distribution of the random effect, given the predictor variables, is unrestricted.

When the distribution function F is logistic, Rasch's (1960, 1961) conditional likelihood approach can be used to obtain a consistent estimator. Andersen (1970) examined the properties of this estimator. See Chamberlain (1984) for a review and additional results.

Manski (1987) showed that consistent estimation is possible without specifying a functional form for F . Furthermore, the form of F can be allowed to depend on the predictor variables in a time-invariant way. Identification does, however, require an unbounded support for at least one of the predictor variables.

This paper has two results. If the support of the observed predictor variables is bounded, then identification is possible only in the logistic case. Even if the support is unbounded, so that (from Manski (1987)) identification holds quite generally, the information bound is zero unless F is logistic. Hence consistent estimation at the standard \sqrt{n} rate is possible only in the logistic case.

¹Financial support was provided by the National Science Foundation.

2. IDENTIFICATION

The random vector $(y_{i1}, y_{i2}, x'_{i1}, x'_{i2}, c_i)$ is independently and identically distributed (i.i.d.) for $i = 1, \dots, n$. We observe $z'_i = (y_{i1}, y_{i2}, x'_{i1}, x'_{i2})$; the latent variable c_i is not observed. The binary variable $y_{it} = 0$ or 1 and $x'_i \equiv (x'_{i1}, x'_{i2})$ has support $X \subset \mathcal{R}^J \times \mathcal{R}^J$. We assume that

$$\text{Prob}(y_{it} = 1|x_i, c_i) = F(\alpha_0 d_{it} + \beta'_0 x_{it} + c_i) \quad (t = 1, 2),$$

where $d_{it} = 1$ if $t = 2$ and = 0 otherwise. The distribution function F is given as part of the prior specification; it is strictly increasing on the whole line with a bounded, continuous derivative, and with $\lim_{s \rightarrow -\infty} F(s) = 0$ and $\lim_{s \rightarrow \infty} F(s) = 1$. Furthermore, y_{i1} and y_{i2} are independent conditional on x_i, c_i . The parameter space $\Theta = \Theta_1 \times \Theta_2$, where Θ_1 is an open subset of \mathcal{R} , Θ_2 is an open subset of \mathcal{R}^J , and $\theta'_0 \equiv (\alpha_0, \beta'_0) \in \Theta$. We assume that Θ_2 contains all $\beta \neq 0$ with $|\beta|$ sufficiently small.

Define

$$p(x, c, \theta) = \begin{pmatrix} [1 - F(\beta' x_1 + c)][1 - F(\alpha + \beta' x_2 + c)] \\ [1 - F(\beta' x_1 + c)]F(\alpha + \beta' x_2 + c) \\ F(\beta' x_1 + c)[1 - F(\alpha + \beta' x_2 + c)] \end{pmatrix}.$$

Let \mathcal{G} consist of the mappings from X into the space of probability measures on \mathcal{R} . We let G_x denote G evaluated at x for $G \in \mathcal{G}$. We shall say that identification fails at θ_0 if

$$\int p(x, c, \theta_0) G_{0x}(dc) = \int p(x, c, \theta^*) G_x^*(dc)$$

for all $x \in X$, where G_0 and $G^* \in \mathcal{G}$, $\theta^* \in \Theta$, and $\theta^* \neq \theta_0$. Then (θ_0, G_0) and (θ^*, G^*) give the same conditional distribution for (y_1, y_2) given x .

The distribution F is logistic if

$$F(s) = \exp(\phi_1 + \phi_2 s) / [1 + \exp(\phi_1 + \phi_2 s)]$$

for some $\phi_1, \phi_2 \in \mathcal{R}$.

THEOREM 1: *If X is bounded, then identification fails for all θ_0 in some open subset of Θ if F is not logistic.*

PROOF: Let $\text{cop}(x, \mathcal{R}, \theta)$ denote the convex hull of the set $\{p(x, c, \theta) : c \in \mathcal{R}\}$. Suppose that for some $\alpha \in \Theta_1$, this convex hull contains an open ball B (in \mathcal{R}^3) when $\beta = 0$. Then for any θ_0 and $\theta^* \in \Theta$ sufficiently close to $(\alpha, 0)$,

$$\text{cop}(x, \mathcal{R}, \theta_0) \cap \text{cop}(x, \mathcal{R}, \theta^*)$$

is nonempty for all $x \in X$. Then for each $x \in X$, there are probability measures G_{0x} and G_x^* such that

$$\int p(x, c, \theta_0) G_{0x}(dc) = \int p(x, c, \theta^*) G_x^*(dc).$$

Hence θ_0 is not identified unless the dimension of $\text{cop}(x, \mathcal{R}, (\alpha, 0))$ is 2 for all $\alpha \in \Theta_1$. In that case, for each $\alpha \in \Theta_1$, there exist scalars ψ_1, \dots, ψ_4 (not all zero) such that

$$\begin{aligned} \psi_1[1 - F(c)][1 - F(\alpha + c)] + \psi_2[1 - F(c)]F(\alpha + c) \\ + \psi_3F(c)[1 - F(\alpha + c)] = \psi_4 \end{aligned}$$

for all $c \in \mathcal{R}$. Letting $c \rightarrow \infty$ gives $\psi_4 = 0$; letting $c \rightarrow -\infty$ gives $\psi_1 = 0$. Hence, with $Q \equiv F/(1 - F)$, we have

$$\psi_2 Q(\alpha + c) + \psi_3 Q(c) = 0$$

and so

$$Q(\alpha + c) = Q(\alpha)Q(c)/Q(0)$$

for all $\alpha \in \Theta_1$ and all $c \in \mathcal{R}$. The only positive, continuously differentiable solution to this form of Cauchy's equation is

$$Q(s) = \exp(\phi_1 + \phi_2 s);$$

then the result follows from $F = Q/(1 + Q)$. *Q.E.D.*

Manski's (1987) model is specified as (dropping the i subscripts)

$$y_t = \begin{cases} 1, & \text{if } \theta'_0 w_t + c + u_t \geq 0, \\ 0, & \text{otherwise} \end{cases} \quad (t = 1, 2),$$

$$u_1|w_1, w_2, c \stackrel{d}{=} u_2|w_1, w_2, c.$$

The key restriction here is that the latent error, u_t , should have an identical distribution in both periods, conditional on the predictor variables w_1, w_2, c . (Also the support of the distribution of u_t conditional on w_1, w_2, c is assumed to be \mathcal{R} .) This allows for a certain kind of heteroskedasticity, but does not permit, for example, the conditional distribution of u_t to be more sensitive to w_t than to w_s ($s, t = 1, 2; s \neq t$).

Our model, with $w'_t = (d_t, x'_t)$, imposes additional restrictions. We require u_1, u_2 to be independent of w_1, w_2, c and to be i.i.d. over time with a known distribution ($-u_t \stackrel{d}{=} F$).

Manski's result is that θ_0 is identified up to scale if a component of $w_2 - w_1$ (with a nonzero coefficient) has positive Lebesgue density on the whole line, conditional on the other components of $w_2 - w_1$ and on c . (In addition, the support of the $K \times 1$ vector $w_2 - w_1$ should not be contained in a proper linear subspace of \mathcal{R}^K .) Our scale normalization is built in to the given specification for the u_i distribution, but we can make comparisons by considering ratios of coefficients. The proof of Theorem 1 shows that a logistic F is necessary for such ratios to be identified. It is the bounded support for the predictor variables (w_1, w_2) that accounts for the difference in results.

In the next section, we shall assume that x_i has positive Lebesgue density on all of \mathcal{R}^{2j} . Then identification is possible in general, but we shall see that the information bound is zero except for the logistic case.

3. INFORMATION

Our semiparametric information bound is based on considering the least favorable parametric subfamily. This idea is owing to Stein (1956) and has been developed by Levit (1975), Begun, Hall, Huang, and Wellner (1983), and Pfanzagl (1982).

The observations consist of independent and identically distributed (i.i.d.) random vectors z_1, \dots, z_n with values in Z , a subset of a Euclidean space. The distribution of z_1 has positive density $f(z; \theta_0, g_0)$ with respect to a σ -finite measure μ . The parametric component θ_0 is an element of Θ , an open subset of \mathcal{R}^K . The nonparametric component g_0 is an element of Γ , an infinite-dimensional set. A path λ through g_0 is a mapping from an open interval $(c, d) \subset \mathcal{R}$ into Γ , where $\lambda(\delta_0) = g_0$ for a unique $\delta_0 \in (c, d)$. The path λ is used to construct a parametric likelihood function

$$f_\lambda(z; \theta, \delta) = f(z; \theta, \lambda(\delta)).$$

Let $\gamma' = (\theta', \delta)$ and $\gamma'_0 = (\theta'_0, \delta_0)$. Then we have mean-square differentiability at γ_0 if

$$f_\lambda^{1/2}(z; \gamma) - f_\lambda^{1/2}(z; \gamma_0) = \sum_{j=1}^{K+1} \psi_{\lambda j}(z)(\gamma_j - \gamma_{0j}) + r(z; \gamma),$$

where

$$\int r^2(z; \gamma) \mu(dz) / |\gamma - \gamma_0|^2 \rightarrow 0$$

as $\gamma \rightarrow \gamma_0$. If the mean-square differential exists and if the partial derivatives exist almost everywhere with respect to μ (a.e. μ), then

$$\psi_{\lambda j}(z) = \frac{1}{2} f_\lambda^{-1/2}(z; \gamma_0) \partial f_\lambda(z; \gamma_0) / \partial \gamma_j \quad (\text{a.e. } \mu).$$

The partial information for θ_{0k} ($k = 1, \dots, K$) in the path λ is

$$I_{\lambda,k} = 4 \inf_{\alpha \in \mathcal{R}^{K+1}, \alpha_k=1} \int \left(\sum_{j=1}^{K+1} \alpha_j \psi_{\lambda j}(z) \right)^2 \mu(dz).$$

Given a set Λ of paths, we define

$$I_{\Lambda,k} = \inf_{\lambda \in \Lambda} I_{\lambda,k}.$$

We let $I_{\Lambda} = 0$ denote that $I_{\Lambda,k} = 0$ for $k = 1, \dots, K$. In that case, no component of θ_0 can be estimated at a \sqrt{n} rate; see Chamberlain (1986, Theorem 2).

In the panel data model from Section 2, we shall assume that the conditional distribution of c given x has a density g with respect to Lebesgue measure. Then the likelihood is based on the density (with respect to μ)

$$f(z; \theta, g) = \int A(z, c, \theta) g(c, x) dc,$$

where

$$A(z, c, \theta) = \prod_{t=1}^2 F(\alpha d_t + \beta' x_t + c)^{y_t} \cdot [1 - F(\alpha d_t + \beta' x_t + c)]^{(1-y_t)}.$$

The measure μ is defined on $Z = Y \times X$, where $Y = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ and X is now $\mathcal{R}^J \times \mathcal{R}^J$. If $B_1 \subset Y$ and B_2 is a Borel subset of X , then

$$\mu(B_1 \times B_2) = \tau(B_1)v(B_2),$$

where τ is the counting measure on Y and the measure v gives the probability distribution of $x_i \equiv (x_{i1}, x_{i2})$.

We shall assume that v has positive Lebesgue density on all of \mathcal{R}^{2J} . In addition, we shall assume that $\beta_{0J} \neq 0$ for $\beta_0 \in \Theta_2$.

We shall use the following specification for Γ :

DEFINITION 1: Γ consists of all measurable functions $g: \mathcal{R} \times X \rightarrow \mathcal{R}$ such that (i) $\inf_{(c,x) \in B} g(c, x) > 0$ for any compact subset B of $\mathcal{R} \times X$; (ii) for all $x \in X$, $\int_{-\infty}^{\infty} g(c, x) dc = 1$; (iii) for each $z \in Z$ and $\theta \in \Theta$, there exists a neighborhood $C \subset \Theta$ of θ and a measurable function $q_1: \mathcal{R} \rightarrow \mathcal{R}$ such that $\int q_1(c)g(c, x) dc < \infty$ and

$$|\partial A(z, c, \eta)/\partial \theta| \leq q_1(c) \quad \text{for all } \eta \in C;$$

(iv) for each $\theta \in \Theta$, there is a neighborhood $C \subset \Theta$ of θ and a function $q_2 : Z \rightarrow \mathcal{R}$ such that $\int q_2(z)\mu(dz) < \infty$ and

$$\begin{aligned} & \left[\int |\partial A(z, c, \eta)/\partial \theta| g(c, x) dc \right]^2 \\ & \quad \int \int A(z, c, \eta) g(c, x) dc \leq q_2(z) \quad \text{for all } \eta \in C. \end{aligned}$$

Part (iii) of Definition 1 ensures that f is continuously differentiable with respect to θ for each $z \in Z$; part (iv) ensures that the mean-square derivative of $f^{1/2}$ with respect to θ exists.

We shall work with the following set of paths:

DEFINITION 2: Λ consists of the paths

$$\lambda(\delta) = g_0[1 + (\delta - \delta_0)h],$$

where $g_0 \in \Gamma$ and $h : \mathcal{R} \times X \rightarrow \mathcal{R}$ is a bounded, measurable function with

$$\int g_0(c, x)h(c, x) dc = 0 \quad \text{for all } x \in X.$$

THEOREM 2: $I_\Lambda = 0$ for all θ_0 in Θ if F is not logistic.

PROOF: It is straightforward to check that $\lambda(\delta) \in \Gamma$ for δ sufficiently close to δ_0 . Define the parametric likelihood function $f_\lambda(z; \theta, \delta) = f(z; \theta, \lambda(\delta))$. Now let $\gamma' = (\theta', \delta)$ and $\gamma'_0 = (\theta'_0, \delta_0)$, and apply the mean-value theorem to obtain

$$f_\lambda^{1/2}(z; \gamma) - f_\lambda^{1/2}(z; \gamma_0) = \frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \gamma'} (\gamma - \gamma_0) + r(z; \gamma),$$

where

$$r(z; \gamma) = \left[\frac{\partial f_\lambda^{1/2}(z; \tilde{\gamma})}{\partial \gamma} - \frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \gamma} \right]' (\gamma - \gamma_0)$$

and $\tilde{\gamma}$ is on the line segment joining γ and γ_0 :

$$\frac{r^2(z; \gamma)}{|\gamma - \gamma_0|^2} \leq \left| \frac{\partial f_\lambda^{1/2}(z; \tilde{\gamma})}{\partial \gamma} - \frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \gamma} \right|^2 \rightarrow 0$$

as $\gamma \rightarrow \gamma_0$ (a.e. μ) since the partial derivatives are continuous in γ . Then the dominated convergence theorem implies that

$$\int r^2(z; \gamma)\mu(dz)/|\gamma - \gamma_0|^2 \rightarrow 0$$

and so the mean-square differentiability condition is satisfied.

Now we need to show that if F is not logistic, then given $k \in \{1, \dots, K\}$ and given $\varepsilon > 0$, there is a path $\lambda \in \Lambda$ such that

$$(1) \quad \begin{aligned} 4 \int \left(\frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \theta_k} - \frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \delta} \right)^2 \mu(dz) \\ = \sum_y \int f^{-1}(z; \theta_0, g_0) \\ \times \left(\frac{\partial f(z; \theta_0, g_0)}{\partial \theta_k} - \int A(z, c, \theta_0) g_0(c, x) h(c, x) dc \right)^2 v(dx) \\ < \varepsilon. \end{aligned}$$

Since $\int f^{-1}(z; \theta_0, g_0) [\partial f(z; \theta_0, g_0)/\partial \theta_k]^2 v(dx) < \infty$, we can choose an $\varepsilon' > 0$ such that (1) is satisfied if there is a compact subset B of \mathcal{R}^{2J} with $v(B) > 1 - \varepsilon'$, $h(c, x) = 0$ for $x \notin B$, and

$$(2) \quad \begin{aligned} \sum_y \int_B f^{-1}(z; \theta_0, g_0) \\ \times \left(\frac{\partial f(z; \theta_0, g_0)}{\partial \theta_k} - \int A(z, c, \theta_0) g_0(c, x) h(c, x) dc \right)^2 v(dx) < \varepsilon'. \end{aligned}$$

Since $f(z; \theta_0, g_0)$ is bounded away from 0 for $x \in B$, there is an $\varepsilon'' > 0$ such that (2) is satisfied if there is a bounded, measurable function $m: \mathcal{R} \times B \rightarrow \mathcal{R}$ such that for all $x \in B: m(c, x) = 0$ for $|c|$ sufficiently large, $\int m(c, x) dc = 0$ and

$$(3) \quad \left[\sum_y \left(\frac{\partial f(z; \theta_0, g_0)}{\partial \theta_k} - \int A(z, c, \theta_0) m(c, x) dc \right)^2 \right]^{1/2} < \varepsilon''.$$

Then we set

$$h(c, x) = 1(x \in B)m(c, x)/g_0(c, x).$$

($1(\cdot)$ is the indicator function that equals 1 if the condition is satisfied and equals 0 otherwise.)

Let $r(x)$ denote the 4×1 vector whose elements are $\partial f(z; \theta_0, g_0)/\partial \theta_k$ for $y = (0, 0), (0, 1), (1, 0), (1, 1)$. Note that $l'r(x) = 0$, where l is a 4×1 vector of 1's. Define

$$a(x, c, \theta) = \begin{pmatrix} [1 - F(\beta'x_1 + c)][1 - F(\alpha + \beta'x_2 + c)] \\ [1 - F(\beta'x_1 + c)]F(\alpha + \beta'x_2 + c) \\ F(\beta'x_1 + c)[1 - F(\alpha + \beta'x_2 + c)] \\ F(\beta'x_1 + c)F(\alpha + \beta'x_2 + c) \end{pmatrix}.$$

Then (3) can be written as

$$(3') \quad \left| r(x) - \int a(x, c, \theta_0) m(c, x) dc \right| < \varepsilon''.$$

Suppose that for all $x \in \mathcal{R}^{2J}$ except for a set with v -probability zero, there exist points $c_j(x) \in \mathcal{R}$ ($j = 1, \dots, 4$) with

$$[a(x, c_1(x), \theta_0), \dots, a(x, c_4(x), \theta_0)]$$

nonsingular. Then for each such x , there is a neighborhood C_x of x such that

$$[a(x', c_1(x), \theta_0), \dots, a(x', c_4(x), \theta_0)]$$

is nonsingular for all x' in the closure of C_x . The C_x provide an open cover of a compact set B with $v(B) > 1 - \varepsilon'$. Hence there is a finite subcover, and we can partition B into Borel subsets D_1, \dots, D_m and choose the c_j to be simple (hence measurable) functions of the form $c_j(x) = \sum_{k=1}^m \kappa_{jk} 1(x \in D_k)$. Furthermore, we can choose the c_j such that

$$H(x) = [a(x, c_1(x), \theta_0), \dots, a(x, c_4(x), \theta_0)]$$

has its determinant bounded away from zero for $x \in B$.

Define $b(x) = H^{-1}(x)r(x)$. Since $l'H(x) = l'$, we have

$$l'b(x) = l'H(x)b(x) = l'r(x) = 0.$$

Then (3') can be written as

$$(3'') \quad \left| \sum_{j=1}^4 a(x, c_j(x), \theta_0) b_j(x) - \int a(x, c, \theta_0) m(c, x) dc \right| < \varepsilon''.$$

Set

$$m(c, x) = \sum_{j=1}^4 1(|c - c_j(x)| < \delta) b_j(x) / (2\delta).$$

Then m is bounded and measurable, $m(c, x) = 0$ for $|c|$ sufficiently large,

$$\int m(c, x) dc = \sum_{j=1}^4 b_j(x) = 0,$$

and (3'') is satisfied if $\delta > 0$ is sufficiently small.

We conclude that $I_A = 0$ unless, for all x in a set S with positive Lebesgue (outer) measure, $\{a(x, c, \theta_0) : c \in \mathcal{R}\}$ lies in a proper linear subspace of \mathcal{R}^4 .

Then for each such x , there exists a nonzero $\psi \in \mathcal{R}^4$ such that $\psi' a(x, c, \theta_0) = 0$ for all $c \in \mathcal{R}$; that is,

$$\begin{aligned} & \psi_1[1 - F(\beta'_0 x_1 + c)][1 - F(\alpha_0 + \beta'_0 x_2 + c)] \\ & + \psi_2[1 - F(\beta'_0 x_1 + c)]F(\alpha_0 + \beta'_0 x_2 + c) \\ & + \psi_3F(\beta'_0 x_1 + c)[1 - F(\alpha_0 + \beta'_0 x_2 + c)] \\ & + \psi_4F(\beta'_0 x_1 + c)F(\alpha_0 + \beta'_0 x_2 + c) = 0. \end{aligned}$$

Taking the limit as $c \rightarrow \infty$ gives $\psi_4 = 0$, and letting $c \rightarrow -\infty$ gives $\psi_1 = 0$. Hence, with $Q \equiv F/(1 - F)$, we have

$$\psi_2 Q(\alpha_0 + \beta'_0 x_2 + c) + \psi_3 Q(\beta'_0 x_1 + c) = 0$$

and so

$$(4) \quad Q(\alpha_0 + \beta'_0 x_2 + c) = Q(\alpha_0 + \beta'_0 x_2)Q(\beta'_0 x_1 + c)/Q(\beta'_0 x_1).$$

Then (4) holds for all x in the closure of S . Define $M(s) = \log Q(s)$ and $\dot{M}(s) = dM(s)/ds$, and take the partial derivative with respect to the J th component of x_2 :

$$\dot{M}(\alpha_0 + \beta'_0 x_2 + c)\beta_{0J} = \dot{M}(\alpha_0 + \beta'_0 x_2)\beta_{0J}$$

for all $c \in \mathcal{R}$. Hence $M(s) = \phi_1 + \phi_2 s$, and the result follows from $F = \exp(M)/(1 + \exp(M))$. *Q.E.D.*

If $0 \notin \Theta_1$, then we can reparameterize in terms of $\tilde{\theta} = (\alpha, \beta/\alpha)$. Then we can apply the proof of Theorem 2 to show that the information bound for $\tilde{\beta}_0 \equiv \beta_0/\alpha_0$ is 0 unless F is logistic. Hence, even though a consistent estimator of $\tilde{\beta}_0$ is available from Manski (1987) (under the additional assumption that the parameter space bounds $|\beta_{0J}|/(|\alpha_0| + |\beta_0|)$ away from 0), estimation at the standard \sqrt{n} rate is possible only in the logistic case.

REFERENCES

- ANDERSEN, E. B. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society, Ser. B*, 32, 283–301. [159]
- BEGUN, J., W. HALL, W. HUANG, AND J. WELLNER (1983): "Information and Asymptotic Efficiency in Parametric–Nonparametric Models," *The Annals of Statistics*, 11, 432–452. [162]
- CHAMBERLAIN, G. (1984): "Panel Data," in *Handbook of Econometrics*, Vol. II, ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland, Chapter 22. [159]
- (1986): "Asymptotic Efficiency in Semiparametric Models With Censoring," *Journal of Econometrics*, 32, 189–218. [163]
- LEVIT, B. (1975): "On the Efficiency of a Class of Nonparametric Estimates," *Theory of Probability and Its Applications*, 20, 723–740. [162]

- MANSKI, C. (1987): "Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data," *Econometrica*, 55, 357–362. [159,161,167]
- PFANZAGL, J. (1982): *Contributions to a General Asymptotic Statistical Theory*. Lecture Notes in Statistics, Vol. 13. New York: Springer-Verlag. [162]
- RASCH, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Denmarks Paedagogiske Institute. [159]
- (1961): "On General Laws and the Meaning of Measurement in Psychology," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4. Berkeley: University of California Press. [159]
- STEIN, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley: University of California Press. [162]

Dept. of Economics, Harvard University, 123 Littauer Center, 1875 Cambridge Street, Cambridge, MA 02138, U.S.A.; gary_chamberlain@harvard.edu.

Manuscript received February, 2009; final revision received August, 2009.

INFERENCE FOR THE IDENTIFIED SET IN PARTIALLY IDENTIFIED ECONOMETRIC MODELS

BY JOSEPH P. ROMANO AND AZEEM M. SHAIKH¹

This paper provides computationally intensive, yet feasible methods for inference in a very general class of partially identified econometric models. Let P denote the distribution of the observed data. The class of models we consider is defined by a population objective function $Q(\theta, P)$ for $\theta \in \Theta$. The point of departure from the classical extremum estimation framework is that it is not assumed that $Q(\theta, P)$ has a unique minimizer in the parameter space Θ . The goal may be either to draw inferences about some unknown point in the set of minimizers of the population objective function or to draw inferences about the set of minimizers itself. In this paper, the object of interest is $\Theta_0(P) = \arg \min_{\theta \in \Theta} Q(\theta, P)$, and so we seek random sets that contain this set with at least some prespecified probability asymptotically. We also consider situations where the object of interest is the image of $\Theta_0(P)$ under a known function. Random sets that satisfy the desired coverage property are constructed under weak assumptions. Conditions are provided under which the confidence regions are asymptotically valid not only pointwise in P , but also uniformly in P . We illustrate the use of our methods with an empirical study of the impact of top-coding outcomes on inferences about the parameters of a linear regression. Finally, a modest simulation study sheds some light on the finite-sample behavior of our procedure.

KEYWORDS: Partially identified model, incomplete model, identified set, identifiable parameter, subsampling, uniform coverage, confidence region, moment inequalities.

1. INTRODUCTION

A PARTIALLY IDENTIFIED MODEL is any model in which the parameter of interest is not uniquely defined by the distribution of the observed data. This paper provides computationally intensive yet feasible methods for inference for one large class of such models. Let P denote the distribution of the observed data. The class of models we consider is defined by a population objective function $Q(\theta, P)$ for $\theta \in \Theta$. The point of departure from the classical extremum estimation framework is that it is not assumed that $Q(\theta, P)$ has a unique minimizer in the parameter space Θ . The goal may be either to draw inferences about some unknown point in the set of minimizers of the population objective function or to draw inferences about the set of minimizers itself. In this paper we consider the second of these two goals. The object of interest is

$$(1) \quad \Theta_0(P) = \arg \min_{\theta \in \Theta} Q(\theta, P).$$

We henceforth refer to $\Theta_0(P)$ as the *identified set*. In this instance, given independent and identically distributed (i.i.d.) data $X_i, i = 1, \dots, n$, generated

¹We would like to thank Michael Wolf for a careful reading of the paper and useful suggestions. We also thank Nese Yildiz for pointing out the need for the nonzero variance condition in Example 2.1.

from P , we seek random sets $\mathcal{C}_n = \mathcal{C}_n(X_1, \dots, X_n)$ that contain the identified set with at least some prespecified probability asymptotically. That is, we require

$$(2) \quad \liminf_{n \rightarrow \infty} P\{\Theta_0(P) \subseteq \mathcal{C}_n\} \geq 1 - \alpha.$$

We refer to such sets as *confidence regions for the identified set that are pointwise consistent in level*. This terminology reflects the fact that the confidence regions are valid only for a *fixed* probability distribution P and helps distinguish this coverage requirement from others discussed later in which we will demand that the confidence regions are valid uniformly in P . We show that the problem of constructing \mathcal{C}_n that satisfy (2) is equivalent to a multiple hypothesis testing problem in which one wants to test the family of null hypotheses $H_\theta : \theta \in \Theta_0(P)$ indexed by $\theta \in \Theta$ while controlling the familywise error rate, the probability of even one false rejection under P . Using this duality, we go on to construct \mathcal{C}_n that satisfy (2) under weak assumptions on P .

In the first goal, the object of interest is some unknown point $\theta \in \Theta_0(P)$. We refer to any $\theta \in \Theta_0(P)$ as an *identifiable parameter*. In this case, given i.i.d. data X_i , $i = 1, \dots, n$, generated from P , we seek random sets $\mathcal{C}_n = \mathcal{C}_n(X_1, \dots, X_n)$ that contain each identifiable parameter with at least some pre-specified probability asymptotically. The problem of constructing such sets is treated in a companion paper (Romano and Shaikh (2008)).

Our results on confidence regions for the identified set build upon the earlier work of Chernozhukov, Hong, and Tamer (2007), who were the first to consider inference for the same class of partially identified models. An important feature of our procedure for constructing confidence regions for the identified set is that it avoids the need for an initial estimate of $\Theta_0(P)$. In general, our procedure is first-order asymptotically equivalent with the procedure proposed by Chernozhukov, Hong, and Tamer (2007). On the other hand, when the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$, our results provide a justification for iterating their procedure until a stopping criterion is met to produce confidence regions that are typically strictly smaller while still maintaining the coverage requirement.

In this paper, we also wish to construct confidence regions whose coverage probability is close to the nominal level not just for a fixed probability distribution P , but rather uniformly over all P in some large class of distributions \mathbf{P} . Confidence regions that fail to satisfy this requirement have the feature that for every sample size n , however large, there is some probability distribution $P \in \mathbf{P}$ for which the coverage probability of the confidence region under P is not close to the prescribed level. Researchers may, therefore, feel that inferences made on the basis of asymptotic approximations are more reliable if the confidence regions exhibit good uniform behavior. Of course, such a requirement will typically require restrictions on P beyond those required for pointwise consistency in level. Bahadur and Savage (1956), for example, showed that if \mathbf{P} is suitably large, then there exists no confidence interval for the mean with finite length

and good uniform behavior. Romano (2004) extended this nonexistence result to a number of other problems. We provide restrictions on \mathbf{P} under which the confidence regions in this paper have good uniform behavior. Concretely, we provide conditions under which \mathcal{C}_n satisfies

$$(3) \quad \liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} P\{\Theta_0(P) \subseteq \mathcal{C}_n\} \geq 1 - \alpha.$$

By analogy with our earlier terminology, sets that satisfy (3) are referred to as *confidence regions for the identified set that are uniformly consistent in level*. Note that if the identified set $\Theta_0(P)$ consists of a single point $\theta_0(P)$, then this definition reduces to the usual definition of confidence regions that are uniformly consistent in level; that is,

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} P\{\theta_0(P) \in \mathcal{C}_n\} \geq 1 - \alpha.$$

Imbens and Manski (2004) analyzed the special case of the above class of partially identified models in which the identified set is an interval whose upper and lower endpoints are means or at least behave like means asymptotically. For this special case, they constructed confidence regions that contain each identifiable parameter with at least some prespecified probability asymptotically and are valid uniformly in P . Romano and Shaikh (2008) constructed confidence regions with this same coverage property for the more general class of models considered here. To the best of our knowledge, this paper is the first to consider confidence regions for the identified set that are valid uniformly in P .

We have so far assumed that the object of interest is the identified set, $\Theta_0(P)$, itself. More generally, the object of interest may be the image of the identified set under a known function. A typical example of such a function is the projection of \mathbf{R}^k onto one of the axes. We extend the above definitions of confidence regions to this setting as follows. Consider a function $f: \Theta \rightarrow \Lambda$. Denote by $\Lambda_0(P)$ the image of $\Theta_0(P)$ under f ; that is,

$$(4) \quad \Lambda_0(P) = \{f(\theta) : \theta \in \Theta_0(P)\}.$$

We refer to a set \mathcal{C}_n^f as a *confidence region for a function of the identified set that is pointwise consistent in level* if it satisfies

$$(5) \quad \liminf_{n \rightarrow \infty} P\{\Lambda_0(P) \in \mathcal{C}_n^f\} \geq 1 - \alpha.$$

As before, we may also demand uniformly good behavior over a class of probability distributions \mathbf{P} by requiring that \mathcal{C}_n^f satisfy

$$(6) \quad \liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} P\{\Lambda_0(P) \in \mathcal{C}_n^f\} \geq 1 - \alpha.$$

By analogy with our earlier terminology, sets that satisfy (6) are referred to as *confidence regions for a function of the identified set that are uniformly consistent*

in level. We adapt our constructions of confidence regions for the identified set to provide constructions of confidence sets that satisfy these alternative coverage requirements.

The remainder of the paper is organized as follows. In Section 2, we consider the problem of constructing confidence regions that satisfy the coverage properties (2) and (3). The construction exploits a useful equivalence between the construction of confidence regions for the identified set and a suitable multiple hypothesis testing problem. We then extend this methodology to construct confidence regions that satisfy (5) and (6). We provide an illustration of our methods in Section 3. In Section 4, we shed some light on the finite-sample behavior of our methodology via a small simulation study. Data and programs are provided as Supplemental Material ([Romano and Shaikh \(2010\)](#)).

2. CONFIDENCE REGIONS FOR THE IDENTIFIED SET

In this section, we consider the problem of constructing confidence regions for the identified set. We begin by treating the construction of sets that satisfy (2) before turning our attention to the problem of constructing sets that satisfy (3).

2.1. Pointwise Consistency in Level

2.1.1. Equivalence With a Multiple Testing Problem

We will first show that the problem of constructing confidence regions that satisfy (2) is equivalent to a certain multiple hypothesis testing problem. The problem is to test the family of hypotheses

$$(7) \quad H_\theta : Q(\theta, P) = 0 \quad \text{for } \theta \in \Theta$$

in a way that asymptotically controls the familywise error rate (FWER_P), the probability of one or more false rejections under P , at level α . Formally,

$$(8) \quad \text{FWER}_P = P\{\text{reject at least 1 null hypothesis } H_\theta \text{ s.t. } Q(\theta, P) = 0\},$$

and by asymptotic control of the FWER_P at level α , we mean the requirement that

$$(9) \quad \limsup_{n \rightarrow \infty} \text{FWER}_P \leq \alpha.$$

The following lemma establishes the equivalence between these two problems.

LEMMA 2.1: *Let P denote the true distribution of the data. Given any procedure for testing the family of null hypotheses (7) which yields a decision for each of the*

null hypotheses, the set of θ values for which the corresponding null hypothesis H_θ is accepted, \mathcal{C}_n , satisfies

$$P\{\Theta_0(P) \subseteq \mathcal{C}_n\} = 1 - \text{FWER}_P,$$

where $\Theta_0(P)$ is defined by (1). Conversely, given any random set \mathcal{C}_n , the procedure for testing the family of hypotheses (7) in which a null hypothesis H_θ is accepted if and only if $\theta \in \mathcal{C}_n$ satisfies

$$\text{FWER}_P = 1 - P\{\Theta_0(P) \subseteq \mathcal{C}_n\}.$$

PROOF: To establish the first conclusion, note that by the definition of $\Theta_0(P)$ we have

$$\begin{aligned} P\{\Theta_0(P) \subseteq \mathcal{C}_n\} &= P\{\text{reject no null hypothesis } H_\theta \text{ s.t. } Q(\theta, P) = 0\} \\ &= 1 - \text{FWER}_P. \end{aligned}$$

The second conclusion follows from the same reasoning. *Q.E.D.*

It follows from Lemma 2.1 that given any procedure for testing the family of null hypotheses (7) that satisfy (9), the set of θ values corresponding to the set of accepted hypotheses, \mathcal{C}_n , satisfies (2). We thus turn to the problem of constructing tests of (7) that satisfy (9).

2.1.2. Single-Step Control of the Familywise Error Rate

First, we briefly discuss a single-step approach to asymptotic control of the FWER_P at level α , since it serves as a building block for the more powerful stepdown procedures that we will develop in the next section. As before, we will require a test statistic for each null hypothesis H_θ such that large values of the test statistic provide evidence against the null hypothesis. The statistic $a_n \hat{Q}_n(\theta)$ for some sequence $a_n \rightarrow \infty$ will be used for this purpose. We assume that the sequence a_n is known and that it does not depend on θ , but both of these requirements can be relaxed using ideas in Chapter 8 of Politis, Romano, and Wolf (1999).

For $K \subseteq \Theta$, let $c_n(K, 1 - \alpha, P)$ denote the smallest $1 - \alpha$ quantile of the distribution of

$$(10) \quad \sup_{\theta \in K} a_n \hat{Q}_n(\theta)$$

under P ; that is,

$$c_n(K, 1 - \alpha, P) = \inf \left\{ x : P \left\{ \sup_{\theta \in K} a_n \hat{Q}_n(\theta) \leq x \right\} \geq 1 - \alpha \right\}.$$

Consider the idealized test in which a null hypothesis H_θ is rejected if and only if $a_n \hat{Q}_n(\theta) > c_n(\Theta_0(P), 1 - \alpha, P)$. This is a single-step method in the sense that each $a_n \hat{Q}_n(\theta)$ is compared with a common value so as to determine its significance. Clearly, such a test satisfies $\text{FWER}_P \leq \alpha$. To see this, note that

$$\begin{aligned} \text{FWER}_P &= P\{a_n \hat{Q}_n(\theta) > c_n(\Theta_0(P), 1 - \alpha, P) \text{ for some } \theta \in \Theta_0(P)\} \\ &= 1 - P\left\{\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \leq c_n(\Theta_0(P), 1 - \alpha, P)\right\} \leq \alpha. \end{aligned}$$

But this test is infeasible, as the critical value depends on the unknown P . A crude solution to this difficulty is available if estimators $\hat{c}_n(K, 1 - \alpha)$ of $c_n(K, 1 - \alpha, P)$ are available that satisfy two properties. First, we require that $\hat{c}_n(K, 1 - \alpha)$ be a “good” estimator when $K = \Theta_0(P)$ in the sense that

$$(11) \quad \limsup_{n \rightarrow \infty} P\left\{\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) > \hat{c}_n(\Theta_0(P), 1 - \alpha)\right\} \leq \alpha.$$

Second, we require that the estimators be monotone in the sense that

$$(12) \quad \hat{c}_n(K, 1 - \alpha) \geq \hat{c}_n(\Theta_0(P), 1 - \alpha) \quad \text{for any } K \supseteq \Theta_0(P).$$

Since $\Theta_0(P) \subseteq \Theta$, it follows that under these assumptions we have that $\hat{c}_n(\Theta, 1 - \alpha)$ asymptotically provides a conservative estimator of $c_n(\Theta_0(P), 1 - \alpha, P)$. Hence, the single-step method in which each statistic $a_n \hat{Q}_n(\theta)$ is compared with the common cutoff $\hat{c}_n(\Theta, 1 - \alpha)$ asymptotically controls the FWER_P at level α provided that these two assumptions are satisfied. We will refrain from stating this result more formally because it will follow from the analysis of the more powerful step-down method in the next section. We will also provide an explicit construction of such estimators in the subsequent section.

2.1.3. Step-Down Control of the Familywise Error Rate

Step-down methods begin by first applying a single-step method, but then additional hypotheses may be rejected after this first stage by proceeding in a stepwise fashion, which we now describe. In the first stage, test the entire family of hypotheses using a single-step procedure; that is, reject all null hypotheses whose corresponding test statistic is too large, where large is determined by some common critical value as described above. If no hypotheses are rejected in this first stage, then stop; otherwise, test the family of hypotheses not rejected in the first stage using a single-step procedure. If no further hypotheses are rejected in this second stage, then stop; otherwise, test the family of hypotheses not rejected in the first and second stages using a single-step procedure. Repeat this process until no further hypotheses are rejected. We now formally define this procedure, which can be viewed as a generalization

of Romano and Wolf (2005), who only considered a finite number of hypotheses.

ALGORITHM 2.1:

1. Let $S_1 = \Theta$. If $\sup_{\theta \in S_1} a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_1, 1 - \alpha)$, then accept all hypotheses and stop; otherwise, set $S_2 = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_1, 1 - \alpha)\}$ and continue.
- ⋮
- j. If $\sup_{\theta \in S_j} a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_j, 1 - \alpha)$, then accept all hypotheses H_θ with $\theta \in S_j$ and stop; otherwise, set $S_{j+1} = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \leq \hat{c}_n(S_j, 1 - \alpha)\}$ and continue.
- ⋮

We now prove that this algorithm provides asymptotic control of the FWER_P under the monotonicity assumption (11) and (12).

THEOREM 2.1: *Let P denote the true distribution generating the data. Consider Algorithm 2.1 with critical values that satisfy (12). Then*

$$(13) \quad \text{FWER}_P \leq P \left\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) > \hat{c}_n(\Theta_0(P), 1 - \alpha) \right\}.$$

Hence, if the critical values also satisfy (11), then

$$\limsup_{n \rightarrow \infty} \text{FWER}_P \leq \alpha.$$

PROOF: To establish (13), denote by \hat{j} the smallest random index for which there is a false rejection; that is, there exists $\theta' \in \Theta_0(P)$ such that $a_n \hat{Q}_n(\theta') > \hat{c}_n(S_{\hat{j}}, 1 - \alpha)$. By definition of \hat{j} , we must have that $\Theta_0(P) \subseteq S_{\hat{j}}$. Thus, by (12) we have that $\hat{c}_n(S_{\hat{j}}, 1 - \alpha) \geq \hat{c}_n(\Theta_0(P), 1 - \alpha)$. Hence, it must be the case that

$$\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \geq a_n \hat{Q}_n(\theta') > \hat{c}_n(\Theta_0(P), 1 - \alpha).$$

The second conclusion follows immediately. *Q.E.D.*

2.1.4. A Subsampling Construction

It follows from Theorem 2.1 that under the two restrictions (11) and (12), the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1, C_n , satisfies (2). We now provide a concrete construction of critical values that satisfy these two properties under a weak assumption on the asymptotic behavior of the test statistics $a_n \hat{Q}_n(\theta)$.

The construction will be based on subsampling. To define the critical values precisely, some further notation is required. Let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \rightarrow 0$. Let $N_n = \binom{n}{b}$ and let $\hat{Q}_{n,b,i}(\theta)$ denote the statistic $\hat{Q}_n(\theta)$ evaluated at the i th subset of data of size b from the n observations. For $K \subseteq \Theta$ and $\alpha \in (0, 1)$, define

$$(14) \quad \hat{r}_n(K, 1 - \alpha) = \inf \left\{ x : \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I \left\{ \sup_{\theta \in K} a_b \hat{Q}_{n,b,i}(\theta) \leq x \right\} \geq 1 - \alpha \right\}.$$

Note that by construction, the critical values defined by (14) satisfy the monotonicity restriction (12). We now provide conditions under which they also satisfy (11).

THEOREM 2.2: *Let $X_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P and let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \rightarrow 0$. Let $J_n(\cdot, P)$ denote the distribution of $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ under P . Suppose $J_n(\cdot, P)$ converges in distribution to a limit distribution $J(\cdot, P)$ and that $J(\cdot, P)$ is continuous at its smallest $1 - \alpha$ quantile. Then the following statements are true:*

- (i) *Condition (11) holds when $\hat{c}_n(\Theta_0(P), 1 - \alpha)$ is given by (14) with $K = \Theta_0(P)$.*
- (ii) *Algorithm 2.1 with $\hat{c}_n(K, 1 - \alpha)$ given by (14) provides asymptotic control of the FWER_P at level α .*
- (iii) *The set of θ values corresponding to accepted hypotheses from Algorithm 2.1 with $\hat{c}_n(K, 1 - \alpha)$ given by (14), \mathcal{C}_n , satisfies (2).*

PROOF: The first result follows from Theorem 2.1.1 of Politis, Romano, and Wolf (1999). The second follows from Theorem 2.1. The third follows from Lemma 2.1. *Q.E.D.*

REMARK 2.1: Because $\binom{n}{b}$ may be large, it is often more practical to use the following approximation to (14). Let the sequence $B_n \rightarrow \infty$ as $n \rightarrow \infty$ and let I_1, \dots, I_{B_n} be chosen randomly with or without replacement from the numbers $1, \dots, N_n$. Then it follows from Corollary 2.4.1 of Politis, Romano, and Wolf (1999) that one may approximate (14) by

$$\inf \left\{ x : \frac{1}{B_n} \sum_{1 \leq i \leq B_n} I \left\{ \sup_{\theta \in K} a_b \hat{Q}_{n,b,I_i}(\theta) \leq x \right\} \geq 1 - \alpha \right\}$$

without affecting the conclusions of Theorem 2.2.

REMARK 2.2: Throughout this paper, we assume that the observations $X_i, i = 1, \dots, n$, are an i.i.d. sequence of random variables with distribution P .

Many of the results, however, can be extended to certain time series settings. Consider, for example, the following extension of Theorem 2.2. Let

$$(15) \quad \hat{r}_n(K, 1 - \alpha) = \inf \left\{ x : \frac{1}{n-b+1} \sum_{1 \leq i \leq n-b+1} I \left\{ \sup_{\theta \in K} a_b \hat{Q}_{n,b,i}(\theta) \leq x \right\} \geq 1 - \alpha \right\},$$

where $i = 1, \dots, n-b+1$ now indexes only the subsets of data of size b whose observations are consecutive. If one assumes that the $X_i, i = 1, \dots, n$, are observations from a distribution P for which the corresponding α -mixing sequence $\alpha_X(m) \rightarrow 0$ as $m \rightarrow \infty$, but otherwise maintains the assumptions of Theorem 2.2, then it follows from Theorem 3.2.1 of Politis, Romano, and Wolf (1999) that the conclusions of the theorem continue to hold.

We now consider two important examples and use Theorem 2.2 to provide conditions under which Algorithm 2.1 with $\hat{c}_n(K, 1 - \alpha)$ given by (14) asymptotically controls the FWER _{P} and thus the set of θ values corresponding to the accepted hypotheses, \mathcal{C}_n , satisfies (2).

EXAMPLE 2.1—Moment Inequalities: Let $X_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P on \mathbf{R}^k . For $j = 1, \dots, m$, let $g_j(x, \theta)$ be a real-valued function on $\mathbf{R}^k \times \mathbf{R}^l$. The identified set is assumed to be $\Theta_0(P) = \{\theta \in \mathbf{R}^l : E_P[g_j(X_i, \theta)] \leq 0 \ \forall j \text{ s.t. } 1 \leq j \leq m\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = \sum_{1 \leq j \leq m} (E_P[g_j(X_i, \theta)])_+^2,$$

where $(x)_+ = \max\{x, 0\}$. The sample analog of $Q(\theta, P)$ is given by

$$\hat{Q}_n(\theta) = \sum_{1 \leq j \leq m} \left(\frac{1}{n} \sum_{1 \leq i \leq n} g_j(X_i, \theta) \right)_+^2.$$

Let $a_n = n$ and suppose P is such that (i)

$$\{g_j(\cdot, \theta) : 1 \leq j \leq m, \theta \in \Theta_0(P)\} \text{ is } P\text{-Donsker},$$

(ii) for every $K \subseteq \{1, \dots, m\}$ and sequence $\theta_n \in \Theta_0(P)$ such that $E_P[g_j(X_i, \theta_n)] \rightarrow 0$ for all $j \in K$, there exists a subsequence n_k and a $\theta \in \Theta_0(P)$ such that $E_P[g_j(X_i, \theta)] = 0$ for all $j \in K$ and $\rho_{P,j}(\theta_{n_k}, \theta) \rightarrow 0$ for all $j \in K$, where $\rho_{P,j}^2(\theta, \theta') = E_P[(g_j(X_i, \theta) - g_j(X_i, \theta'))^2] \rightarrow 0$. To rule out degenerate situations, assume further that (iii) there exist $1 \leq j^* \leq m$ and $\theta^* \in \Theta_0(P)$ such that $E_P[g_{j^*}(X_i, \theta^*)] = 0$ and $\text{Var}_P[g_{j^*}(X_i, \theta^*)] > 0$. Assumption (i) is known to hold provided that the class of functions is not too large; for general results to this

end and numerous applications, see van der Vaart and Wellner (1996). The conditions of Theorem 2.2 are verified for $\alpha < \frac{1}{2}$ under these assumptions in Section A.2 of the Appendix.

EXAMPLE 2.2—Regression With Interval Outcomes: The following example allows for inference in a linear regression model in which the dependent variable is interval-censored. Let $(X_i, Y_{1,i}, Y_{2,i}, Y_i^*), i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P^* on $\mathbf{R}^k \times \mathbf{R} \times \mathbf{R} \times \mathbf{R}$. The parameter of interest, θ_0 , is known to satisfy $E_{P^*}[Y_i^*|X_i] = X_i'\theta_0$, but Y_i^* is unobserved, which precludes conventional estimation of θ_0 . Let P denote the distribution of the observed random variables $(X_i, Y_{1,i}, Y_{2,i})$. The random variables $(Y_{1,i}, Y_{2,i})$ are known to satisfy $Y_{1,i} \leq Y_i^* \leq Y_{2,i}$ with probability 1 under P^* . Thus, $\theta_0 \in \Theta_0(P) = \{\theta \in \mathbf{R}^k : E_P[Y_{1,i}|X_i] \leq X_i'\theta \leq E_P[Y_{2,i}|X_i] \text{ } P\text{-a.s.}\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = E_P[(E_P[Y_{1,i}|X_i] - X_i'\theta)_+^2 + (X_i'\theta - E_P[Y_{2,i}|X_i])_+^2].$$

Manski and Tamer (2002) characterized the identified set in this setting and also considered the case where Y_i^* is observed, but X_i is interval-censored.

Let $a_n = n$ and suppose P is such that (i) $\text{supp}_P(X_i) = \{x_1, \dots, x_m\}$ and (ii) the variances of Y_1 and Y_2 , $\sigma_1^2(P)$ and $\sigma_2^2(P)$, exist. To rule out degenerate situations, assume further that (iii) there exist $\theta^* \in \Theta$, $\ell^* \in \{1, 2\}$, and $j^* \in \{1, \dots, m\}$ such that $E_P[Y_{\ell^*,i}|X_i = x_{j^*}] = x_{j^*}'\theta^*$ and $\text{Var}_P[Y_{\ell^*,i}|X_i = x_{j^*}] > 0$. For $\ell \in \{1, 2\}$ and $j \in \{1, \dots, m\}$, let $\tau_\ell(x_j, P) = E_P[Y_{\ell,i}|X_i = x_j]$ and

$$\hat{\tau}_\ell(x_j) = \frac{1}{n(x_j)} \sum_{1 \leq i \leq n: X_i = x_j} Y_{\ell,i},$$

where $n(x_j) = |\{1 \leq i \leq n : X_i = x_j\}|$. Let

$$\hat{Q}_n(\theta) = \sum_{1 \leq j \leq m} \frac{n(x_j)}{n} \{(\hat{\tau}_1(x_j) - x_j'\theta)_+^2 + (x_j'\theta - \hat{\tau}_2(x_j))_+^2\}.$$

We now verify the conditions of Theorem 2.2 under these assumptions for $\alpha < \frac{1}{2}$.

To this end, note that

$$\begin{aligned} a_n \hat{Q}_n(\theta) &= \sum_{1 \leq j \leq m} \sum_{1 \leq \ell \leq 2} \left(\sqrt{\frac{n}{n(x_j)}} \frac{1}{\sqrt{n}} \right. \\ &\quad \times \left. \sum_{1 \leq i \leq n} (-1)^{\ell-1} (Y_{\ell,i} - x_j'\theta) I\{X_i = x_j\} \right)_+^2. \end{aligned}$$

For $K \subseteq \{1, \dots, m\} \times \{1, 2\}$ let

$$\Theta_0(K, P) = \{\theta \in \Theta_0(P) : E_P[Y_{\ell,i}|X_i = x_j] = x'_j \theta \text{ for all } (j, \ell) \in K\}.$$

Hence, except for the multiplicative factors $\sqrt{n/n(x_j)}$, which are asymptotically constant anyway, the structure here is the same as the structure of Example 2.1. As a result, we may use arguments nearly identical to those given for Example 2.1 above in Section A.2 of the Appendix to show that the limiting behavior of

$$(16) \quad \sup_{\theta \in \Theta_0(K, P)} a_n \hat{Q}_n(\theta)$$

is equal to the limiting behavior of

$$\begin{aligned} & \max_K \sup_{\theta \in \Theta_0(K, P)} \sum_{(j, \ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} \frac{1}{\sqrt{n}} \right. \\ & \quad \times \left. \sum_{1 \leq i \leq n} (-1)^{\ell-1} (Y_{\ell,i} - x'_j \theta) I\{X_i = x_j\} \right)_+^2 \\ &= \max_K \sup_{\theta \in \Theta_0(K, P)} \sum_{(j, \ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} \frac{1}{\sqrt{n}} \right. \\ & \quad \times \left. \sum_{1 \leq i \leq n} (-1)^{\ell-1} (Y_{\ell,i} - E_P[Y_{\ell,i}|X_i = x_j]) I\{X_i = x_j\} \right)_+^2, \end{aligned}$$

where $p(x_j) = P\{X_i = x_j\}$, the maximum over K is understood to be over all subsets of $\{1, \dots, m\} \times \{1, 2\}$, and the supremum over the empty set is understood to be zero. The vector whose (j, ℓ) component is given by

$$\frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (-1)^{\ell-1} (Y_{\ell,i} - E_P[Y_{\ell,i}|X_i = x_j]) I\{X_i = x_j\}$$

tends in distribution to a multivariate normal random variable. Let $Z_{j,\ell}(P)$ denote the (j, ℓ) component of this limiting multivariate normal random variable. It follows by the continuous mapping theorem that (16) tends in distribution to

$$(17) \quad \max_K \sup_{\theta \in \Theta_0(K, P)} \sum_{(j, \ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} Z_{j,\ell}(P) \right)_+^2.$$

To determine for which α (17) is continuous at its $1 - \alpha$ quantile, first note that (17) is a convex function of $Z(P)$. By Theorem 11.1 of Davydov, Lifshits, and

Smorodina (1998), the distribution of (17) is continuous everywhere except possibly at zero, but

$$\begin{aligned} P\left\{\max_K \sup_{\theta \in \Theta_0(K, P)} \sum_{(j, \ell) \in K} \left(\frac{1}{\sqrt{p(x_j)}} Z_{j, \ell}(P)\right)_+^2 \leq 0\right\} \\ \leq P\left\{\left(\frac{1}{\sqrt{p(x_{j^*})}} Z_{j^*, \ell^*}(P)\right)_+^2 \leq 0\right\} \leq \frac{1}{2}, \end{aligned}$$

where j^* , ℓ^* , and θ^* are as in assumption (iii) above. Hence, (17) is continuous at its $1 - \alpha$ quantile for $\alpha < \frac{1}{2}$.

REMARK 2.3: A Tobit-like model is a special case of the above setup if we suppose further that $Y_{2,i} = Y_i^*$ and $Y_{1,i} = Y_i^*$ if $Y_i^* > 0$, and $Y_{2,i} = 0$ and $Y_{1,i} = -\infty$ (or some large negative number if there is a plausible lower bound on Y_i^*) if $Y_i^* \leq 0$.

REMARK 2.4: Our construction of critical values has used subsampling. Following Andrews (2000), it is possible to show that a naive bootstrap construction fails to approximate the distribution of (10) when $K = \Theta_0(P)$. It may still be the case that (11) is satisfied, but in simulations it seems to be too conservative in practice. Bugni (2007) showed that a suitably modified version of the bootstrap can be used to estimate the distribution of (10) when $K = \Theta_0(P)$. His approximation depends crucially on the structure of Example 2.1 and does not extend easily to more general models, but it is worthwhile to note that it can be used as an ingredient in Algorithm 2.1. Specifically, we may replace $\hat{c}_n(K, 1 - \alpha)$ with the $1 - \alpha$ quantile of his bootstrap approximation to the distribution of (10). It follows from the analysis of Bugni (2007) that these critical values will satisfy (11) under weak assumptions. Since they also satisfy (12), the conclusions of Theorem 2.1 follow.

REMARK 2.5: When the set of minimizers of $\hat{Q}_n(\theta)$ provides a consistent estimate of $\Theta_0(P)$, Chernozhukov, Hong, and Tamer (2007) proposed constructing confidence regions that satisfy (2) using a single-step method in which

$$(18) \quad S_1 = \tilde{\Theta}_{0,n} = \arg \min_{\theta \in \Theta} \hat{Q}_n(\theta)$$

and critical values are given by (14); that is,

$$\mathcal{C}_n = \{\theta \in \Theta : a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\tilde{\Theta}_{0,n}, 1 - \alpha)\}.$$

Such an approach can be shown by example to fail to lead to confidence regions that satisfy (2) when the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$. See Example 2.7 of Romano and Shaikh (2006) for details.

REMARK 2.6: When the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$, Chernozhukov, Hong, and Tamer (2007) proposed constructing confidence regions that satisfy (2) using a single-step method in which

$$(19) \quad S_1 = \hat{\Theta}_{0,n} = \{\theta \in \Theta : \hat{Q}_n(\theta) < \varepsilon_n\},$$

where ε_n is a positive sequence of constants tending to zero slowly. Because of this restriction on the rate at which ε_n tends to zero, they were able to show that

$$(20) \quad P\{\Theta_0(P) \subseteq S_1\} \rightarrow 1.$$

The proof of Theorem 2.1 requires that the initial set S_1 be such that $\Theta_0(P) \subseteq S_1$, but one can allow for S_1 to be random provided that it satisfies (20) without affecting the argument in any way. Hence, using our results, it follows that this construction satisfies (2). Unfortunately, the specific choice of ε_n in finite samples is arbitrary and the confidence region resulting from application of their method may thus be very large or very small depending on the choice of ε_n . Our results provide a justification of iterating their procedure until a stopping criterion is met, thereby removing this arbitrariness, and produce typically smaller confidence regions while still maintaining the coverage requirement.

REMARK 2.7: It follows from the discussion in Remark 2.5 that there are no first-order differences between the confidence regions from our step-down procedure with S_1 given by (19) and those of Chernozhukov, Hong, and Tamer (2007). Even with such a delicate choice of S_1 , we expect the iterative approach to perform better in finite samples. To this end, it is worthwhile to examine second-order differences. In Section A.3 of the Appendix, we show in the context of a simple example that our confidence region is smaller to second order than the one proposed by Chernozhukov, Hong, and Tamer (2007). In the example we consider, it is important to note that the set of minimizers of $\hat{Q}_n(\theta)$ provides a consistent estimate of $\Theta_0(P)$, so one could instead use a single-step procedure with S_1 given by (18). Compared with this procedure, our confidence region is not smaller to second order. We simply use the example to illustrate a phenomenon that we expect to persist even when the set of minimizers of $\hat{Q}_n(\theta)$ does not provide a consistent estimate of $\Theta_0(P)$.

2.2. Uniform Consistency in Level

We now provide conditions under which the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1, \mathcal{C}_n , satisfies (3).

THEOREM 2.3: Let $X_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P and let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \rightarrow 0$. Let $J_n(\cdot, P)$ denote the distribution of $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ under P . Suppose $P \in \mathbf{P}$ and

$$(21) \quad \limsup_{n \rightarrow \infty} \sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} \{J_b(x, P) - J_n(x, P)\} \leq 0.$$

Then the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1 using critical values given by (14), \mathcal{C}_n , satisfies (3).

PROOF: By Theorem 2.1, we have that

$$\text{FWER}_P \leq 1 - P \left\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\Theta_0(P), 1 - \alpha) \right\}.$$

By Theorem 3.1(iv) of Romano and Shaikh (2008), it follows that

$$(22) \quad \liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} P \left\{ \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\Theta_0(P), 1 - \alpha) \right\} \geq 1 - \alpha.$$

Thus,

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathbf{P}} \text{FWER}_P \leq \alpha.$$

The asserted claim now follows immediately from Lemma 2.1. *Q.E.D.*

The intuition behind condition (21) is as follows. Lemma A.1 shows under weak conditions that the subsampling estimator of the distribution of $\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta)$ approximates $J_b(x, P)$ well uniformly over both $x \in \mathbf{R}$ and $P \in \mathbf{P}$. Condition (21) implies that critical values from $J_b(x, P)$ are no smaller than those from $J_n(x, P)$. Hence, under this assumption, Theorem 3.1 of Romano and Shaikh (2008) implies that subsampling behaves well over $P \in \mathbf{P}$ in the sense that (22) holds.

We now apply Theorem 2.3 to construct confidence regions that satisfy the coverage requirement (3) for the two examples considered in Section 2.

EXAMPLE 2.3—Moment Inequalities: Recall the setup of Example 2.1. We will now use Theorem 2.3 to show that for this example the set of θ values corresponding to the accepted hypotheses from Algorithm 2.1, \mathcal{C}_n , satisfies (3) for a large class of distributions \mathbf{P} . To this end, let $a_n = n$, let \mathbf{P} be such that (i)

$$\{g_j(\cdot, \theta) : 1 \leq j \leq m, \theta \in \Theta\} \quad \text{is } P\text{-Donsker and pre-Gaussian}$$

uniformly in $P \in \mathbf{P}$ and (ii) Θ is compact with respect to the metric

$$\bar{\rho}(\theta, \theta') = \sup_{P \in \mathbf{P}} \max_{1 \leq j \leq m} \rho_{P,j}(\theta, \theta'),$$

where $\rho_{P,j}^2(\theta, \theta') = E_P[(g_j(X, \theta) - g_j(X, \theta'))^2]$. To rule out degenerate situations, assume further that (iii) there exists $\varepsilon > 0$ such that for each $P \in \mathbf{P}$ there exist $1 \leq j^* \leq m$ and $\theta^* \in \Theta_0(P)$ such that $E_P[g_{j^*}(X_i, \theta^*)] = 0$ and $\text{Var}_P[g_{j^*}(X_i, \theta^*)] \geq \varepsilon$. Assumption (i) is again known to hold provided that the class of functions is not too large; for general results to this end and numerous applications, see [van der Vaart and Wellner \(1996\)](#). In the [Appendix](#), we verify that the required condition (21) holds under these assumptions.

EXAMPLE 2.4—Regression With Interval Outcomes: Recall the setup of [Example 2.2](#). As argued there, the structure of this example is similar to that of [Example 2.3](#). Since we provide details in the case of [Example 2.3](#) above, we do not do so here.

2.3. Confidence Regions for Functions of the Identified Set

In this section, we consider the problem of constructing sets that satisfy (5) and (6). Let $f : \Theta \rightarrow \Lambda$ be given. Our construction again relies on equivalence with an appropriate multiple testing problem, but in this case the family of null hypotheses is given by

$$(23) \quad H_\lambda : \lambda \in \Lambda_0(P) \quad \text{for } \lambda \in \Lambda,$$

where $\Lambda_0(P)$ is defined by (4). The alternative hypotheses are understood to be

$$K_\lambda : \lambda \notin \Lambda_0(P) \quad \text{for } \lambda \in \Lambda.$$

As before, it suffices to consider the problem of testing this family of null hypotheses in a way that controls the FWER_P at level α .

For $\lambda \in \Lambda$, let $f^{-1}(\lambda) = \{\theta \in \Theta : f(\theta) = \lambda\}$. Note that

$$\begin{aligned} \lambda \in \Lambda_0(P) &\iff \exists \theta \in f^{-1}(\lambda) \quad \text{s.t. } Q(\theta, P) = 0 \\ &\implies \inf_{\theta \in f^{-1}(\lambda)} Q(\theta, P) = 0. \end{aligned}$$

This suggests a natural test statistic for each of these null hypotheses H_λ :

$$(24) \quad \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta),$$

where $a_n \hat{Q}_n(\theta)$ is the test statistic used earlier to test the null hypothesis that $Q(\theta, P) = 0$.

We may now proceed as before, but with this test statistic in place of our earlier test statistic $a_n \hat{Q}_n(\theta)$. For $K \subseteq \Lambda$, let $\hat{c}_n^f(K, 1 - \alpha)$ be an estimator of the $1 - \alpha$ quantile of distribution of

$$\sup_{\lambda \in K} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta)$$

and consider the following modification of Algorithm 2.1.

ALGORITHM 2.2:

1. Let $S_1 = \Lambda$. If $\sup_{\lambda \in S_1} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_1, 1 - \alpha)$, then accept all H_λ and stop; otherwise, set $S_2 = \{\lambda \in \Lambda : \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_1, 1 - \alpha)\}$ and continue.
- ⋮
- j. If $\sup_{\lambda \in S_j} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_j, 1 - \alpha)$, then accept all H_λ with $\lambda \in S_j$ and stop; otherwise, set $S_{j+1} = \{\lambda \in \Lambda : \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta) \leq \hat{c}_n^f(S_j, 1 - \alpha)\}$ and continue.
- ⋮

We now provide conditions under which the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 leads to confidence regions that satisfy (5) and (6). For $K \subseteq \Lambda$ and $\alpha \in (0, 1)$, let

$$(25) \quad \begin{aligned} \hat{r}_n^f(K, 1 - \alpha) \\ = \inf \left\{ x : \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I \left\{ \sup_{\lambda \in K} \inf_{\theta \in f^{-1}(\lambda)} a_b \hat{Q}_{n,b,i}(\theta) \leq x \right\} \geq 1 - \alpha \right\}. \end{aligned}$$

THEOREM 2.4: *Let $X_i, i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P and let $b = b_n < n$ be a sequence of positive integers tending to infinity, but satisfying $b/n \rightarrow 0$. Let $J_n(\cdot, P)$ denote the distribution of $\sup_{\lambda \in \Lambda_0(P)} \inf_{\theta \in f^{-1}(\lambda)} a_n \hat{Q}_n(\theta)$ under P . Let \mathcal{C}_n^f denote the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 when $\hat{c}_n^f(K, 1 - \alpha)$ is given by (25).*

- (i) *Suppose $J_n(\cdot, P)$ converges in distribution to $J(\cdot, P)$ and that $J(\cdot, P)$ is continuous at its smallest $1 - \alpha$ quantile. Then \mathcal{C}_n^f satisfies (5).*
- (ii) *Suppose $P \in \mathbf{P}$ and*

$$(26) \quad \limsup_{n \rightarrow \infty} \sup_{P \in \mathbf{P}} \sup_{x \in \mathbf{R}} \{J_b(x, P) - J_n(x, P)\} \leq 0.$$

Then \mathcal{C}_n^f satisfies (6).

The proof follows immediately from the arguments given in Sections 2.1 and 2.2.

We now provide a simple illustration of the use of Theorem 2.4.

EXAMPLE 2.5: Let (X_i, Y_i) , $i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P on \mathbf{R}^2 . The parameter of interest, θ_0 , is known to satisfy $\theta_{0,1} \geq \mu_X(P)$ and $\theta_{0,2} \geq \mu_Y(P)$. The identified set is therefore given by $\Theta_0(P) = \{\theta \in \mathbf{R}^2 : \theta_1 \geq \mu_X(P) \text{ and } \theta_2 \geq \mu_Y(P)\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = (\mu_X(P) - \theta_1)_+^2 + (\mu_Y(P) - \theta_2)_+^2.$$

The sample analog of $Q(\theta, P)$ is given by $\hat{Q}_n(\theta) = (\bar{X}_n - \theta_1)_+^2 + (\bar{Y}_n - \theta_2)_+^2$. Suppose the object of interest is the projection of $\Theta_0(P)$ onto its first component rather than the entire set $\Theta_0(P)$; that is, the object of interest is $\Lambda_0(P) = f(\Theta_0(P))$, where $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ is defined by $f(\theta) = \theta_1$ instead of $\Theta_0(P)$. Note that $\Lambda_0(P)$ is simply $\{\theta_1 \in \mathbf{R} : \theta_1 \geq \mu_X(P)\}$.

First consider the problem of constructing sets that satisfy (5). Let $a_n = n$ and suppose P is such that $\sigma_X^2(P)$ exists. Assume without loss of generality that $\mu_X(P) = 0$. Then

$$\begin{aligned} \sup_{\theta_1 \in \Lambda_0(P)} \inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta) &= \sup_{\theta_1 \geq 0} \inf_{\theta_2 \in \mathbf{R}} n(\bar{X}_n - \theta_1)_+^2 + n(\bar{Y}_n - \theta_2)_+^2 \\ &= n(\bar{X}_n)_+^2 \xrightarrow{\mathcal{L}} (\sigma_X(P)Z)_+^2, \end{aligned}$$

where Z is a standard normal random variable. It now follows from Theorem 2.4(i) that the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 when $\hat{c}_n^f(K, 1 - \alpha)$ is given by (25), \mathcal{C}_n^f , satisfies (5).

Now consider the problem of constructing sets that satisfy (6). As before, let $a_n = n$ and let \mathbf{P} be a set of distributions for which the marginal distribution of X satisfies

$$(27) \quad \lim_{\lambda \rightarrow \infty} \sup_{P \in \mathbf{P}} E_P \left[\frac{|X - \mu(P)|^2}{\sigma^2(P)} I \left\{ \frac{|X - \mu(P)|}{\sigma(P)} > \lambda \right\} \right] = 0.$$

After noting that $\sup_{\theta_1 \in \Lambda_0(P)} \inf_{f^{-1}(\theta_1)} a_n \hat{Q}_n(\theta)$ is simply $n(\bar{X}_n)_+^2$, it is straightforward to apply Lemma 11.4.1 of Lehmann and Romano (2005) to show that (26) holds. Therefore, it follows from Theorem 2.4(ii) that the set of θ values corresponding to accepted hypotheses from Algorithm 2.2 when $\hat{c}_n^f(K, 1 - \alpha)$ is given by (25), \mathcal{C}_n^f , satisfies (6).

REMARK 2.8: Given a confidence region for the identified set \mathcal{C}_n , one construction of a confidence region for a function of the identified set is the image

of \mathcal{C}_n under the function of interest. Such a construction will typically be conservative in the sense that the coverage probability will exceed the nominal level.

3. EMPIRICAL ILLUSTRATION

In this section, we use the techniques developed above to examine the impact of top-coding outcomes on the inferences that can be made about the parameters of a linear regression. By top-coding a random variable, we mean the practice of recording the realization of the random variable if and only if it is below a certain threshold. This model is a special case of our Example 2.2, and so the theory developed above applies here under the appropriate assumptions. A similar empirical example can be found in [Chernozhukov, Hong, and Tamer \(2004\)](#).

The motivation for our exercise stems from the following observation. To study changes in the wage structure and earnings inequality, researchers often regress the logarithm of hourly wages on various demographic characteristics. Data sets used for this purpose invariably top-code wages for reasons of confidentiality. One approach to deal with the top-coding of wages is to replace all of the top-coded outcomes with a common value. In practice, this common value is often taken to be a scalar multiple of the threshold. This approach is justified theoretically under the assumption that the distribution of wages conditional on top-coding is distributed as a Pareto random variable. See, for example, [Katz and Autor \(1999\)](#), wherein the scalar used for this purpose is taken to be 1.5. Of course, we do not wish to impose any parametric assumptions.

To examine this issue, we begin with a sample of observations from the Annual Demographic Supplement of the Current Population Survey for the year 2000. For each individual in the survey, the survey records a variety of demographic variables as well as information on wages and salaries. We select observations with the following demographic characteristics: (1) race is white; (2) age is between 20 and 24 years; (3) at least college graduates; (4) primary source of income is wages and salaries; (5) worked at least 2 hours per week on average. There are 305 such observations, none of which suffers from top-coding of wages and salaries. We treat this sample of individuals as the distribution of the observed data P and draw an i.i.d. sample of $n = 1000$ observations from this P . We will analyze these data both for the benchmark case of no top-coding and for cases in which some amount of top-coding has been artificially imposed on the data.

Recall the setup of Example 2.2. To allow for graphical illustration of the confidence regions, we consider only a model in which $k = 2$; specifically, we take $X_i = (1, D_i)$, where D_i is 1 if the sex is female and 0 otherwise. The latent outcome variable $Y_i^* = \log(\text{wage}_i^*/H_i)$, where wage_i^* is total wages and salaries, which is possibly unobserved in the presence of top-coding, and H_i is total

hours worked. We assume that wage_i^* is bounded above by $\overline{\text{wage}} = \10^8 . In the benchmark case in which there is no top-coding, we will let $Y_{1,i} = Y_{2,i} = Y_i^*$. In the cases in which there is some top-coding, let $\underline{\text{wage}}$ be the threshold above which wages are not observed. Define $Y_{1,i} = Y_{2,i} = Y_i^*$ if $\text{wage}_i^* \leq \underline{\text{wage}}$; otherwise, let $Y_{1,i} = \underline{Y}_i = \log(\text{wage}/H_i)$ and $Y_{2,i} = \overline{Y}_i = \log(\overline{\text{wage}}/H_i)$.

Below we will construct confidence regions of level $1 - \alpha = 0.95$ for the identified set for each of three different scenarios. For the sake of completeness, we will also construct confidence regions for identifiable parameters, as described in [Romano and Shaikh \(2008\)](#). More specifically, following [Romano and Shaikh \(2008\)](#), we consider

$$\{\theta \in \mathbf{R}^2 : a_n \hat{Q}_n(\theta) \leq \hat{r}_n(\{\theta\}, 1 - \alpha)\},$$

where $\hat{r}_n(\{\theta\}, 1 - \alpha)$ is given by (14). We will compare the inferences that can be drawn from these confidence regions with those that can be drawn from regressing Y_i^a on X_i , where, in the benchmark case of no top-coding, $Y_i^a = Y_i^*$, and, in cases with top-coding, $Y_i^a = Y_i^*$ if $\text{wage}_i^* \leq \underline{\text{wage}}$ and $Y_i^a = 1.5 \times \underline{Y}_i$ otherwise.

Before proceeding, we discuss some computational details. First, consider the choice of b . In practice, one would like to use a data-dependent subsample size; see [Politis, Romano, and Wolf \(1999\)](#) for a review of several algorithms for choosing the subsample size in this way. For the purposes of this exercise, however, we use the same subsample size, $b = 30$, in each of the constructions. As a result, differences among the confidence regions below are not driven by variation in the choice of subsample size. Note that the results below remain similar for subsample sizes between 20 and 40, the range of subsample sizes for which the simulation results in the following section suggest that the procedure behaves well in finite samples. Second, when computing critical values, we also used an approximation as described in Remark 2.1 with $B = 200$ because $\binom{n}{b}$ is too large to compute critical values exactly. Finally, following the discussion in Remark 2.5, in the first step of Algorithm 2.1, we let

$$S_1 = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 1000\}.$$

The results below remain similar for much larger choices of S_1 .

We first consider the case in which there is no top-coding. Algorithm 2.1 converged after 11 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_{11} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0055\}.$$

We also regress Y_i^a on X_i and obtain a Wald-style confidence region of the form

$$\{\theta \in \mathbf{R}^2 : (\hat{\theta}_n - \theta)' \hat{\Sigma}_n^{-1} (\hat{\theta}_n - \theta) \leq 5.99\},$$

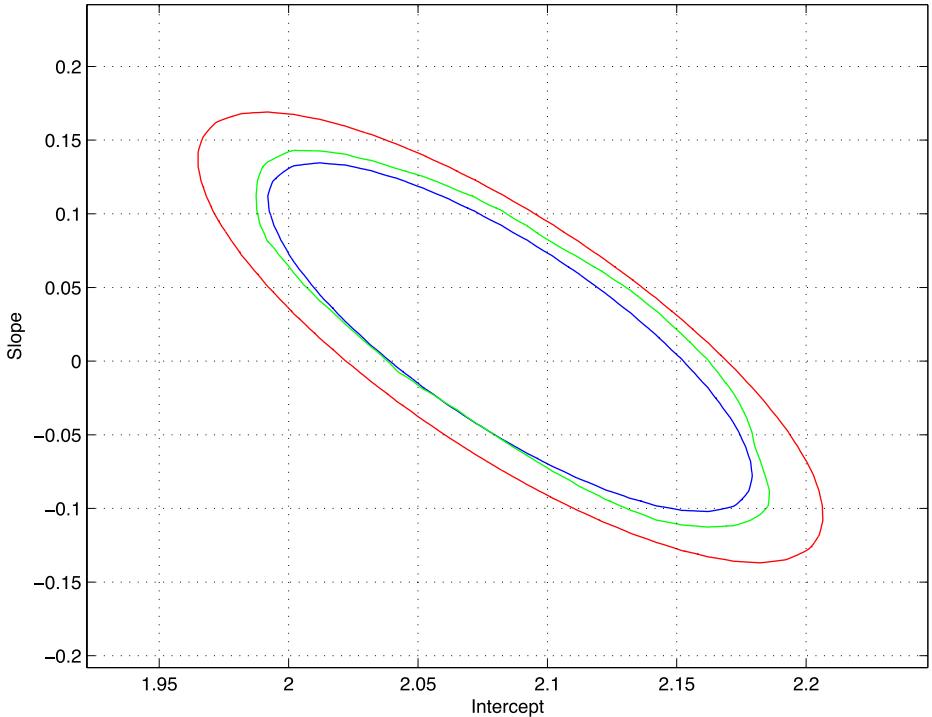


FIGURE 1.—Confidence regions with no top-coding and $\overline{\text{wage}} = \10^8 : green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region.

where $\hat{\Sigma}_n$ is the usual heteroskedasticity-robust estimator of the variance of $\hat{\theta}_n$. These two confidence regions together with the confidence region for identifiable parameters are displayed in Figure 1. Since the true P that generates the data is known, it is also possible to calculate the identified set, which in this case is a singleton. It is given by $\Theta_0(P) = \{(2.047, 0.042)\}$. As one would expect, in this instance all three confidence regions are of similar shape and size. The largest is the confidence region for the identified set and the smallest is the Wald-style confidence region. The confidence region for identifiable parameters is contained strictly within the confidence region for the identified set.

Next, we consider a case in which there is some amount of top-coding and repeat the exercise above. For concreteness, we choose $\overline{\text{wage}} = \$41,000$, which corresponds to 5% of the population being subject to top-coding. Algorithm 2.1 converged after 12 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_{12} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0182\}.$$

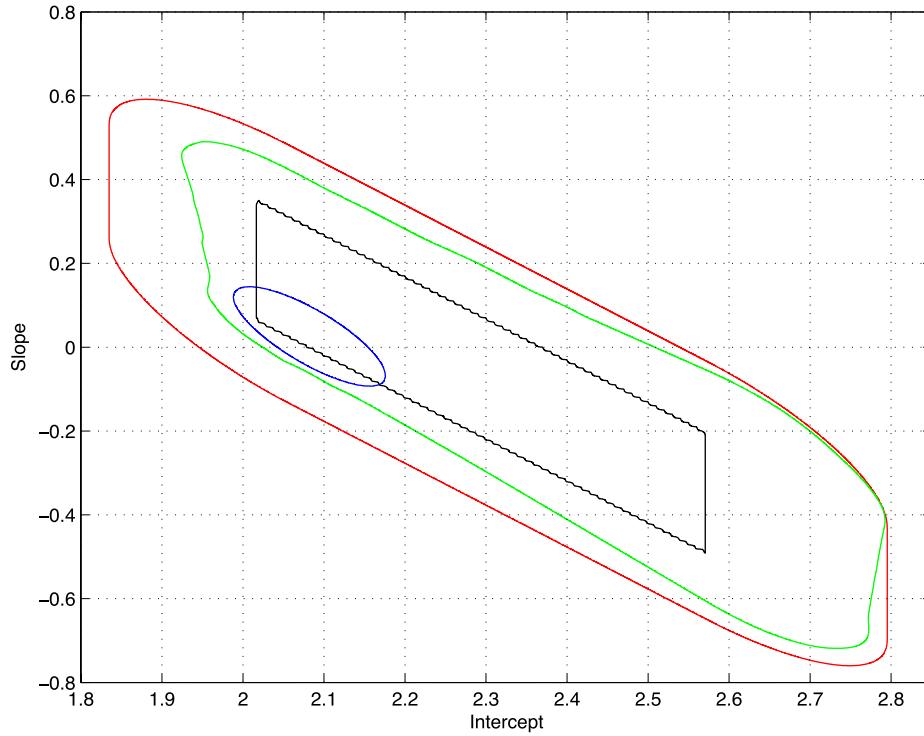


FIGURE 2.—Confidence regions with 5% top-coding and $\bar{\text{wage}} = \$10^8$: green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region, black = identified set.

The resulting confidence regions are displayed in Figure 2. Again, we may also calculate the identified set, which is no longer a singleton due to top-coding. It is given by

$$\begin{aligned} \{\theta \in \mathbf{R}^2 : E\{Y_{1,i}|D_i = 0\} &\leq \theta_1 \leq E\{Y_{2,i}|D_i = 0\}, \\ E\{Y_{1,i}|D_i = 1\} &\leq \theta_1 + \theta_2 \leq E\{Y_{2,i}|D_i = 1\} \end{aligned}$$

and is therefore a parallelogram. This set is also displayed in Figure 2. Both the confidence region for the identified set and the confidence region for identifiable parameters contain the identified set, but, as before, the confidence region for identifiable parameters is contained strictly within the confidence region for the identified set. The Wald-style confidence region, though still the smallest, covers only a small portion of the identified set. As a result, inferences based on the Wald-style confidence region might be very misleading if the assumptions used to achieve identification are not correct.

To make this point more forcefully, we carry out the same exercise for the case in which there is even more top-coding. Specifically, we reduce wage to \$35,000, which corresponds to 10% of the population being subject to top-coding. Algorithm 2.1 converged after 9 steps and the confidence region for the identified set is given by

$$\mathcal{C}_n = S_9 = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0361\}.$$

The resulting confidence regions along with the identified set are displayed in Figure 3. The qualitative features of this figure are the same as before, except now the Wald-style confidence region covers an even smaller portion of the identified set, and so inferences based on it may be even more misleading.

Of course, so far we have assumed a very generous upper bound on annual wages and salaries of $\overline{\text{wage}} = \10^8 . To assess how sensitive the qualitative results described above are to the value of $\overline{\text{wage}}$, we reexamine the previous case

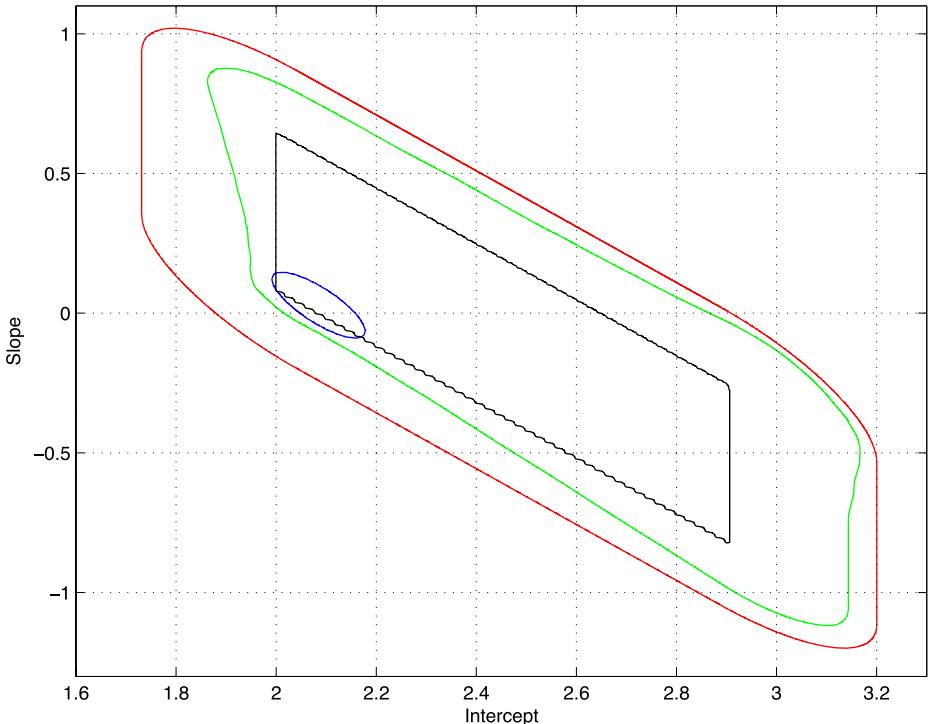


FIGURE 3.—Confidence regions with 10% top-coding and $\overline{\text{wage}} = \10^8 : green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region, black = identified set.

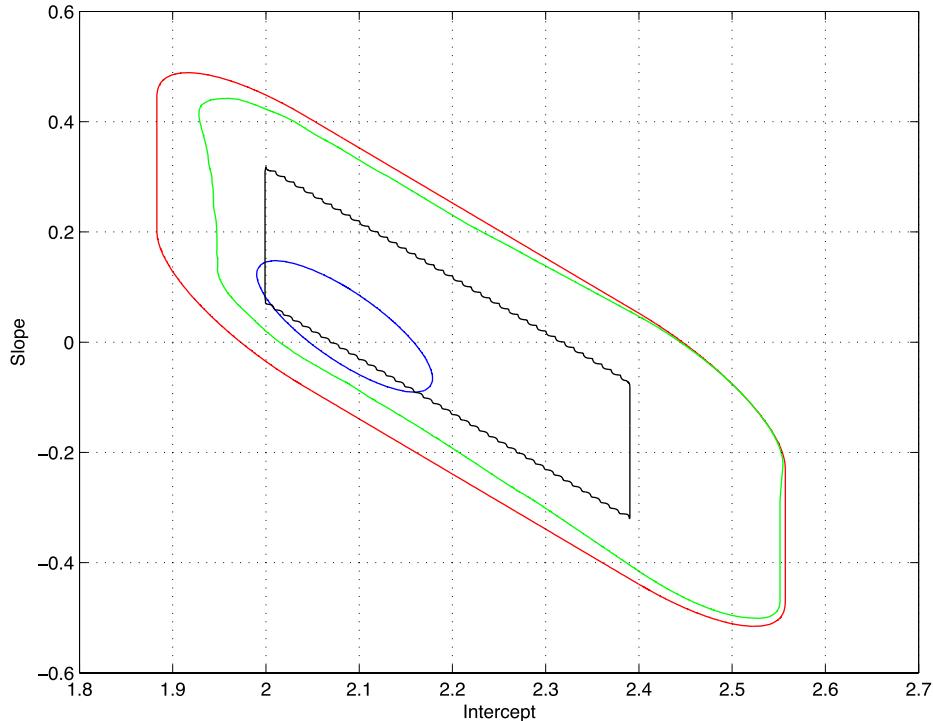


FIGURE 4.—Confidence regions with 5% top-coding and $\overline{\text{wage}} = \10^6 : green = confidence region for identifiable parameters, red = confidence region for identified set, blue = Wald-style confidence region, black = identified set.

in which 10% of the population is subject to top-coding with the much lower value of $\overline{\text{wage}} = \10^6 . Algorithm 2.1 converged after 12 steps and the confidence region for the identified set is then given by

$$\mathcal{C}_n = S_{12} = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 0.0094\}.$$

The confidence regions from this exercise along with the identified set are displayed in Figure 4. Again, the qualitative features of this figure are the same as before, but, as one would expect, the identified set is smaller than before. This suggests that in applications, the choice of $\overline{\text{wage}}$ is important, as it will noticeably impact the sharpness of inferences in such a setting.

4. SIMULATION RESULTS

In this section, we shed some light on the finite-sample behavior of our step-down procedure via a small simulation study. For the simulation study, we

use the final specification of the empirical illustration in which $\overline{\text{wage}} = \10^6 . As in the empirical illustration, the sample size n is 1000 and $\alpha = 0.05$. Following the discussion in Remark 2.5, in the first step of Algorithm 2.1, we set $S_1 = \{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq 1000\}$. To assess the sensitivity of our procedure to the choice of subsample size, we consider values of b in $\{15, 20, \dots, 85\}$. Finally, as described in Remark 2.1, we approximate the critical values with $B = 200$.

For each of 100 simulations, we compute the variables (i) j^* , the iteration at which Algorithm 2.1 converged, (ii) $\hat{c}_n(S_{j^*}, 1 - \alpha)$, the critical value that defines \mathcal{C}_n , (iii) $\hat{c}_n(S_{j^*-1}, 1 - \alpha)$, (iv) $\sup_{\Theta_0(P)} \hat{Q}_n(\theta)$, and (v) $I\{\Theta_0(P) \subseteq \mathcal{C}_n\}$.

In Table I, we present, for each value of b , (i) the average number of iterations needed for Algorithm 2.1 to converge, that is, the average value of \hat{j}^* , and (ii) the simulated probability that the identified set is covered by \mathcal{C}_n . The simulation results show that the average number of iterations increases with the subsample size, but it is typically between 7 and 10. The simulation results also show that the coverage probabilities are close to the nominal level, $1 - \alpha$, for values of b ranging from 20 to 40.

In Table II, we present, for each value of b , (i) the mean of $\hat{c}_n(S_{j^*}, 1 - \alpha)$, (ii) the mean of $\hat{c}_n(S_{j^*-1}, 1 - \alpha)$, and (iii) the simulated $1 - \alpha$ quantile of $\sup_{\Theta_0(P)} \hat{Q}_n(\theta)$. We label the third column “ideal” because it represents the best possible critical value. Of course, it is infeasible, since it depends on P , which is typically unknown. Fortunately, the simulation results show that for values of b between 20 and 40, $\hat{c}_n(S_{j^*}, 1 - \alpha)$ is close to this ideal value. The simulation results also allow for a comparison with single-step procedures. To see this, recall

TABLE I
SIMULATION RESULTS

b	Coverage Probability	Average Number of Iterations
15	0.98	6.99
20	0.94	7.02
25	0.97	7.64
30	0.91	7.98
35	0.98	8.04
40	0.96	8.14
45	0.97	8.87
50	1.00	8.98
55	0.98	9.08
60	0.99	9.20
65	0.99	9.77
70	0.98	9.95
75	0.99	10.02
80	1.00	10.08
85	1.00	10.24

TABLE II
SIMULATION RESULTS—CRITICAL VALUES

b	Avg. $\hat{c}_n(S_{j^*}, 1 - \alpha)$	Avg. $\hat{c}_n(S_{j^*-1}, 1 - \alpha)$	Ideal
15	0.0063	0.0066	0.0057
20	0.0069	0.0073	0.0066
25	0.0074	0.0077	0.0057
30	0.0076	0.0078	0.0077
35	0.0079	0.0082	0.0052
40	0.0084	0.0088	0.0064
45	0.0084	0.0086	0.0062
50	0.0089	0.0092	0.0047
55	0.0090	0.0093	0.0056
60	0.0093	0.0096	0.0060
65	0.0097	0.0099	0.0052
70	0.0098	0.0100	0.0071
75	0.0102	0.0105	0.0054
80	0.0103	0.0105	0.0058
85	0.0107	0.0109	0.0052

that for a single-step procedure to lead to a smaller confidence region than our step-down procedure, we would have to choose S_1 to be smaller than

$$\{\theta \in \mathbf{R}^2 : \hat{Q}_n(\theta) \leq \hat{c}_n(S_{j^*-1}, 1 - \alpha)\}.$$

Since $\hat{c}_n(S_{j^*-1}, 1 - \alpha)$ is typically very small in the simulation results, we would have to choose S_1 to be very small as well for the single-step procedure to lead to a smaller confidence region than our step-down procedure. Moreover, such a confidence region, though smaller than ours, may have poor coverage probabilities in finite samples.

APPENDIX

A.1. Auxiliary Results

LEMMA A.1: *Let X_1, \dots, X_n be a sequence of i.i.d. random variables with distribution P . Denote by $J_n(\cdot, P)$ the distribution of the statistic $\tau_n(\hat{\theta}_n - \theta(P))$. Suppose $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ is a symmetric function of its arguments and that $\hat{\theta}_n \in S$. Let \mathcal{V} be a Vapnik–Chervonenkis (VC) class of subsets of S with VC index v and assume that \mathcal{V} is permissible. For $0 < b < n$, let $N_n = \binom{n}{b}$ and let $k_n = \lfloor \frac{n}{b} \rfloor$. Then, for any $\varepsilon > 0$, we have that*

$$(28) \quad P \left\{ \sup_{V \in \mathcal{V}} \left| \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\{\tau_b(\hat{\theta}_{n,b,i} - \theta(P)) \in V\} - J_b(V, P) \right| > \varepsilon \right\}$$

is bounded above by

$$(29) \quad \frac{1}{\varepsilon} \left(\left(\sqrt{\frac{2}{k_n}} \vee 4 \sqrt{\frac{2 \log(8k_n^v)}{k_n}} \right) \wedge 1 \right) + \frac{32k_n^v}{\varepsilon} \sqrt{\frac{2\pi}{k_n}} \left[\Phi\left(\frac{4}{\sqrt{k_n}}\right) - \Phi\left(\left(\frac{1}{\sqrt{k_n}} \vee \sqrt{2 \log(8k_n^v)}\right) \wedge \frac{4}{\sqrt{k_n}}\right) \right],$$

where $\Phi(\cdot)$ is the standard normal distribution. We also have that for any $0 < \delta < 1$, (28) is bounded above by

$$(30) \quad \frac{\delta}{\varepsilon} + \frac{1}{\varepsilon} 8k_n^v \exp\left\{-\frac{k_n \delta^2}{32}\right\}$$

whenever $k_n \delta^2 \geq 2$.

PROOF: For $V \in \mathcal{V}$ define

$$\begin{aligned} S_n(V, P; X_1, \dots, X_n) \\ = \frac{1}{k_n} \sum_{1 \leq i \leq k_n} I\{\tau_b(\hat{\theta}_b(X_{b(i-1)+1}, \dots, X_{bi}) - \theta(P)) \in V\} - J_b(V, P). \end{aligned}$$

Denote by \mathcal{S}_n the symmetric group with n elements. Note that using this notation, we may rewrite

$$\frac{1}{N_n} \sum_{1 \leq i \leq N_n} I\{\tau_b(\hat{\theta}_{n,b,i} - \theta(P)) \in V\} - J_b(V, P)$$

as

$$Z_n(V, P; X_1, \dots, X_n) = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} S_n(V, P; X_{\pi(1)}, \dots, X_{\pi(n)}).$$

Note further that

$$\sup_{V \in \mathcal{V}} |Z_n(V, P; X_1, \dots, X_n)| \leq \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \sup_{V \in \mathcal{V}} |S_n(V, P; X_{\pi(1)}, \dots, X_{\pi(n)})|,$$

which is a sum of $n!$ identically distributed random variables. Let $\varepsilon > 0$ be given. It follows that

$$(31) \quad \begin{aligned} P\left\{\sup_{V \in \mathcal{V}} |Z_n(V, P; X_1, \dots, X_n)| > \varepsilon\right\} \\ \leq P\left\{\frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \sup_{V \in \mathcal{V}} |S_n(V, P; X_{\pi(1)}, \dots, X_{\pi(n)})| > \varepsilon\right\}. \end{aligned}$$

Using Markov's inequality, the right-hand side of (31) can be bounded by

$$(32) \quad \frac{1}{\varepsilon} E \left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| \right\}$$

$$(33) \quad = \frac{1}{\varepsilon} \int_0^1 P \left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| > u \right\} du.$$

Recall that the generalized Glivenko–Cantelli theorem asserts that

$$P \left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| > u \right\}$$

is bounded above by $8k_n^v \exp\{-k_n u^2/32\}$ whenever $k_n u^2 \geq 2$ and by 1 otherwise. It follows that (33) is bounded above by

$$\begin{aligned} & \frac{1}{\varepsilon} \int_0^{(\sqrt{2/k_n} \vee 4\sqrt{(2\log(8k_n^v))/k_n}) \wedge 1} 1 du \\ & + \frac{1}{\varepsilon} \int_{(\sqrt{2/k_n} \vee 4\sqrt{(2\log(8k_n^v))/k_n}) \wedge 1}^\infty 8k_n^v \exp\left\{-\frac{k_n u^2}{32}\right\} du. \end{aligned}$$

Evaluating this last expression yields the bound (29). To establish (30), note that for any $0 < \delta < 1$, we have that

$$\begin{aligned} (34) \quad & E \left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| \right\} \\ & \leq \delta + P \left\{ \sup_{V \in \mathcal{V}} |S_n(V, P; X_1, \dots, X_n)| > \delta \right\}. \end{aligned}$$

The result (30) now follows immediately by using an exponential inequality as found in the proof of the Glivenko–Cantelli theorem for VC classes (see Section 2 of Pollard (1984)) to bound the second term on the right-hand side in (34). *Q.E.D.*

LEMMA A.2: *Let F and $F_n, n \geq 1$, be distribution functions on \mathbf{R} . Suppose $F_n(x) \rightarrow F(x)$ for all $x \geq 0$ and that F is continuous on $(0, \infty)$. Then*

$$\sup_{x \geq 0} |F_n(x) - F(x)| \rightarrow 0.$$

For brevity, we omit the proof of this generalization of Polya's theorem.

A.2. Technical Details for Example 2.1

Note that we may write

$$(35) \quad \sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) = \sup_{\theta \in \Theta_0(P)} \sum_{1 \leq j \leq m} (Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)])_+^2,$$

where

$$(36) \quad Z_{n,j}(\theta, P) = \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (g_j(X_i, \theta) - E_P[g_j(X_i, \theta)]).$$

Let $0 > \lambda_n \rightarrow 0$, but so slowly that $\sqrt{n}\lambda_n \rightarrow -\infty$ and, for $K \subseteq \{1, \dots, m\}$, define

$$\begin{aligned} \Theta'_n(K, P) &= \{\theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \iff j \in K\}, \\ \Theta_n(K, P) &= \{\theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \text{ for all } j \in K\}, \\ \Theta_0(K, P) &= \{\theta \in \Theta_0(P) : E_P[g_j(X_i, \theta)] = 0 \text{ for all } j \in K\}. \end{aligned}$$

Note that

$$\begin{aligned} \Theta_0(P) &= \bigcup \{\Theta_0(K, P) : K \subseteq \{1, \dots, m\}\} \\ &= \bigcup \{\Theta'_n(K, P) : K \subseteq \{1, \dots, m\}\} \end{aligned}$$

and adopt the convention that the sum over the empty set and the supremum over the empty set are zero. Hence, (35) can be bounded from below as

$$\begin{aligned} (37) \quad &\max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{1 \leq j \leq m} (Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)])_+^2 \\ &\geq \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)])_+^2 \\ &= \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2. \end{aligned}$$

On the other hand, (35) can be bounded from above as

$$\begin{aligned} &\max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P)} \sum_{1 \leq j \leq m} (Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)])_+^2 \\ &\leq \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2_+ \\ &\quad + \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P)} \sum_{j \notin K} (Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)])_+^2 \end{aligned}$$

$$(38) \quad \leq \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2_+ \\ + \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P)} \sum_{j \notin K} (Z_{n,j}(\theta, P) + \sqrt{n}E_P[g_j(X_i, \theta)])^2_+$$

$$(39) \quad = \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2_+ + o_P(1).$$

To see the equality (39), note that

$$\sup_{\theta \in \Theta'_n(K, P)} \sum_{j \notin K} (Z_{n,j}(\theta, P) + \sqrt{n}E_P[g_j(X_i, \theta)])^2_+ \\ \leq m \left(\sup_{\theta \in \Theta_0(P)} \max_{1 \leq j \leq m} |Z_{n,j}(\theta, P)| + \sqrt{n}\lambda_n \right)^2_+.$$

By assumption (i),

$$(40) \quad \sup_{\theta \in \Theta_0(P)} \max_{1 \leq j \leq m} |Z_{n,j}(\theta, P)| = O_P(1).$$

It thus follows from the assumption that $\sqrt{n}\lambda_n \rightarrow -\infty$ that (38) tends in probability to zero, which in turn implies the equality (39).

Next, we argue that

$$(41) \quad \sup_{\theta \in \Theta_n(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2_+ - \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2_+ \xrightarrow{P} 0.$$

Since $\Theta_0(K, P) \subseteq \Theta_n(K, P)$, the left-hand side of (41) is bounded from below by zero. It therefore suffices to show that (41) is bounded from above by zero in probability. To this end, note that the left-hand side of (41) is bounded from above by

$$(42) \quad \sum_{j \in K} (Z_{n,j}(\hat{\theta}_n, P))^2_+ + \varepsilon_n - \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2_+,$$

where $0 < \varepsilon_n \rightarrow 0$ and $\hat{\theta}_n \in \Theta_n(K, P)$ are such that

$$\sup_{\theta \in \Theta_n(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))^2_+ < \sum_{j \in K} (Z_{n,j}(\hat{\theta}_n, P))^2_+ + \varepsilon_n.$$

It is therefore enough to show that (42) is bounded from above in probability by zero. To see this, suppose by way of contradiction that there exists $\delta > 0$ such that

$$P \left\{ \sum_{j \in K} (Z_{n,j}(\hat{\theta}_n, P))_+^2 - \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_{n,j}(\theta, P))_+^2 > \delta \right\} \not\rightarrow 0.$$

By assumption (ii), it follows that there exists a subsequence n_k and a $\hat{\theta} \in \Theta_0(K, P)$ such that $\rho_{P,j}(\hat{\theta}_{n_k}, \hat{\theta}) \rightarrow 0$ for all $j \in K$ and

$$(43) \quad P \left\{ \sum_{j \in K} (Z_{n_k,j}(\hat{\theta}_{n_k}, P))_+^2 - \sum_{j \in K} (Z_{n_k,j}(\hat{\theta}, P))_+^2 > \delta \right\} \not\rightarrow 0.$$

Since $f_K : \mathbf{R}^m \rightarrow \mathbf{R}$ given by $f_K(x) = \sum_{j \in K} (x_j)_+^2$ is a continuous function, it is uniformly continuous on a compact set. Hence, for any $M > 0$ there exists a $\omega = \omega(M) > 0$ for which the left-hand side of (43) is bounded from above by

$$(44) \quad P \left\{ \sup_{\theta \in \Theta_0(P)} \max_{j \in K} |Z_{n_k,j}(\theta, P)| > M \right\} \\ + P \left\{ \max_{j \in K} |Z_{n_k,j}(\hat{\theta}_{n_k}, P) - Z_{n_k,j}(\hat{\theta}, P)| > \omega \right\}.$$

But it follows from assumption (i) that $Z_n(\theta, P)$ is asymptotically equicontinuous in the sense that for every $1 \leq j \leq m$ and $\delta > 0$,

$$\lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} P \left\{ \sup_{\rho_{P,j}(\theta, \theta') < \eta} |Z_{n,j}(\theta, P) - Z_{n,j}(\theta', P)| > \delta \right\} = 0.$$

It thus follows from (40) and assumption (iii) that (44) tends to zero as $M \rightarrow \infty$ and $k \rightarrow \infty$. This yields a contradiction to (43), so (41) is established.

The asymptotic behavior of (35) is therefore given by the asymptotic behavior of (37). By assumption (i) and the continuous mapping theorem, (37) tends in distribution to

$$(45) \quad \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_j(\theta, P))_+^2,$$

where $Z(\theta, P) = (Z_1(\theta, P), \dots, Z_m(\theta, P))$ is a mean-zero multivariate Gaussian process with covariance kernel

$$\text{Cov}(Z_j(\theta, P), Z_{j'}(\theta', P)) = E_P[g_j(X_i, \theta)g_{j'}(X_i, \theta')].$$

It remains to show that (45) is continuous at its $1 - \alpha$ quantile for appropriate values of α . To analyze this question, first note that (45) is a convex function

of $Z(\theta, P)$. It therefore follows from Theorem 11.1 of [Davydov, Lifshits, and Smorodina \(1998\)](#) that the distribution of (45) is continuous everywhere except possibly at zero. Next, note that

$$P\left\{\max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_0(K, P)} \sum_{j \in K} (Z_j(\theta, P))_+^2 \leq 0\right\} \leq P\{(Z_{j^*}(\theta^*, P))_+^2 \leq 0\} \leq \frac{1}{2},$$

where j^* and θ^* are as in assumption (iii). Hence, (45) is continuous at its $1 - \alpha$ quantile for all $\alpha < \frac{1}{2}$.

A.3. Technical Details for Remark 2.7

Let (X_i, Y_i) , $i = 1, \dots, n$, be an i.i.d. sequence of random variables with distribution P on \mathbf{R}^2 . Let $\mu_X(P)$ denote the mean of the first component of the distribution P and let $\mu_Y(P)$ denote the mean of the second component of the distribution P . The parameter of interest, θ_0 , is known to satisfy $\mu_X(P) \leq \theta_0 \leq \mu_Y(P)$. The identified set is therefore given by $\Theta_0(P) = \{\theta \in \mathbf{R} : \mu_X(P) \leq \theta \leq \mu_Y(P)\}$. This set may be characterized as the set of minimizers of

$$Q(\theta, P) = (\mu_X(P) - \theta)_+^2 + (\theta - \mu_Y(P))_+^2.$$

The sample analog of $Q(\theta, P)$ is given by $\hat{Q}_n(\theta) = (\bar{X}_n - \theta)_+^2 + (\theta - \bar{Y}_n)_+^2$.

Let $a_n = n$ and suppose P is such that $E[|(X_i, Y_i)|^4] < \infty$ and that Cramer's condition holds, that is,

$$\limsup_{|s| \rightarrow \infty} |\psi_P(s)| < 1,$$

where $\psi_P(s)$ denotes the characteristic function of P . Assume further that $X_i \leq Y_i$ with probability 1 under P . This assumption is not essential, but it simplifies the analysis, while still allowing us to make our comparison. See Remark A.1 below for further discussion.

Let

$$\hat{\Theta}_{0,n} = \{\theta \in \mathbf{R} : n(\bar{X}_n - \theta)_+^2 + n(\theta - \bar{Y}_n)_+^2 \leq \lambda_n\},$$

where $\lambda_n > 0$ is an increasing sequence tending to infinity, but so slowly that $\lambda_n/n \rightarrow 0$. [Chernozhukov, Hong, and Tamer \(2007\)](#) suggested, for example, $\lambda_n = \log(n)$. Consider the confidence region given by

$$\mathcal{C}'_n = \{\theta \in \mathbf{R} : n(\bar{X}_n - \theta)_+ + n(\theta - \bar{Y}_n)_+^2 \leq \hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)\}.$$

To obtain a second-order accurate expression for $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$, first note that

$$(46) \quad \hat{\Theta}_{0,n} = \left\{ \theta \in \mathbf{R} : \bar{X}_n - \sqrt{\frac{\lambda_n}{n}} \leq \theta \leq \bar{Y}_n + \sqrt{\frac{\lambda_n}{n}} \right\}.$$

We therefore have that $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$ is the $1 - \alpha$ quantile of

$$L_n(x) = \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I \left\{ \max \left\{ b \left(\bar{X}_{n,b,i} - \bar{X}_n + \sqrt{\frac{\lambda_n}{n}} \right)_+^2, \right. \right.$$

$$\left. \left. b \left(\bar{Y}_n - \bar{Y}_{n,b,i} + \sqrt{\frac{\lambda_n}{n}} \right)_+^2 \right\} \leq x \right\},$$

which, for $x \geq 0$, we may rewrite as

$$\frac{1}{N_n} \sum_{1 \leq i \leq N_n} I \left\{ \sqrt{b} (\bar{X}_{n,b,i} - \bar{X}_n) \leq \sqrt{x} - \sqrt{\frac{\lambda_n b}{n}}, \right.$$

$$\left. \sqrt{b} (\bar{Y}_n - \bar{Y}_{n,b,i}) \leq \sqrt{x} - \sqrt{\frac{\lambda_n b}{n}} \right\}.$$

Now consider

$$\tilde{L}_n(x, y) = \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I \left\{ \sqrt{b} (\bar{X}_{n,b,i} - \bar{X}_n) \leq x, \sqrt{b} (\bar{Y}_n - \bar{Y}_{n,b,i}) \leq y \right\}$$

and the related

$$U_n(x, y) = \frac{1}{N_n} \sum_{1 \leq i \leq N_n} I \left\{ \sqrt{b} (\bar{X}_{n,b,i} - \mu_X(P)) \leq x, \right.$$

$$\left. \sqrt{b} (\mu_Y(P) - \bar{Y}_{n,b,i}) \leq y \right\}.$$

It follows from Lemma A.1 that

$$U_n(x, y) = J_b(x, y, P) + O_P \left(\sqrt{\frac{\log(k_n)}{k_n}} \right)$$

uniformly in x and y , where $J_b(x, y, P) = P \{ \sqrt{b} (\bar{X}_n - \mu_X(P)) \leq x, \sqrt{b} (\mu_Y(P) - \bar{Y}_n) \leq y \}$. To see this, simply take $\delta = 4\sqrt{5}\sqrt{\log(k_n)/k_n}$ in equation (30) of Lemma A.1. From Lemma 5.4 of Hall (1992), we have, under our above assumptions on P , that

$$J_b(x, y, P) = \Phi_{\sigma_X(P), \sigma_Y(P), \rho(P)}(x, y) + \frac{1}{\sqrt{b}} f(x, y, P) + O_P \left(\frac{1}{b} \right)$$

uniformly in x and y , where $\Phi_{\sigma_X(P), \sigma_Y(P), \rho(P)}(x, y)$ is the bivariate normal cumulative distribution function with variances $\sigma_X^2(P)$ and $\sigma_Y^2(P)$ and covariance $\rho(P)$, and $f(x, y, P)$ is a smooth function of x and y that depends on P

through its second- and third-order cumulants. Now, since $\sqrt{b/n}$ is of smaller order than $\sqrt{\log(k_n)/k_n}$,

$$\sqrt{b}|\bar{X}_n - \mu_X(P)| = O_P(\sqrt{b/n}) \leq O_P(\sqrt{\log(k_n)/k_n}),$$

and similarly for \bar{Y}_n . Using these facts, we can argue as in the proof of Lemma 2 of Bertail (1997) that

$$\tilde{L}_n(x, y) = U_n(x, y) = O_P(\sqrt{\log(k_n)/k_n}).$$

Therefore,

$$\begin{aligned}\tilde{L}_n(x, y) &= \Phi_{\sigma_X(P), \sigma_Y(P), \rho(P)}(x, y) + \frac{1}{\sqrt{b}}f(x, y, P) \\ &\quad + O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log(k_n)}{k_n}}\right)\end{aligned}$$

uniformly in x and y . From this expression for $\tilde{L}_n(x, y)$, we can deduce that $\hat{c}_n = \inf\{x \in \mathbf{R} : \tilde{L}_n(x, x) \geq 1 - \alpha\}$ satisfies

$$\hat{c}_n = c + \frac{\delta}{\sqrt{b}} + O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log(k_n)}{k_n}}\right),$$

where c is such that $\Phi_{\sigma_X(P), \sigma_Y(P), \rho(P)}(c, c) = 1 - \alpha$ and

$$\delta = -\frac{f(c, c)}{\nabla_{x,y}\Phi_{\sigma_X(P), \sigma_Y(P), \rho(P)}(x, y)|_{x=c, y=c}(1, 1)}.$$

Hence,

$$(47) \quad \hat{r}_n(\hat{\theta}_{0,n}, 1 - \alpha) = \left(\hat{c}_n + \sqrt{\frac{\lambda_n b}{n}}\right)^2$$

$$(48) \quad = \left(c + \frac{\delta}{\sqrt{b}} + \sqrt{\frac{\lambda_n b}{n}} + O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log k_n}{k_n}}\right)\right)^2.$$

Now consider the confidence region \mathcal{C}_n given by Algorithm 2.1. First note that

$$\sup_{\theta \in \hat{\theta}_{0,n}} a_n \hat{Q}_n(\theta) = \lambda_n.$$

From the above expression for $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$, we therefore have that

$$P\left\{\sup_{\theta \in \hat{\Theta}_{0,n}} a_n \hat{Q}_n(\theta) > \hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)\right\} \rightarrow 1.$$

It follows that \mathcal{C}_n is no larger than

$$\{\theta \in \mathbf{R} : n(\bar{X}_n - \theta)_+^2 + n(\theta - \bar{Y}_n)_+^2 \leq \hat{r}_n(\mathcal{C}'_n, 1 - \alpha)\}$$

with probability tending to 1. From the above analysis, we have immediately that a second-order accurate expression for $\hat{r}_n(\mathcal{C}'_n, 1 - \alpha)$ is

$$(49) \quad \left(c + \frac{\delta}{\sqrt{b}} + \sqrt{\frac{b}{n}} \left(c + \frac{\delta}{\sqrt{b}} + \sqrt{\frac{\lambda_n b}{n}}\right) + O_P\left(\frac{1}{b} \vee \sqrt{\frac{\log(k_n)}{k_n}}\right)\right)^2.$$

Since $\sqrt{\frac{b}{n}}(c + \frac{\delta}{\sqrt{b}} + \sqrt{\lambda_n b/n})$ is of smaller order than $\sqrt{\lambda_n b/n}$, we expect that $\hat{r}_n(\mathcal{C}'_n, 1 - \alpha)$ will be smaller to second order than $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$. To illustrate this, consider the special case in which $\lambda_n = \log(n)$, as suggested by Chernozhukov, Hong, and Tamer (2007), and $b = n^{1/3}$, as suggested by Politis, Romano, and Wolf (1999). In this case, (48) simplifies to

$$c + \frac{\delta}{\sqrt{b}} + O_P\left(\sqrt{\frac{\log(n)}{n^{2/3}}}\right),$$

whereas (49) simplifies to

$$c + \frac{\delta}{\sqrt{b}} + O_P\left(\sqrt{\frac{\log(n^{2/3})}{n^{2/3}}}\right).$$

Hence, $\hat{r}_n(\mathcal{C}'_n, 1 - \alpha)$ is smaller to second order than $\hat{r}_n(\hat{\Theta}_{0,n}, 1 - \alpha)$, as expected.

REMARK A.1: The assumption that $X_i \leq Y_i$ with probability 1 under P was only used in the second-order comparison of Remark 2.7 to express $\hat{\Theta}_{0,n}$ as in (46). If we were to change $\hat{Q}_n(\theta)$, then this assumption can be removed as well. For example, if we consider

$$\hat{Q}_n(\theta) = \max\{n(\bar{X}_n - \theta)_+^2, n(\theta - \bar{Y}_n)_+^2\},$$

then we can express $\hat{\Theta}_{0,n}$ as in (46) regardless of whether $X_i \leq Y_i$ with probability 1 under P or not. The rest of the analysis can be followed without any changes to illustrate the second-order benefits of the iterative approach.

A.4. Technical Details for Example 2.3

We begin with some preliminaries. We can write

$$\sup_{\theta \in \Theta_0(P)} a_n \hat{Q}_n(\theta) = \sup_{\theta \in \Theta_0(P)} \sum_{1 \leq j \leq m} (Z_{n,j}(\theta, P) + \sqrt{n} E_P[g_j(X_i, \theta)])_+^2,$$

where

$$Z_{n,j}(\theta, P) = \frac{1}{\sqrt{n}} \sum_{1 \leq i \leq n} (g_j(X_i, \theta) - E_P[g_j(X_i, \theta)]).$$

Let $0 > \lambda_n \rightarrow 0$, but in a way that $\sqrt{b}\lambda_n \rightarrow 0$ and $\sqrt{n}\lambda_n \rightarrow -\infty$. For $K \subseteq \{1, \dots, m\}$, define

$$\begin{aligned} \Theta'_n(K, P) &= \{\theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \iff j \in K\}, \\ \Theta_n(K, P) &= \{\theta \in \Theta_0(P) : \lambda_n < E_P[g_j(X_i, \theta)] \text{ for all } j \in K\}, \\ \Theta_0(K, P) &= \{\theta \in \Theta_0(P) : E_P[g_j(X_i, \theta)] = 0 \text{ for all } j \in K\}. \end{aligned}$$

Note that

$$\begin{aligned} \Theta_0(P) &= \bigcup \{\Theta_n(K, P) : K \subseteq \{1, \dots, m\}\} \\ &= \bigcup \{\Theta'_n(K, P) : K \subseteq \{1, \dots, m\}\}. \end{aligned}$$

Below we will adopt the convention that the sum over the empty set and the supremum over the empty set are zero.

Next, note that for any sequence $P_n \in \mathbf{P}$, we have that

$$\begin{aligned} (50) \quad & \sup_{\theta \in \Theta_0(P_n)} a_b \hat{Q}_b(\theta) \\ &= \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{1 \leq j \leq m} (Z_{b,j}(\theta, P_n) + \sqrt{b} E_{P_n}[g_j(X_i, \theta)])_+^2 \\ &\geq \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n) + \sqrt{b} E_{P_n}[g_j(X_i, \theta)])_+^2 \\ &\geq \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n)_+^2 \\ &= \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2 + \Delta_{1,n}(P_n), \end{aligned}$$

where

$$\begin{aligned}\Delta_{1,n}(P_n) &= \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n)_+^2 \\ &\quad - \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2.\end{aligned}$$

Moreover, $\Delta_{1,n}(P_n) = o_{P_n}(1)$. To see this, it suffices to show that for any $K \subseteq \{1, \dots, m\}$,

$$\begin{aligned}(51) \quad &\sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2 \\ &\quad - \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n)_+^2 = o_{P_n}(1).\end{aligned}$$

Since the left-hand side of (51) is bounded from below by

$$\sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2 - (Z_{b,j}(\theta, P_n) + \sqrt{b}\lambda_n)_+^2 \geq 0,$$

it suffices to show that (51) is bounded from above by zero in probability. To this end, let $\hat{\theta}_n \in \Theta_n(K, P_n)$ be such that

$$\sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{b,j}(\theta, P_n))_+^2 \leq \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n))_+^2 + \varepsilon_n$$

for $0 < \varepsilon_n \rightarrow 0$. Thus, (51) is bounded from above by

$$\sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n))_+^2 + \varepsilon_n - \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n) + \sqrt{b}\lambda_n)_+^2.$$

Since the function $f_K : \mathbf{R}^m \rightarrow \mathbf{R}$ defined by $f_K(x) = \sum_{j \in K} (x_j)_+^2$ is continuous, it is uniformly continuous on a compact set. Hence, for any $\delta > 0$ and $M > 0$, there exists $\omega = \omega(M) > 0$ such that

$$\begin{aligned}(52) \quad &P_n \left\{ \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n))_+^2 - \sum_{j \in K} (Z_{b,j}(\hat{\theta}_n, P_n) + \sqrt{b}\lambda_n)_+^2 > \delta \right\} \\ &\leq P_n \left\{ \max_{1 \leq j \leq m} |Z_{b,j}(\hat{\theta}_n, P_n)| > M \right\} + P_n \{ \sqrt{b}\lambda_n < -\omega \} \\ &\leq P_n \left\{ \sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{b,j}(\theta, P_n)| > M \right\} + I\{ \sqrt{b}\lambda_n < -\omega \}.\end{aligned}$$

By assumption (i),

$$\sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{b,j}(\theta, P_n)| = O_{P_n}(1).$$

It therefore follows from the assumption that $\sqrt{b}\lambda_n \rightarrow 0$ that (52) tends to zero as $n \rightarrow \infty$ and then $M \rightarrow \infty$. The desired claim thus follows.

Similarly, we have for any sequence $P_n \in \mathbf{P}$ that

$$\begin{aligned} (53) \quad & \sup_{\theta \in \Theta_0(P_n)} a_n \hat{Q}_n(\theta) \\ &= \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P_n)} \sum_{1 \leq j \leq m} (Z_{n,j}(\theta, P_n) + \sqrt{n}E_{P_n}[g_j(X_i, \theta)])_+^2 \\ &\leq \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P_n)} \sum_{j \in K} (Z_{n,j}(\theta, P_n))^2_+ \\ &\quad + \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P_n)} \sum_{j \notin K} (Z_{n,j}(\theta, P_n) + \sqrt{n}E_{P_n}[g_j(X_i, \theta)])_+^2 \\ &\leq \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_n(K, P_n)} \sum_{j \in K} (Z_{n,j}(\theta, P_n))^2_+ + \Delta_{2,n}(P_n), \end{aligned}$$

where

$$\Delta_{2,n}(P_n) = \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta'_n(K, P_n)} \sum_{j \notin K} (Z_{n,j}(\theta, P_n) + \sqrt{n}E_{P_n}[g_j(X_i, \theta)])_+^2.$$

Moreover, $\Delta_{2,n}(P_n) = o_{P_n}(1)$. To see this, note that

$$\begin{aligned} & \sup_{\theta \in \Theta'_n(K, P_n)} \sum_{j \notin K} (Z_{n,j}(\theta, P_n) + \sqrt{n}E_{P_n}[g_j(X_i, \theta)])_+^2 \\ &\leq m \left(\sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{n,j}(\theta, P_n)| + \sqrt{n}\lambda_n \right)_+^2. \end{aligned}$$

By assumption (i),

$$\sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{n,j}(\theta, P_n)| = O_{P_n}(1).$$

The desired claim thus follows from the assumption that $\sqrt{n}\lambda_n \rightarrow -\infty$. In fact, we have that $\Delta_{2,n}(P_n)$ is identically equal to zero with probability tending to 1.

We now use these facts to argue by contradiction that the required condition (21) holds. If the result were false, then there would exist a subsequence n_k and a corresponding sequence $P_{n_k} \in \mathbf{P}$ such that

$$\sup_{x \in \mathbb{R}} \{J_{b_{n_k}}(x, P_{n_k}) - J_{n_k}(x, P_{n_k})\} \rightarrow \delta$$

for some $\delta > 0$. It follows from (50) and (53) that

$$(54) \quad \sup_{x \in \mathbb{R}} \{ \tilde{J}_{b_{n_k}}(x, P_{n_k}) - \bar{J}_{n_k}(x, P_{n_k}) \} \not\rightarrow 0,$$

where

$$\begin{aligned} \tilde{J}_{b_{n_k}}(x, P_{n_k}) &= P_{n_k} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_k}(K, P_{n_k})} \sum_{j \in K} (Z_{b_{n_k}, j}(\theta, P_{n_k}))_+^2 \right. \\ &\quad \left. + \Delta_{1, n_k}(P_{n_k}) \leq x \right\}, \\ \bar{J}_{n_k}(x, P_{n_k}) &= P_{n_k} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_k}(K, P_{n_k})} \sum_{j \in K} (Z_{n_k, j}(\theta, P_{n_k}))_+^2 \right. \\ &\quad \left. + \Delta_{2, n_k}(P_{n_k}) \leq x \right\}. \end{aligned}$$

By assumption (ii) and Theorem 3.85 of Aliprantis and Border (2006), the set of all nonempty closed subsets of Θ is a compact metric space with respect to the Hausdorff metric

$$\bar{\rho}_H(A, B) = \inf\{\eta > 0 : A \subseteq B^\eta, B \subseteq A^\eta\},$$

where

$$A^\eta = \bigcup \{a' \in \Theta : \bar{\rho}(a', a) < \eta \text{ for some } a \in A\}.$$

Hence, for each $K \subseteq \{1, \dots, m\}$ for which $\Theta_{n_k}(K, P_{n_k})$ is nonempty infinitely often, there exists $\emptyset \neq \Theta^*(K) \subseteq \Theta$ such that $\Theta_{n_k}(K, P_{n_k})$ converges to $\Theta^*(K)$ under $\bar{\rho}_H$ along a subsequence. For any $K \subseteq \{1, \dots, m\}$ for which $\Theta_{n_k}(K, P_{n_k})$ is nonempty only finitely often, let $\Theta^*(K) = \emptyset$. Note that by assumption (iii) and the pigeonhole principle there exists $1 \leq j^* \leq m$ such that infinitely often there exists $\theta_{n_k}^* \in \Theta_{n_k}(\{j^*\}, P_{n_k})$ such that $\text{Var}_{P_{n_k}}[g_{j^*}(X_i, \theta_{n_k}^*)] \geq \varepsilon$.

By assumption (i), we have that

$$\sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{b_{n_k}, j}(\theta, P_{n_k})| = O_{P_{n_k}}(1),$$

$$\sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{n_k, j}(\theta, P_{n_k})| = O_{P_{n_k}}(1).$$

Hence, along a subsequence, both $Z_{b_{n_k}}(\theta, P_{n_k})$ and $Z_{n_k}(\theta, P_{n_k})$ converge to mean-zero multivariate Gaussian processes. Since the covariance kernel of these limiting processes is given by the limit of the covariance kernels of $Z_{b_{n_k}}(\theta, P_{n_k})$ and $Z_{n_k}(\theta, P_{n_k})$, which are identical, the covariance kernels of

the limiting processes must also be identical. The distributions of the limiting processes must therefore be identical.

In summary, there exists a subsequence n_{k_ℓ} along which $\rho_H(\Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}}), \Theta^*(K)) \rightarrow 0$ for all $K \subseteq \{1, \dots, m\}$, and $Z_{b_{n_{k_\ell}}}(\theta, P_{n_{k_\ell}})$ and $Z_{n_{k_\ell}}(\theta, P_{n_{k_\ell}})$ both converge in distribution $Z^*(\theta)$, a mean-zero multivariate Gaussian process. Let $J^*(x)$ denote the distribution function of

$$(55) \quad \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} (Z_j^*(\theta))^2_+.$$

Furthermore, there exist $1 \leq j^* \leq m$ and $\theta^* \in \Theta^*(\{j^*\})$ such that $Z_{j^*}^*(\theta^*)$ is not degenerate. Since (55) is a convex function of $Z^*(\theta)$, it follows from Theorem 11.1 of Davydov, Lifshits, and Smorodina (1998) that $J^*(x)$ is continuous on $(0, \infty)$.

To complete the argument, we argue that

$$\begin{aligned} \sup_{x \in \mathbf{R}} |\tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}}) - J^*(x)| &\rightarrow 0, \\ \sup_{x \in \mathbf{R}} |\tilde{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}}) - J^*(x)| &\rightarrow 0, \end{aligned}$$

which will lead to a contradiction of (54). Note that the distribution functions are identically equal to zero for $x < 0$, so it suffices to show that

$$\begin{aligned} \sup_{x \geq 0} |\tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}}) - J^*(x)| &\rightarrow 0, \\ \sup_{x \geq 0} |\tilde{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}}) - J^*(x)| &\rightarrow 0. \end{aligned}$$

For this purpose, we will use Lemma A.2. To apply this lemma, we first argue that

$$(56) \quad \begin{aligned} &\max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}))^2_+ \\ &- \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}))^2_+ \end{aligned}$$

and

$$(57) \quad \begin{aligned} &\max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}))^2_+ \\ &- \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} (Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}))^2_+ \end{aligned}$$

both converge in probability to zero. We only establish the result for (56), as essentially the same argument will establish the result for (57). It suffices to show that

$$(58) \quad \begin{aligned} & \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}))_+^2 \\ & - \sup_{\theta \in \Theta^*(K)} \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}))_+^2 = o_{P_{n_{k_\ell}}}(1). \end{aligned}$$

We first argue that the left-hand side of (58) is bounded above by zero in probability. To this end, let $\hat{\theta}_\ell \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})$ be such that

$$\sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}}))_+^2 \leq \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell, P_{n_{k_\ell}}))_+^2 + \varepsilon_\ell$$

for $0 < \varepsilon_\ell \rightarrow 0$, and let $\hat{\theta}_\ell^* \in \Theta^*(K)$ be such that $\bar{\rho}(\hat{\theta}_\ell, \hat{\theta}_\ell^*) \rightarrow 0$. With $\hat{\theta}_\ell$ and $\hat{\theta}_\ell^*$ so defined, the left-hand side of (58) is bounded above by

$$\sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell, P_{n_{k_\ell}}))_+^2 + \varepsilon_\ell - \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell^*, P_{n_{k_\ell}}))_+^2.$$

Since the function $f_K : \mathbf{R}^m \rightarrow \mathbf{R}$ defined by $f_K(x) = \sum_{j \in K} (x_j)_+^2$ is continuous, it is uniformly continuous on a compact set. Hence, for any $\delta > 0$ and $M > 0$, there exists $\omega = \omega(M) > 0$ such that

$$(59) \quad \begin{aligned} & P_{n_{k_\ell}} \left\{ \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell, P_{n_{k_\ell}}))_+^2 - \sum_{j \in K} (Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell^*, P_{n_{k_\ell}}))_+^2 > \delta \right\} \\ & \leq P_{n_{k_\ell}} \left\{ \max_{1 \leq j \leq m} |Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell, P_{n_{k_\ell}})| > M \right\} \\ & \quad + P_{n_{k_\ell}} \left\{ \max_{j \in K} |Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell, P_{n_{k_\ell}}) - Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell^*, P_{n_{k_\ell}})| > \omega \right\} \\ & \leq P_{n_{k_\ell}} \left\{ \sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}})| > M \right\} \\ & \quad + P_{n_{k_\ell}} \left\{ \max_{1 \leq j \leq m} |Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell, P_{n_{k_\ell}}) - Z_{b_{n_{k_\ell}}, j}(\hat{\theta}_\ell^*, P_{n_{k_\ell}})| > \omega \right\}. \end{aligned}$$

By assumption (i),

$$(60) \quad \sup_{\theta \in \Theta} \max_{1 \leq j \leq m} |Z_{b_{n_{k_\ell}}, j}(\theta, P_{n_{k_\ell}})| = O_{P_{n_{k_\ell}}}(1).$$

Assumption (i) also implies that for any $\omega > 0$ and all $1 \leq j \leq m$,

$$\lim_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}} P \left\{ \sup_{\rho_{P,j}(\theta, \theta') < \eta} |Z_{n,j}(\theta, P) - Z_{n,j}(\theta', P)| > \omega \right\} = 0.$$

It therefore follows from the assumption that $\bar{\rho}(\hat{\theta}_\ell, \hat{\theta}_\ell^*) \rightarrow 0$ that (59) tends to zero as $\ell \rightarrow \infty$ and then $M \rightarrow \infty$. By interchanging the roles of $\Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})$ and $\Theta^*(K)$, we see that the left-hand side of (58) is also bounded below by zero in probability. The desired result follows.

Since $\Delta_{1,n_{k_\ell}}(P_{n_{k_\ell}})$ and $\Delta_{2,n_{k_\ell}}(P_{n_{k_\ell}})$ both converge in probability to zero, we have by the continuous mapping theorem that $\tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}})$ and $\bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}})$ both converge in distribution to $J^*(x)$. Since $J^*(x)$ is continuous on $(0, \infty)$, we have for every $x > 0$ that

$$\begin{aligned} \tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}}) &\rightarrow J^*(x), \\ \bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}}) &\rightarrow J^*(x). \end{aligned}$$

To apply Lemma A.2, we therefore need only show that these convergences hold at $x = 0$.

Consider first $\tilde{J}_{b_{n_{k_\ell}}}(x, P_{n_{k_\ell}})$. Since $\Delta_{1,n_{k_\ell}}(P_{n_{k_\ell}}) \leq 0$, we have that

$$\begin{aligned} \tilde{J}_{b_{n_{k_\ell}}}(0, P_{n_{k_\ell}}) &\geq P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}))_+^2 \leq 0 \right\} \\ &= P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}) \leq 0 \right\}. \end{aligned}$$

Conversely, since $\Delta_{1,n_{k_\ell}}(P_{n_{k_\ell}}) = o_{P_{n_{k_\ell}}}(1)$, we have for any $\varepsilon > 0$ that

$$\begin{aligned} &\limsup_{\ell \rightarrow \infty} \tilde{J}_{b_{n_{k_\ell}}}(0, P_{n_{k_\ell}}) \\ &\leq \limsup_{\ell \rightarrow \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}))_+^2 \leq \varepsilon \right\} \\ &\leq \limsup_{\ell \rightarrow \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} (Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}))_+^2 \leq \varepsilon \right\} \\ &= \limsup_{\ell \rightarrow \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{b_{n_{k_\ell}},j}(\theta, P_{n_{k_\ell}}) \leq \sqrt{\varepsilon} \right\}. \end{aligned}$$

But arguing as was done to establish that (56) and (57) both tended to zero in probability, we have by the continuous mapping theorem that

$$(61) \quad \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}) \xrightarrow{\mathcal{L}} \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta^*(K)} \max_{j \in K} Z_j^*(\theta).$$

By letting $\varepsilon \rightarrow 0$ and noting that $J^*(0)$ is equal to the probability that (61) is less than or equal to zero, the desired result follows.

Now consider $\bar{J}_{n_{k_\ell}}(x, P_{n_{k_\ell}})$. Since $\Delta_{2,n_{k_\ell}}(P_{n_{k_\ell}})$ is identically equal to zero with probability tending to 1, we have that

$$\begin{aligned} & \lim_{\ell \rightarrow \infty} \bar{J}_{n_{k_\ell}}(0, P_{n_{k_\ell}}) \\ & \leq \lim_{\ell \rightarrow \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \sum_{j \in K} (Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}))_+^2 \leq 0 \right\} \\ & = \lim_{\ell \rightarrow \infty} P_{n_{k_\ell}} \left\{ \max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}}) \leq 0 \right\}. \end{aligned}$$

But arguing as was done to establish that (56) and (57) both tended to zero in probability, we have by the continuous mapping theorem that

$$\max_{K \subseteq \{1, \dots, m\}} \sup_{\theta \in \Theta_{n_{k_\ell}}(K, P_{n_{k_\ell}})} \max_{j \in K} Z_{n_{k_\ell}, j}(\theta, P_{n_{k_\ell}})$$

converges in distribution to the right-hand side of (61). To complete the proof, note that $J^*(0)$ is equal to the probability that the right-hand side of (61) is less than or equal to zero.

REFERENCES

- ALIPRANTIS, C. D., AND K. C. BORDER (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*. New York: Springer. [206]
- ANDREWS, D. W. K. (2000): "Inconsistency of the Bootstrap When a Parameter Is on the Boundary of the Parameter Space," *Econometrica*, 68, 399–405. [180]
- BAHADUR, R. R., AND L. J. SAVAGE (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Annals of Mathematical Statistics*, 27, 1115–1122. [170]
- BERTAIL, P. (1997): "Second-Order Properties of an Extrapolated Bootstrap Without Replacement Under Weak Assumptions," *Bernoulli*, 3, 149–179. [201]
- BUGNI, F. (2007): "Bootstrap Inference in Partially Identified Models," Working Paper. [180]
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2004): "Inference for Identified Parameter Sets in Econometric Models," Working Paper, Department of Economics, MIT. [186]
- (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284. [170, 180, 181, 199, 202]
- DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local Properties of Distributions of Stochastic Functionals*. Providence, RI: American Mathematical Society. [179, 199, 207]
- HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*. New York: Springer. [200]

- IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857. [171]
- KATZ, L. F., AND D. H. AUTOR (1999): "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, Vol. 3A, ed. by O. Ashenfelter and D. Card. Amsterdam: North-Holland, 1463–1555. [186]
- LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer. [185]
- MANSKI, C. F., AND E. TAMER (2002): "Inference on Regressions With Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–546. [178]
- POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer. [173, 176, 177, 187, 202]
- POLLARD, D. (1984): *Convergence of Stochastic Processes*. New York: Springer. [195]
- ROMANO, J. P. (2004): "On Non-Parametric Testing, the Uniform Behavior of the t -Test, and Related Problems," *Scandinavian Journal of Statistics*, 31, 567–584. [171]
- ROMANO, J. P., AND A. M. SHAIKH (2006): "Inference for the Identified Set in Partially Identified Econometric Models," Technical Report 2006-10, Department of Statistics, Stanford University. [180]
- _____, (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference—Special Issue in Honor of Ted Anderson*, 138, 2786–2807. [170, 171, 182, 187]
- _____, (2010): "Supplement to 'Inference for the Identified Set in Partially Identified Econometric Models,'" *Econometrica Supplemental Material*, 78, http://www.econometricsociety.org/ecta/Supmat/6706_data_and_programs.zip. [172]
- ROMANO, J. P., AND M. WOLF (2005): "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing," *Journal of the American Statistical Association*, 100, 94–108. [175]
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer. [178, 183]

Depts. of Economics and Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305-4065, U.S.A.; romano@stanford.edu

and

Dept. of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, U.S.A.; amshaikh@uchicago.edu.

Manuscript received September, 2006; final revision received September, 2009.

A MODEL OF DELEGATED PROJECT CHOICE

BY MARK ARMSTRONG AND JOHN VICKERS¹

We present a model in which a principal delegates the choice of project to an agent with different preferences. The principal determines the set of projects from which the agent may choose. The principal can verify the characteristics of the project chosen by the agent, but does not know which other projects were available to the agent. We consider situations where the collection of available projects is exogenous to the agent but uncertain, where the agent must invest effort to discover a project, where the principal can pay the agent to choose a desirable project, and where the principal can adopt more complex schemes than simple permission sets.

KEYWORDS: Delegation, principal–agent, rules, merger policy.

1. INTRODUCTION

IN THE MAIN MODEL in this paper, we present an analysis of a principal–agent problem in which the principal can influence the agent’s behavior not by outcome-contingent rewards, but by specifying what the agent is and is not allowed to do. The agent, whose preferences differ from those of the principal, will select from her available projects the permitted project that best serves her interests. The principal can verify whether or not the selected project is indeed within the permitted set, but cannot observe the number or characteristics of the other projects available to the agent. This paper investigates how the principal should best specify the set of projects from which the agent can choose.

An application of our analysis is to an important issue in competition policy, which is the appropriate welfare standard to use when evaluating mergers (or some other form of conduct). The two leading contenders are a *total welfare* standard, where mergers are evaluated according to whether they decrease the unweighted sum of producer and consumer surplus, and a *consumer welfare* standard, where mergers detrimental to consumers are blocked. Many economic commentators feel that antitrust policy should aim to maximize total welfare, whereas in many jurisdictions the focus is more on consumer welfare alone. See [Farrell and Katz \(2006\)](#) for an overview of the issues. One purpose of this paper is to examine a particular strategic reason, discussed previously by [Lyons \(2002\)](#) and [Fridolfsson \(2007\)](#), to depart from the regulator’s true welfare standard, which is that a firm may have a *choice* of merger possibilities. A less profitable merger might be better for total welfare, but will not be chosen under a total welfare standard.

¹We are grateful to Daron Acemoglu, V. Bhaskar, Sandro Brusco, Martin Cripps, James Dow, Florian Englmaier, Bengt Holmstrom, Michael Katz, Niko Matouschek, Meg Meyer, Tima Mylovanov, Barry Nalebuff, Eric Rasmusen, David Sappington, Karl Schlag, Aggey Semenov, Joel Sobel, Glen Weyl, Jidong Zhou, and to three referees for comments and discussion. Armstrong gratefully acknowledges the support of the Economic and Social Research Council (U.K.).

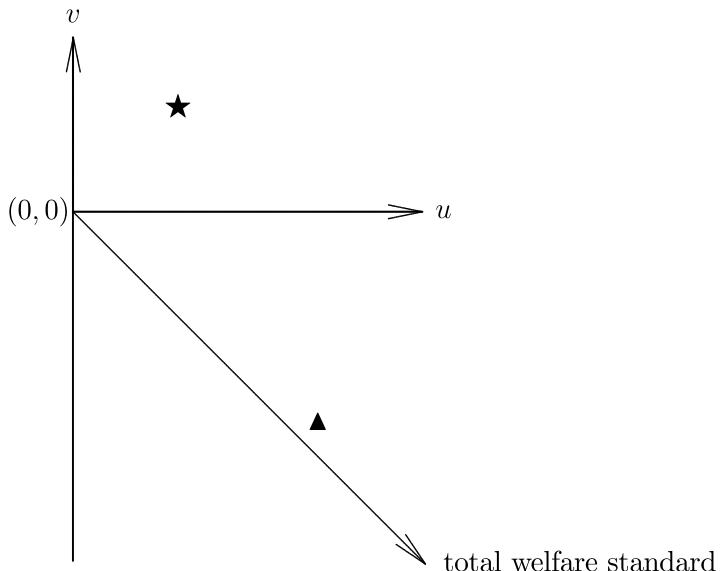


FIGURE 1.—The impact of welfare standard on chosen mergers.

To illustrate, consider Figure 1. Here, u represents the gain in total profit resulting from a particular merger, while v measures the resulting gain (which may be negative) to consumers. Suppose that u and v are verifiable once a merger is proposed to the competition authority. If the regulator follows a total welfare standard, he will permit any merger which lies above the negatively sloped line in the figure. Suppose the firm can choose between the two mergers depicted by \blacktriangle and \star on the figure. With a total welfare standard, the firm will choose the merger with the higher u payoff, that is, the \blacktriangle merger. However, the regulator would prefer the alternative \star since that yields higher total welfare. If the regulator instead imposed a consumer welfare standard, so that only those mergers which lie above the horizontal line $v = 0$ are permitted, then the firm will be forced to choose the preferred merger. In this case, a regulator wishing to maximize total welfare is better off if he imposes a consumer welfare standard. As Farrell and Katz (2006, p. 17) put it, “if we want to maximize gains in total surplus (northeasterly movements as shown in Figure 1) and firms always push eastwards, there is something to be said for someone adding a northerly force.” Nevertheless, there is a potential cost to adopting a consumer welfare standard: if the \blacktriangle merger turns out to be the *only* merger possibility, then a consumer welfare standard will not permit this, even though the merger will improve total welfare. Thus, the choice of welfare standard will depend

on the likely number of possible mergers and the distribution of profit and consumer surplus gains for a possible merger.²

Another application of our analysis could be to project choice within an organization. While shareholders wish to choose the available project which yields the highest net present value (NPV), a manager might prefer larger, more capital intensive projects. If the manager sometimes has a choice of project and has the ability to hide her less preferred projects, shareholders may wish to limit the kinds of projects which can be implemented. This question was analyzed by Berkovitch and Israel (2004, p. 241), who wrote: “If headquarters cannot observe all available projects, then the manager may manipulate the selection process by presenting projects such that managerial utility is maximized. [...] While NPV is the best way to *measure* value added, in many situations, it is not a good way to *implement* the selection of the highest NPV projects.”

More generally, our analysis addresses an aspect of the theory of optimal *rules*: the relationship between the ultimate objective of the rule-setter and the optimal rule to commit to. That relationship is not straightforward inasmuch as the likely consequences of a rule, including for the attainment of the ultimate objective, depend on the responses of agents seeking to maximize, within the rules, their own objectives. The interplay between rules and the responses that they induce is at the heart of our analysis. Our goal is to characterize optimal rules—which are sometimes strikingly simple—in terms of the fundamentals of the models.

Our benchmark model in Section 2 analyzes a setting in which the agent chooses one project from an exogenous, but uncertain, finite set of available projects. Monetary incentives are ruled out, and the principal optimally restricts agent choice in a way that forbids some projects that are moderately good, in the hope of inducing the agent instead to choose a project that is better for the principal. This bias is akin to putting less weight on the agent’s payoff than is in the true welfare function.

In Section 3 we present three variants of this benchmark model. In Section 3.1 we analyze a setting where the agent influences the likelihood of finding a project by exerting costly effort. Here, we show that the principal optimally sets a *linear* permission rule. In addition, to induce greater effort, the principal allows some projects that are detrimental to his interests; this bias is akin to putting more weight on the agent’s payoff than is in his true welfare function. Second, in Section 3.2 we discuss the impact of monetary incentives

²Besides administrative filing fees, merging firms do not make monetary payments to competition authorities, still less payments contingent on merger approval. Approval is sometimes made conditional upon the implementation of (nonmonetary) remedies, such as the sale to third parties of some assets. Our schematic framework would need adaptation to allow for this, although the issue of the appropriate welfare standard nevertheless applies to the questions about merger remedies in that they are designed to ensure that mergers are not likely to be detrimental in terms of the chosen standard.

to choose a good project. When the agent is liquidity constrained, it may be preferable to restrict the agent's freedom to choose projects than to reward her for choosing a good project. Finally, in Section 3.3 we consider the possible benefits to the principal of using a more complex delegation scheme. For instance, the principal may sometimes do better if he permits the implementation of a mediocre project when the agent reports she has several other mediocre projects available. However, when the number of projects follows a Poisson distribution, the principal can do no better than to offer a simple permission set.

Some other papers have examined situations in which a principal delegates decision-making to a (potentially) better-informed agent whose preferences differ from those of the principal, and where contingent transfers between principal and agent are ruled out. Aghion and Tirole (1997) showed how, depending on information structure and payoff alignment, it may be optimal for a principal to delegate full decision-making power to a potentially better-informed agent. The principal's loss of control over project choice can be outweighed by advantages in terms of encouraging the agent's initiative to develop and gather information about projects. In like vein Baker, Gibbons, and Murphy (1999), though they deny formal delegation of authority, examined informal delegation through repeated-game relational contracts. Even an informed principal able to observe project payoffs may refrain from vetoing ones that yield him poor payoffs so as to promote search incentives for the agent.

Our work is closer to the models which analyze *constrained* delegation, where the agent can make decisions but only within specified limits and the principal's problem is to decide how much leeway to give the agent. (For instance, a judge sets a convicted criminal's punishment, but only within mandatory minimum and/or maximum limits for the type of crime.) This literature was initiated by Holmstrom (1984), and the elements of his model go as follows. There is a set of decisions, indexed by a scalar variable d which takes values in some interval D , one of which needs to be made. Unlike our model where the agent must choose from a finite set of projects, here *any* decision is feasible. A given decision generates payoffs to the two parties which depend on the state of the world, represented by θ , and only the agent observes this parameter. The preferences of the principal and agent differ, and if decision d is made when the state is θ , the principal obtains payoff $V(d, \theta)$ and the agent has payoff $U(d, \theta)$. The principal's problem is to choose a set, say $\mathcal{D} \subset D$, from which the agent is permitted to choose her decision. This permission set is chosen to maximize the principal's expected payoff (given his prior on θ), given that the agent will make her preferred decision from \mathcal{D} given the state θ .³

³This "delegation problem" coincides with the "mechanism design problem" where the agent makes an announcement about the claimed state of the world, $\hat{\theta}$, and the principal commits to a rule $d(\hat{\theta})$ which maps the announcement to the implemented decision. The two approaches

Holmstrom mostly limits attention to cases where the permission set \mathcal{D} is an interval. Subject to this assumption (and other regularity conditions), he shows that an agent whose preferences are closer to the principal's will be given wider discretion. (This result has subsequently sometimes been termed the “ally principle.”) Following Holmstrom's initial contribution, subsequent papers have analyzed when interval delegation is optimal for the principal, making the additional assumption that θ is a scalar variable.⁴ Melumad and Shibano (1991) were the first to calculate optimal permission sets, in the special case where preferences were quadratic and where θ was uniformly distributed. They found that interval delegation was optimal when principal and agent have ideal policies which are similarly responsive to the state θ , but that otherwise it could be optimal to have “holes” in \mathcal{D} . Martimort and Semenov (2006) found a sufficient condition on the distribution of θ for interval delegation to be optimal. Alonso and Matouschek (2008) systematically investigated when interval delegation is optimal, and they generalized Melumad and Shibano's insight that the relative responsiveness of preferred decisions to the state is the key factor for this. They showed that when interval delegation is suboptimal, the ally principle need not hold and an agent with preferences more aligned with those of the principal might optimally be given less discretion.

Those models in the Holmstrom tradition differ from ours with respect to the actions which are feasible and the form of asymmetric information. In particular, they characterize each decision by a scalar parameter (such as the length of a prison sentence), all decisions are always feasible, and the agent has private information about a payoff-relevant state of the world. In our model, by contrast, payoffs of the chosen project to both principal and agent are known, but only a finite number of projects are feasible (such as the possible mergers for a firm) and only the agent knows what those possible projects are. Like the papers discussed above, our aim is to characterize the optimal permission set from which the agent can choose, but in a two-dimensional setting where the principal can observe both his own and the agent's payoff from the project chosen by the agent.⁵

are equivalent since the principal never directly observes the true θ and by making a suitable announcement $\hat{\theta}$ the agent can implement any decision in the range of the rule $d(\cdot)$.

⁴Szalay (2005) presented an interesting variant on this delegation problem in which interval delegation is often suboptimal. In his model, there is no divergence in preferences between the principal and agent, but the agent incurs a private cost to observe θ . He showed that it can be optimal for the principal to remove intermediate policies from \mathcal{D} so that the agent is forced to choose between relatively extreme options, as this sharpens the agent's incentive to discover θ .

⁵A paper which also investigates a two-dimensional delegation problem is Amador, Werning, and Angeletos (2006). There, an agent with quasihyperbolic preferences has wealth which she consumes over two periods. If there were no uncertainty about her preferences, she would gain by committing to a fixed consumption path at time zero. However, she will receive a utility shock in period 1 and this uncertainty gives a motive to allow some flexibility in consumption. Amador, Werning, and Angeletos found a condition which implies that the optimal permission set simply involves placing a ceiling on first-period consumption.

2. BENCHMARK MODEL: CHOOSING A PROJECT

A principal (“he”) delegates the choice of project to an agent (“she”). There may be several projects for the agent to choose from, although only one can be implemented over the relevant time horizon. A project is fully described by two scalar parameters, u and v . The agent’s payoff if the type- (u, v) project is implemented is u , while the payoff to the principal, who is assumed to be risk-neutral, is $v + \alpha u$. (If no project is implemented, each party obtains a payoff of zero.) Here, $\alpha \geq 0$ represents the weight the principal places on the agent’s interests and v represents factors specific to the principal’s interests. The parameter α might reflect a true regard for the agent’s payoff (as in the merger application when profits carry some weight in social welfare) and/or it might reflect a trade-off between allowing the agent wider project choice—and so a greater chance of on-the-job benefits u —and paying her a higher (noncontingent) salary.⁶ For example, if the principal jointly chooses the permitted set of projects and the salary to meet the participation constraint of a risk-neutral agent, and u is measured in money terms, then the following analysis applies with $\alpha = 1$.⁷

Each project is an independent draw from the same distribution for (u, v) . Since without contingent money rewards the agent will never propose a project with a negative payoff, without loss of generality we suppose that only nonnegative u are realized. The marginal density of $u \geq 0$ is $f(u)$. The conditional density of v given u is denoted $g(v, u)$ and the associated conditional distribution function for v is $G(v, u)$. Here, v can be positive or negative. Suppose that the support of (u, v) is a rectangle $[0, u_{\max}] \times [v_{\min}, v_{\max}]$, where $v_{\min} \leq 0 \leq v_{\max}$ so that $(0, 0)$ lies in the support of (u, v) . Finally, suppose that both f and g are continuously differentiable on the support of (u, v) .

In this benchmark model, the number of projects is random and the probability that the agent has exactly $n \geq 0$ available projects is q_n . (Our analysis covers the case where there are surely N projects, but the analysis is no easier for that case. Indeed, we will see that n being a Poisson variable is the easiest example to analyze.) Suppose that the project characteristics (u, v) are distributed independently of n .

The principal delegates the choice of project to the agent. We assume in this benchmark model that the principal cannot offer contingent monetary incentives to the agent to choose a desirable project, and also that the characteristics of the chosen project—and *only* that project—are verifiable. This latter assumption may be appropriate if the principal can verify the claimed project characteristics only after the project has been implemented. (Suppose that the

⁶Aghion and Tirole (1997) referred to this benefit of giving the agent freedom to choose projects as the “participation” benefit of delegation.

⁷At the other extreme, if the agent is infinitely risk-averse and cares only about her minimum income over all outcomes, then the following analysis applies when $\alpha = 0$. Such an agent is unwilling to trade off her salary against the uncertain prospect of on-the-job benefits.

principal has the ability to punish the agent if it turns out that the agent misrepresented the payoffs.) Alternatively, the assumption is reasonable if the principal has substantial costs associated with auditing each project's characteristics, and/or the agent has significant costs associated with preparing a credible proposal for an additional project. (Each of these is arguably the case in the merger scenario, for example.) We also assume that the principal considers only deterministic policies.⁸

Under the assumption that only the chosen project's characteristics are verifiable, the principal's (deterministic) problem reduces to the delegation problem of choosing a set of permitted projects.⁹ That is, before the agent has any private information, the principal commits to a (measurable) permission set of projects, denoted $\mathcal{D} \subset [0, u_{\max}] \times [v_{\min}, v_{\max}]$, and the agent can then implement any project that lies in \mathcal{D} .¹⁰ In Section 3.3 we discuss an alternative, and sometimes superior, delegation scheme which can be used when the principal can cheaply verify a *list* of projects which the agent reveals to be available.

Given a permission set \mathcal{D} , for each u let $\mathcal{D}_u = \{v \text{ such that } (u, v) \in \mathcal{D}\}$ be the set of type- u projects which are permitted and let

$$p(u) = \int_{v \in \mathcal{D}_u} g(v, u) dv$$

⁸We restrict attention to deterministic policies mainly because it is hard to imagine being able to commit to or implement a stochastic mechanism in practice. It is possible that a stochastic scheme, if feasible, could do better than a deterministic scheme. For example, suppose that $\alpha = 0$, that $n = 2$ for sure, and that $(u, v) = (0.5, 1)$ with probability 0.5 and $(u, v) = (0.9, 0.1)$ with the same probability. In this case, the optimal deterministic scheme only permits the principal's favored project, $(0.5, 1)$. However, permitting the project $(0.9, 0.1)$ with probability 0.5 yields the principal a higher expected payoff than banning it altogether: in both cases the agent would choose the principal's preferred project if she could, but if that is not available the stochastic scheme would still allow some chance of a desirable project being chosen.

⁹Under this assumption, similarly to footnote 3, this delegation approach is equivalent to a mechanism design approach in which the principal commits to a rule that determines which project is chosen as a function of the agent's report of her private information, that is, the number and characteristics of available projects. To see this, note that the set of projects can be partitioned into two subsets: the set of projects, say \mathcal{D} , which, by making suitable reports of other projects, could be chosen for implementation under the principal's decision rule and those projects which are never implemented by the principal's rule. Faced with this rule, the agent will simply choose her preferred available project (if any) in the former set and announce any other projects required to implement that choice. Clearly, this mechanism is equivalent to the delegation problem where the agent can directly choose any project in the set \mathcal{D} .

¹⁰It is important to emphasize the assumption here, as in delegation problems more generally, of commitment. (However, see Section 3.3 below for discussion about how the principal does not want to renegotiate the permission set in the case where the number of projects follows a Poisson distribution.) Baker, Gibbons, and Murphy (1999) and Alonso and Matouschek (2007) showed how the principal's commitment power can be endogenously generated with repeated interaction (as is the case in the merger context, for instance).

be the proportion of type- u projects which are permitted. Let

$$x(u) = 1 - \int_u^{u_{\max}} p(z)f(z) dz$$

be the probability that any given project either has agent payoff less than u or is not permitted. Note that $x(0)$ is the fraction of project types which are banned, that $x(\cdot)$ is continuous, and, when differentiable, its derivative is

$$(1) \quad x'(u) = p(u)f(u).$$

(Since $x(\cdot)$ is weakly increasing, it is differentiable almost everywhere.) If there are n available projects, the probability that each project is either banned or generates agent payoff less than u is $(x(u))^n$. Summing over n implies that the probability that each available project is either banned or generates agent payoff less than u is $\phi(x(u))$, where

$$\phi(x) \equiv \sum_{n=0}^{\infty} q_n x^n$$

is the *probability generating function* (PGF) associated with the random variable n . It follows that the density of the agent's preferred permitted project (where this exists) is $\frac{d}{du}\phi(x(u))$. Useful properties of PGFs are that they are well defined on the relevant interval $0 \leq x \leq 1$ and are smooth, convex, and increasing over this interval.

The principal's payoff with permission set \mathcal{D} is therefore

$$(2) \quad \begin{aligned} & \int_0^{u_{\max}} \left\{ E[v \mid (u, v) \in \mathcal{D}] + \alpha u \right\} \frac{d}{du} \phi(x(u)) du \\ &= \int_0^{u_{\max}} \left\{ \int_{v \in \mathcal{D}_u} v g(v, u) dv + \alpha u p(u) \right\} f(u) \phi'(x(u)) du. \end{aligned}$$

The principal's problem is to maximize expression (2), taking into account the relationship between p and x in (1) and the endpoint constraint $x(u_{\max}) = 1$. The following lemma shows that the optimal permission set takes a "threshold" form:

LEMMA 1: *In the optimal policy, there exists a threshold rule $r(\cdot)$ such that*

$$(u, v) \in \mathcal{D} \quad \text{if and only if} \quad v \geq r(u).$$

PROOF: From (1), the function $x(\cdot)$ depends on \mathcal{D} only via the "sufficient statistic" $p(u)$, not on the particular v projects which are permitted given u .

Therefore, for any candidate function $p(u)$, the principal might as well permit those particular v projects which maximize the term $\{\cdot\}$ in (2), subject to the constraint that the proportion of type- u projects is $p(u)$. But the problem of choosing the set \mathcal{D}_u so as to

$$\text{maximize } \int_{v \in \mathcal{D}_u} vg(v, u) dv \quad \text{subject to } \int_{v \in \mathcal{D}_u} g(v, u) dv = p(u)$$

is solved by permitting the projects with the highest v so that the proportion of permitted projects is $p(u)$, that is, that $\mathcal{D}_u = \{v \text{ such that } v \geq r(u)\}$ for some $r(u)$. *Q.E.D.*

Thus, the problem simplifies to the choice of threshold rule $r(\cdot)$ rather than the choice of permission set \mathcal{D} . (A similar argument is valid in the project discovery model in Section 3.1.) Figure 2 depicts a threshold rule $r(\cdot)$ and shows $x(u)$ as the measure of the shaded area.

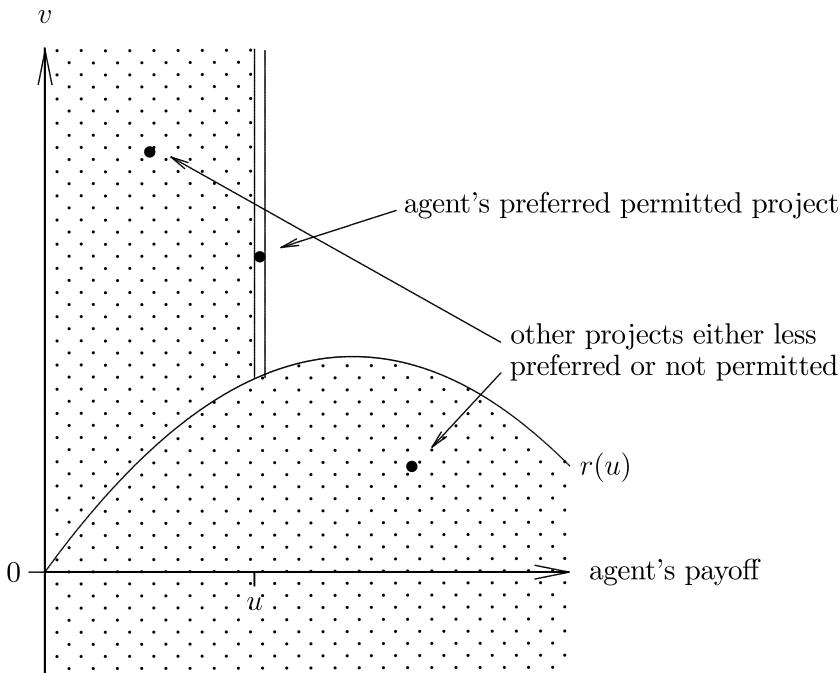


FIGURE 2.—The agent's preferred permitted project.

Define $V(r, u)$ to be the expected value of v given that the project has agent payoff u and that v is at least r . Recasting (2) in terms of $r(\cdot)$ rather than \mathcal{D} , the principal's problem is to choose $r(\cdot)$ to maximize

$$(3) \quad \int_0^{u_{\max}} [V(r(u), u) + \alpha u] [1 - G(r(u), u)] f(u) \phi'(x(u)) du$$

subject to the "equation of motion"

$$(4) \quad x'(u) = [1 - G(r(u), u)] f(u)$$

and the endpoint condition

$$(5) \quad x(u_{\max}) = 1.$$

This optimal control problem is solved formally in the Appendix, but its solution can be understood intuitively with the following argument. Consider a point $(u, r(u))$ on the frontier of the permitted set. For the rule to be optimal, it is necessary that the principal be indifferent between his payoff $[r(u) + \alpha u]$ at that point and his expected payoff from the agent's next-best permitted alternative, conditional on the agent's best permitted project being $(u, r(u))$.

To calculate the latter expected payoff, note that for a given project, the density of payoff vector $(u, r(u))$ is $f(u)g(r(u), u)$. Then the probability that one out of n projects has payoffs $(u, r(u))$ and all the others have agent utility no greater than $z \leq u$ or are not permitted is $nf(u)g(r(u), u)[x(z)]^{n-1}$. Taking the sum across n , the probability that one project has payoffs $(u, r(u))$ and all other permitted projects have utility no greater than z is therefore $f(u)g(r(u), u)\phi'(x(z))$. In particular, the probability that the agent's preferred permitted project is $(u, r(u))$ is $f(u)g(r(u), u)\phi'(x(u))$. Conditional on that event, the probability that the next-best permitted alternative for the agent has agent utility no greater than $z \leq u$ is

$$\frac{f(u)g(r(u), u)\phi'(x(z))}{f(u)g(r(u), u)\phi'(x(u))} = \frac{\phi'(x(z))}{\phi'(x(u))},$$

which has associated density $\phi''(x(z))x'(z)/\phi'(x(u))$. Therefore, the indifference condition required for optimality is

$$(6) \quad r(u) + \alpha u = \frac{1}{\phi'(x(u))} \int_0^u [V(r(z), z) + \alpha z] \phi''(x(z)) x'(z) dz$$

for all $u \in [0, u_{\max}]$.

In particular, we see from (6) that $r(0) = 0$. This implies that the principal does not wish to restrict the desirable projects available to the agent whose best project has only zero payoff for her, that is, there is "no distortion at the bottom." The reason for this is that when $u = 0$, there is no strategic benefit to

restricting choice. (The strategic effect of raising $r(u)$ above $-\alpha u$ is to increase the probability that the agent will choose a smaller z , and this effect cannot operate when $u = 0$.) Differentiating (6) with respect to u and using (4) implies the Euler equation for the principal's problem, which is expression (7) below.

PROPOSITION 1: *The principal's problem of maximizing (3) subject to (4) and (5) over all piecewise-continuous threshold functions $r(\cdot)$ has a solution. This solution is differentiable and satisfies the Euler equation*

$$(7) \quad r'(u) + \alpha = [V(r(u), u) - r(u)][1 - G(r(u), u)]f(u) \frac{\phi''(x(u))}{\phi'(x(u))}$$

with initial condition $r(0) = 0$. A sufficient condition for a threshold function which satisfies the Euler equation to be a global optimum is that

$$(8) \quad \zeta(x) \equiv \frac{\phi''(x)}{\phi'(x)} \quad \text{weakly decreases with } x.$$

The proofs of this and of subsequent propositions are given in the [Appendix](#). Expression (7) reveals that ζ in (8) is important for the form of the solution. A short list of examples for this term follows:

- If n is known to be exactly $N \geq 1$ for sure (so $q_N = 1$), then $\phi(x) = x^N$ and $\zeta(x) = (N - 1)/x$.
- If n is Poisson with mean μ (so $q_n = e^{-\mu} \mu^n / n!$ for $n \geq 0$), then $\phi(x) = e^{-\mu(1-x)}$ and $\zeta(x) \equiv \mu$.
- If n is binomial (the sum of N Bernoulli variables with success probability a), then $\phi(x) = (1 - a(1 - x))^N$ and $\zeta(x) = a(N - 1)/(1 - a(1 - x))$. The “known- n ” case is a special case of the binomial with $a = 1$. The Poisson is a limit case of the binomial when $aN = \mu$ and $a \rightarrow 0$.
- If n is geometric (so $q_n = (1 - a)a^{n-1}$ for $n \geq 1$ and some parameter $a \in (0, 1)$), then $\phi(x) = (1 - a)x/(1 - ax)$ and $\zeta(x) = 2a/(1 - ax)$.

Assumption (8), which states that $\phi'(x)$ is a log-concave function, is valid for the binomial distribution—and hence for the known- n and Poisson subcases—but not for the geometric distribution.

Define the “naive” threshold rule to be $r_{\text{naive}}(u) = -\alpha u$. This is the threshold rule which permits all desirable projects, that is, those projects such that $v + \alpha u \geq 0$. This rule might be implemented by a principal who ignored the strategic effect that the agent will choose the project with the highest u whenever she has a choice. As such, the naive rule is optimal when the agent never has a choice of project, that is, when $q_0 + q_1 = 1$. (In this case, $\phi'' \equiv 0$, the right-hand side of (7) vanishes, and so $r(\cdot) \equiv r_{\text{naive}}(\cdot)$ is optimal.) Outside this case, though, the right-hand side of (7) is strictly positive. Since $r'(u) + \alpha > 0$ and $r(0) = 0$, it follows that $r(u) > r_{\text{naive}}(u)$ when $u > 0$. Therefore, the prin-

cipal forbids some strictly desirable projects (and never permits an undesirable project). Moreover, the gap between the optimal and the naive rule, $r(u) - r_{\text{naive}}(u)$, strictly increases. We state this formally:

COROLLARY 1: *Suppose the agent sometimes has a choice of project (i.e., $q_0 + q_1 < 1$). Then it is optimal for the principal to forbid some strictly desirable projects, and the gap between the optimal threshold rule $r(u)$ and the naive threshold rule $r_{\text{naive}}(u)$ widens with u . In particular, when $\alpha = 0$, the optimal threshold rule increases with u .*

What is the intuition for why the principal wishes to exclude some desirable projects from the permitted set, whenever the agent sometimes has a choice of project? Suppose the principal initially allows all desirable projects, so that $r(u) \equiv r_{\text{naive}}(u)$. If the principal increases $r(\cdot)$ slightly at some $u > 0$, the direct cost is approximately zero, since the principal then excludes projects about which he is almost indifferent (since $r(u) + \alpha u = 0$). But there is a strictly beneficial strategic effect: there is some chance that the agent's highest- u project is excluded by the modified permitted set, in which case there is a chance that she chooses another project which is permitted, say with $z < u$. This alternative project is unlikely to be marginal for the principal, and the principal will expect to get payoff $V(r(z), z) + \alpha z$, which is strictly positive when $r(z) = -\alpha z$. This argument indicates that it is beneficial to restrict desirable projects and not to permit any undesirable projects. Moreover, it is intuitive that the strategic effect is more important for higher u , since it applies over a wider range $z < u$, and this explains why $r(u) - r_{\text{naive}}(u)$ increases with u .

We next discuss some comparative statics for this problem.

PROPOSITION 2: *Let α_L and α_H be two weights placed by the principal on the agent's payoff, where $\alpha_L < \alpha_H$. Let $r_i(\cdot)$ and $x_i(\cdot)$ solve the Euler equation (7) when $\alpha = \alpha_i$ for $i = L, H$. If assumption (8) holds, then $x_L(0) \geq x_H(0)$, that is, the fraction of permitted projects increases with α .*

Thus we see that the more the principal cares about the utility of the agent, the more discretion—in the sense of a greater fraction of projects being permitted—the agent is given.¹¹ This result is similar to the “ally principle” in the Holmstrom-type models mentioned in Section 1, where the more likely the agent's preferences were to be close to the principal's, the more discretion the agent was given.

A second way in which the ally principle might be expected to be seen concerns the extent of correlation between u and v . Intuitively, when u is positively

¹¹It is not necessarily the case that the threshold rules are *nested* so that $r_H(\cdot) \leq r_L(\cdot)$; one can find examples where the two threshold rules cross for some positive u .

correlated with v , the agent's incentives are likely to be aligned with those of the principal. In the limit of perfect positive correlation, since the agent's best project is always the principal's best project, it is optimal to give the agent complete freedom to choose a project. (By contrast, with strong negative correlation, the agent's best permitted project is likely to be the principal's worst permitted project, at least when α is small.) However, it is not obvious how formally to define a notion of "more correlation" which could be used as a basis for general comparative statics analysis. Instead, in Section 2.2 we discuss an example which confirms this intuition.

It is also intuitive that when the agent is likely to have more projects to choose from, the principal will further constrain the permitted set of projects. With more projects available, the agent is likely to have at least one which lies close to the principal's preferred project. A notion of "more projects" which ensures that this intuition is valid is the familiar *monotone likelihood ratio property* (MLRP). The details are provided in the next result:

PROPOSITION 3: *Let $(q_0^L, q_1^L, q_2^L, \dots)$ and $(q_0^H, q_1^H, q_2^H, \dots)$ describe two probability distributions for the number of projects which satisfy MLRP, that is, q_n^H/q_n^L weakly increases with n . Let $\phi_i(\cdot)$ be the PGF corresponding to $(q_0^i, q_1^i, q_2^i, \dots)$, and suppose $\phi_L(x)$ satisfies (8). Let $r_i(\cdot)$ and $x_i(\cdot)$ solve the Euler equation (7) when the number of projects is governed by $(q_0^i, q_1^i, q_2^i, \dots)$ for $i = L, H$. Then $x_H(0) \geq x_L(0)$.*

Thus, the fraction of permitted projects falls when the agent is likely to have more projects available.¹² The requirement that the number of projects be ordered by MLRP is a stronger requirement than first-order stochastic dominance. Indeed, there are examples where stochastic dominance leads to a *smaller* fraction of projects being excluded.¹³ Moreover, it is not necessarily the case that the principal benefits when the agent has access to more projects. When there is strong negative correlation between u and v , an agent choosing

¹²As emphasized in Lyons (2002), in the merger application it is more likely that the more stringent consumer standard is superior to a total welfare standard in large, complex economies where merger possibilities may be more plentiful.

¹³An example where adding more projects widens the optimal set of permitted projects is the following. Suppose initially the agent has no projects at all with probability $1 - \varepsilon$ and exactly two projects with probability ε . Because the state when no projects materialize plays no role in the determination of $r(\cdot)$, the optimal threshold rule for this agent is just as if there were two projects for sure. Such a threshold rule will strictly exclude some desirable projects. Consider next the situation in which the agent has exactly one more project than the previous situation (i.e., $n = 1$ with probability $1 - \varepsilon$ and $n = 3$ with probability ε). Whenever ε is small, the state where there is only one project will dominate the choice of $r(\cdot)$, and almost all desirable projects will be permitted, thus widening the set of permitted projects. One can check that this pair of probability distributions does not satisfy MLRP.

from more projects is likely, all else equal, to choose a worse project from the principal's perspective.¹⁴

Without making further assumptions, it is hard to make more progress in characterizing the solution to (7). In the remainder of Section 2, we examine further properties of the solution in three special cases.

2.1. Independent Payoffs and $\alpha = 0$

Suppose that the distribution of v is independent of u and that $\alpha = 0$. Then the principal does not care about the agent's choice of u , either directly (since $\alpha = 0$) or indirectly (since the distribution of his payoff v does not depend on u).

Write $G(v)$ and $V(r)$ as functions which do not depend on u . It follows that

$$\begin{aligned} & \frac{d}{du} \left[\frac{V(r(u)) - r(u)}{f(u)} \frac{d}{du} \phi(x(u)) \right] \\ &= \frac{d}{du} [(V(r(u)) - r(u))(1 - G(r(u)))\phi'(x(u))] \\ &= \frac{d}{du} \left[\left(\int_{r(u)}^{v_{\max}} (1 - G(v)) dv \right) \phi'(x(u)) \right] \\ &= -r'(u)(1 - G(r(u)))\phi'(x(u)) \\ &\quad + (V(r(u)) - r(u))(1 - G(r(u)))^2 f(u)\phi''(x(u)), \end{aligned}$$

which equals zero at the optimum from (7). Therefore, the second-order Euler equation reduces to the first-order equation

$$(9) \quad [V(r(u)) - r(u)] \frac{d}{du} \phi(x(u)) = kf(u)$$

for some positive constant k .

It follows that the principal obtains the same expected payoff with all density functions $f(\cdot)$ for u . To see this, change variables in (9) from u to $F(u)$. That is to say, write $\hat{r}(F(u)) \equiv r(u)$ and $\hat{x}(F(u)) \equiv x(u)$, so that \hat{r} represents the

¹⁴Our benchmark model assumes that each realization of project characteristics (u, v) is independent across projects. If, instead, project characteristics were positively correlated across projects, it is plausible that the effect of correlation would be similar to having fewer independent projects. (For instance, in the extreme case where all projects had the same realization of (u, v) , the situation is just as if the agent had a single project to choose from, in which case the naive rule is optimal.)

threshold rule expressed in terms of the cumulative *fraction* of u projects F . Then (9) becomes

$$(10) \quad [V(\hat{r}(F)) - \hat{r}(F)] \frac{d}{dF} \phi(\hat{x}(F)) \equiv k,$$

with endpoint conditions $\hat{r}(0) = 0$ and $\hat{x}(1) = 1$. Here, the optimal threshold rule $\hat{r}(\cdot)$ does not depend on the distribution for u , as long as u is continuously distributed.¹⁵ As such, only *ordinal* rankings of u matter for the principal in this case.

2.2. Exponential Distribution for v

Suppose next that v given u is exponentially distributed on $[0, \infty)$ with mean $\lambda(u)$, so that $G(v, u) = 1 - e^{-v/\lambda(u)}$. Suppose that $\alpha = 0$. Since $V(r, u) \equiv r + \lambda(u)$ in this example, the Euler equation (7) is

$$r'(u) = \lambda(u) \frac{d}{du} \log \phi'(x(u))$$

with initial condition $r(0) = 0$. Since we wish to compare policies across different distributions for (u, v) , the threshold rule r is not in itself insightful. Rather, we study the *fraction* of permitted type- u projects and, given r , write $p(u) = 1 - G(r(u), u) = e^{-r(u)/\lambda(u)}$ for this fraction. Writing the Euler equation in terms of p rather than r implies that

$$(11) \quad \frac{d}{du} \log p(u) = - \left[\frac{d}{du} \log \phi'(x(u)) + \frac{\lambda'(u)}{\lambda(u)} \log p(u) \right]$$

with initial condition $p(0) = 1$.

Consider first the case where λ is constant, so that u and v are independent. Expression (11) implies that $p\phi'(x)$ does not vary with u , and it follows that $\phi(x(u)) = k_1 F(u) + k_2$ for constants k_1 and k_2 . Since $\phi(x(u_{\max})) = 1$, it follows that $k_1 + k_2 = 1$. Since $p(0) = 1$, it follows that $k_1 = \phi'(x_0)$, where $x_0 = x(0)$ is the fraction of banned projects at the optimum when u and v are independent. In sum, at the optimum, $\phi(x(u)) = 1 - \phi'(x_0)(1 - F(u))$. Evaluating this at $u = 0$ implies that the fraction of banned projects is the unique solution to

$$(12) \quad \phi(x_0) + \phi'(x_0) = 1.$$

¹⁵Note that this argument requires us to change variables in expression (9), and so $F(u)$ needs to be differentiable and, in particular, the distribution for u has no “atoms.” If there were atoms, then we would need to consider what project the agent would choose in the event of a tie, when there were two projects which yielded the same maximal agent payoff u .

Next, suppose that u and v are positively correlated in the (strong) sense that $\lambda(u)$ increases with u . Write $h(u) \equiv p(u)\phi'(x(u))$, which from (11) is an increasing function. It follows that

$$\begin{aligned}\phi(x(u)) &= 1 - \int_u^{u_{\max}} h(\tilde{u})f(\tilde{u})d\tilde{u} \\ &= 1 - h(0)(1 - F(u)) - \int_u^{u_{\max}} [h(\tilde{u}) - h(0)]f(\tilde{u})d\tilde{u}.\end{aligned}$$

Since $h(0) = \phi'(\tilde{x}_0)$, where \tilde{x}_0 denotes the fraction of banned projects in this case with positive correlation, and h is increasing, it follows that $\phi(\tilde{x}_0) + \phi'(\tilde{x}_0) < 1$. Since $\phi(\cdot) + \phi'(\cdot)$ is an increasing function, it follows that the fraction of permitted projects is higher with positive correlation than with independence. A parallel argument establishes that when there is negative correlation, in the sense that λ decreases with u , the fraction of permitted projects is smaller than with independence. In this exponential example, then, positive correlation between u and v is associated with a greater number of permitted projects than negative correlation.

2.3. Poisson Distribution for the Number of Projects

As our third special case, suppose that the number of projects follows a Poisson distribution with mean μ , in which case the Euler equation (7) reduces to a first-order differential equation in $r(u)$:

$$(13) \quad r'(u) + \alpha = \mu[V(r(u), u) - r(u)][1 - G(r(u), u)]f(u), \quad r(0) = 0.$$

The next result shows that the comparative statics of $r(\cdot)$ with respect to α and μ are stronger than the corresponding results in the general setting reported above in Propositions 2 and 3.

PROPOSITION 4: *With a Poisson distribution for the number of available projects, the optimal threshold rule $r(\cdot)$ is pointwise decreasing in α and increasing in μ .*

To obtain some explicit solutions for the threshold rule, suppose that (u, v) is uniformly distributed on the rectangle $[0, 1] \times [-1, 1]$. In this case, (13) becomes the homogeneous equation

$$(14) \quad r'(u) = \frac{1}{4}\mu(1 - r(u))^2 - \alpha, \quad r(0) = 0.$$

Note that if $\mu = 4\alpha$, then the solution to (14) is simply the flat rule $r(u) \equiv 0$. Thus, in the merger context, if the regulator wishes to maximize total welfare

(so $\alpha = 1$), then when the expected number of merger possibilities is $\mu = 4$, the regulator should optimally enforce a consumer welfare standard.

The solution to (14) when $\mu \neq 4\alpha$ is given implicitly by

$$(15) \quad \int_0^{r(u)} \frac{1}{\frac{1}{4}\mu(1-r)^2 - \alpha} dr = u.$$

When $\alpha = 0$, expression (15) yields the simple formula

$$(16) \quad r(u) = \frac{\mu u}{4 + \mu u}.$$

When $\alpha > 0$, expression (15) can be integrated using partial fractions to give

$$(17) \quad r(u) = \left(1 - \frac{4\alpha}{\mu}\right) \frac{e^{u\sqrt{\alpha\mu}} - 1}{(1 + \sqrt{4\alpha/\mu})e^{u\sqrt{\alpha\mu}} - (1 - \sqrt{4\alpha/\mu})}.$$

Figure 3 plots the rule (17) for $\alpha = 1$ and various μ . Here, higher curves correspond to higher μ as in Proposition 4. The straight line depicted for $\mu = 0$ is just the naive rule which permits any desirable project.

A final observation about the Poisson distribution concerns the principal's expected payoff, which from (6) evaluated at u_{\max} is equal to $r(u_{\max}) + \alpha u_{\max}$. (Recall that the density of the agent's choice of u is $\frac{d}{du}\phi(x(u))$ and that the Poisson case entails $\phi''(x) \equiv \mu\phi'(x)$.) For instance, in the uniform example with $\alpha = 0$, where the threshold rule is given by (16), it follows that the principal's maximum expected payoff is $r(1) = \mu/(4 + \mu)$.

3. VARIANTS OF THE BENCHMARK MODEL

3.1. Incentives to Find a Project

The benchmark model in Section 2 assumed that the number of projects was exogenous to the agent. In such a framework the agent does not need to be given an incentive to discover projects. In this variant, we suppose that the agent needs to exert effort to find a project. We do this in the simplest possible way, so that by exerting effort e , the agent finds a single project with probability e , while with remaining probability $1 - e$, no project emerges.¹⁶ If she

¹⁶A richer model would involve the agent being able to affect the expected number of projects, so that the agent may end up with a choice of project. (For instance, if the number of projects follows a Poisson distribution, the agent could choose μ by incurring cost $C(\mu)$, say.) The principal's optimal policy in this situation has some similarities to the policy when the number of projects was exogenous: the threshold rule is nonlinear and involves $r(0) = 0$. However, like the model of costly discovery analyzed in this section, the threshold rule reflects the need to give the agent an incentive to find more projects (which typically benefits the principal as well as the agent), and it may be optimal ex ante to permit projects which are undesirable ex post.

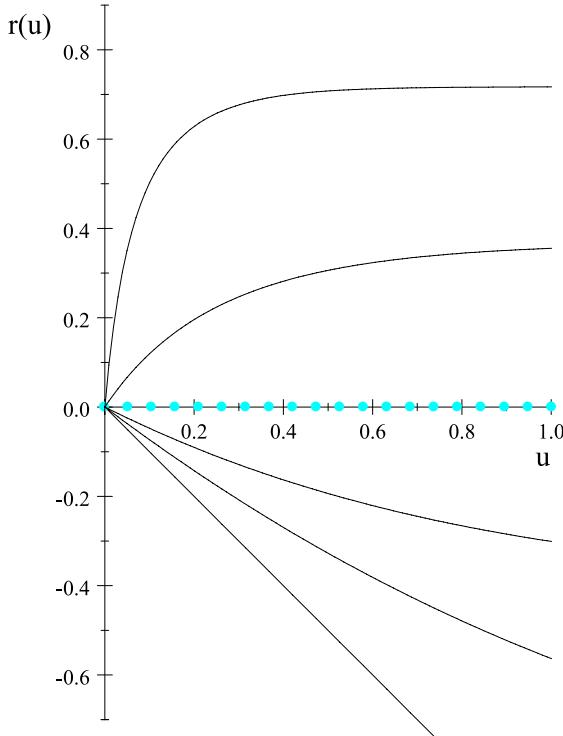


FIGURE 3.—Uniform-Poisson case with $\alpha = 1$ and $\mu = 0, 1, 2, 4$ (dotted), 10, and 50.

finds a project, that project's characteristics (u, v) are realized according to the same density functions f and g as in the benchmark model. To achieve success probability e , the agent incurs the private cost $c(e)$. Here, $c(\cdot)$ is assumed to be convex, with $c(0) = c'(0) = 0$ and $c'(1) = \infty$.

Since the agent's effort incentives depend on her *expected* payoff across all permitted projects, her attitude toward risk is relevant, and in this section we assume the agent is risk-neutral. The principal's payoff is a weighted sum of the agent's payoff (including her cost of effort) and the expected value of v , where the weight on the agent's payoff by the principal is α . The principal determines a piecewise-continuous function $r(\cdot)$ such that any (u, v) project with $v \geq r(u)$ is permitted.

If she discovers a project, the agent's expected payoff excluding effort cost is

$$(18) \quad A = \int_0^{u_{\max}} u[1 - G(r(u), u)]f(u) du,$$

and the agent will choose effort e to maximize her net payoff $eA - c(e)$. Clearly, a reduction in the threshold rule $r(\cdot)$ induces a higher value of A

in (18), which in turn leads unambiguously to greater effort from the agent.¹⁷ Since high effort benefits the principal as well as the agent, the principal has a reason (beyond the weight α placed on the agent's interests) to increase the leeway given to the agent.

If we write

$$\sigma(A) \equiv \max_{e \geq 0} : eA - c(e)$$

for the agent's maximum payoff given A , then σ is a convex increasing function and $\sigma'(A)$ is the agent's choice of effort e given her reward A . The principal chooses $r(\cdot)$ to maximize his expected payoff

$$(19) \quad \alpha\sigma(A) + B\sigma'(A),$$

where

$$B = \int_0^{u_{\max}} V(r(u), u)[1 - G(r(u), u)]f(u) du.$$

The principal's optimal policy is described in the next result:

PROPOSITION 5: *The principal's optimal policy takes the form*

$$(20) \quad r(u) + \left[B \frac{\sigma''(A)}{\sigma'(A)} + \alpha \right] u \equiv 0.$$

Thus, the optimal threshold rule is a ray emanating from the origin. This ray is downward sloping and weakly steeper than the principal's naive rule, $r_{\text{naive}}(u) \equiv -\alpha u$. The only situation in which the principal implements his naive rule is when $\sigma'' = 0$, which applies when the agent's success probability does not respond to incentives, that is, there is an *exogenous* success probability e . Outside this case, though, the principal allows some projects which are strictly undesirable ($v + \alpha u < 0$) so as to stimulate the agent's effort.¹⁸ The more that the agent responds to incentives (in the sense that the function $\sigma''(\cdot)/\sigma'(\cdot)$ is shifted upward), the more leeway she should have to choose a project. This distortion is the opposite to the bias in the benchmark model in Section 2, where the principal forbade some desirable projects.

Some intuition for the linearity of $r(u)$ comes from the following argument. The principal's payoff in (19) is a function of both A (the expected value of v

¹⁷This is akin to the “initiative effect” of delegation in Aghion and Tirole (1997).

¹⁸This feature is also seen in Aghion and Tirole (1997) and Baker, Gibbons, and Murphy (1999). By contrast, as discussed in Section 1, Szalay (2005) presented a model where information gathering incentives are enhanced by forbidding projects which both the principal and agent might often wish to implement.

from a single project given that the project is only implemented if it is permitted) and B (the expected value of u from a project given that the project is only implemented if it is permitted), and A and B are, in turn, functions of the principal's permission rule $r(\cdot)$. The problem of choosing $r(\cdot)$ to maximize a (nonlinear) function of A and B has the same first-order condition as maximizing a linear sum $A + \gamma B$ for some constant γ . That is to say, the solution to the principal's problem is obtained by choosing $r(\cdot)$ to maximize

$$\int_0^{u_{\max}} \int_{r(u)}^{v_{\max}} [v + \gamma u] g(v, u) f(u) dv du$$

for some constant γ , the solution to which is clearly to set $r(u) = -\gamma u$ so that only the positive $[v + \gamma u]$ are contained in the integral.

In earlier work we analyzed a more complicated version of this problem in which the agent searches sequentially for a satisfactory project and can influence the arrival rate of new projects by incurring effort. Then the agent might not implement the first permitted project which emerges, but rather wait until she finds a permitted project which achieves a reservation utility, where this reservation utility will depend on the threshold rule $r(\cdot)$ as well as her discount rate. A linear threshold rule is again optimal for the principal, although not necessarily a rule which starts at the origin. When the principal and agent are more impatient, the threshold rule is shifted downward, so that the principal is willing to accept a less good project and with less delay. In the limit of extreme impatience, the dynamic search problem essentially reduces to the framework discussed in this section where the agent tries to discover a single project.

3.2. Paying for a Good Project

Most of our analysis presumes that monetary incentives to choose a desirable project are not available or desirable, for reasons outside the model. In this second variant, we briefly discuss the principal's optimal policy when he can condition the agent's payment on her performance. We will see that, even within the confines of the model, monetary incentives are not always desirable.

First, suppose that the agent is risk-neutral and is able to bear large losses ex post. As in most principal–agent models, the principal here is able to attain his first-best outcome with the use of monetary incentives. The first-best outcome is obtained when (i) he does not restrict the agent's choice of project, (ii) he pays the agent v when a type- (u, v) project is implemented (and allows the agent to keep her benefit u), and (iii) he extracts the agent's entire expected surplus from this scheme in the form of a payment to the principal up front. Such a scheme is akin to "selling the firm" to the agent, and gives the agent ideal incentives to choose the best available project while leaving the agent with zero expected rent.

Outside this extreme case, however, the first-best solution will not be attainable, and there may again be a role for restricting the agent's discretion.

Moreover, the use of monetary incentives will not always be optimal for the principal.¹⁹ To illustrate most simply, consider the situation in which the agent is liquidity constrained in the sense that she must receive a nonnegative salary (excluding her payoff u from an implemented project) in all outcomes.²⁰ For simplicity, suppose that there are two possible kinds of projects, one of which is preferred by the agent while the other is preferred by the principal. Specifically, the “bad project” has payoffs (u_H, v_L) and the “good project” has payoffs (u_L, v_H) , where $0 < u_L < u_H$ and $0 < v_L < v_H$.²¹ Write $\Delta_u = u_H - u_L$ and $\Delta_v = v_H - v_L$, and suppose $\Delta_v > \Delta_u$ so that (u_L, v_H) is indeed socially the good project. The difference Δ_u can be interpreted as the “bias” of the agent. Since there are just two types of project, what matters is the probability that the agent has only the good project available (denoted P_G), the probability she has only the bad project (denoted P_B), and the probability she has a choice of project types (denoted P_{GB}).

If the principal bans the bad project, his payoff is

$$\pi_1 = (P_G + P_{GB})(v_H + \alpha u_L).$$

If the principal allows both projects but does not use monetary rewards, his payoff is

$$\pi_2 = P_G(v_H + \alpha u_L) + (P_{GB} + P_B)(v_L + \alpha u_H).$$

(Here, the agent will choose the bad project whenever that project is available.) The remaining policy is to give the agent a monetary incentive equal to Δ_u to choose the good project (and not to fetter her discretion), which entails payoff

$$\pi_3 = (P_G + P_{GB})(v_H - \Delta_u + \alpha u_H) + P_B(v_L + \alpha u_H).$$

(Here, the agent will implement the good project whenever such a project is available.)

We require $\pi_3 \geq \max\{\pi_1, \pi_2\}$ for monetary incentives to be optimal. Now

$$\pi_3 - \pi_1 > 0 \Leftrightarrow P_B(v_L + \alpha u_H) > (P_G + P_{GB})(1 - \alpha)\Delta_u$$

and

$$\pi_3 - \pi_2 > 0 \Leftrightarrow P_{GB}(\Delta_v - \Delta_u) > P_G(1 - \alpha)\Delta_u.$$

¹⁹For instance, if there are very many projects available to the agent, the first-best outcome is approximately achieved by permitting the agent to choose only the best projects for the principal and making no monetary payments to the agent.

²⁰A similar restriction to nonnegative payments is made in Aghion and Tirole (1997), Berkovitch and Israel (2004), and Alonso and Matouschek (2008, Section 8.1).

²¹If $u_L < 0$, then without monetary compensation the agent will not reveal the good project, even when that is the only option available to her. In this case, the use of money rewards enables the good project to be implemented when available, which is an important benefit of using money rewards relative to restricting the agent’s choice.

These inequalities are jointly satisfied when the agent's bias Δ_u is sufficiently small (i.e., when little money needs to be paid to change agent behavior) or when α is close to 1 (so that payments to the agent are not costly for the principal). By contrast, monetary incentives should not be used when the agent's bias is large or $v_L + \alpha u_H$ is small. Finally, note that increasing the number of project draws will make it more likely the agent has a choice of project types, so that P_{GB} rises, and this makes it less likely that $\pi_3 > \pi_1$.²² Thus, all else equal, we expect that a greater number of project opportunities, or a larger agent bias, will make the use of money rewards less attractive.²³

Berkovitch and Israel (2004) analyzed a related model, also with binary project types. Provided the manager's bias is not too large, they showed [Proposition 1(a)] that (i) if the "good" (i.e., less capital intensive) project is relatively likely to be available, then the optimal policy is (stochastically) to ban the bad project rather than to reward the manager when she brings forward a good project, and (ii) if a good project is less likely to emerge, it becomes optimal to permit both projects but to pay the agent for a good project.

This example illustrates a more general trade-off between banning mediocre projects and rewarding the choice of good ones. When he bans mediocre projects, the principal suffers the cost that such projects are not implemented when they are the only ones available. Rewarding the choice of good projects avoids this cost, but instead involves paying the reward whenever at least one good project is available. Restricting choice is therefore preferred when the chance of having only mediocre projects is small, which is more likely to be true when the agent can choose from many projects.²⁴ In richer settings than the illustrative binary example above, it may be optimal both to ban mediocre projects and reward the choice of good projects. In addition, if the agent is not liquidity constrained, it is possible to financially penalize her choice of bad projects, which could well be preferable to an outright ban. We leave a more complete analysis of the interactions between restricting choice and monetary incentives as a topic for further work.

3.3. A More Complex Delegation Scheme

Our benchmark scheme simply involves specifying a set of permitted projects and the agent choosing her preferred available project in this set. In particular, only the agent's chosen project is subject to verification. If, however, the

²²For instance, if there are N project opportunities for sure and each opportunity has probability P of yielding a bad project, then $P_B = P^N$, $P_G = (1 - P)^N$, and $P_{GB} = 1 - P^N - (1 - P)^N$.

²³See Figure 9 in **Alonso and Matouschek (2008)** for an illustration in their framework of the limited gains to the principal in being able to make contingent payments to the agent rather than restricting discretion.

²⁴Another reason why monetary incentives are not always given to an agent is that the agent performs several tasks, and giving incentives to do one task well might induce the agent to underperform on other, unmeasured, aspects of her job (see **Holmstrom and Milgrom (1991)**).

principal could easily determine the genuine feasibility of all reported projects (so there are no significant costs of auditing reported but unchosen projects), the principal may be able to do better by inducing the agent to list more than one project. Since the listed projects have characteristics which can be verified by the principal, the agent can only reveal true projects. But, except in the implausible case in which the number of available projects is known in advance (the known- n case), the agent need only report those projects she wishes to report and she cannot be made to reveal the “whole truth.”²⁵ In such cases, the most general (deterministic) delegation scheme takes the following form: if the agent reveals a list of feasible projects, the principal picks (in a pre-determined way) one of these projects or implements no project.²⁶

For simplicity, we analyze this issue in the context of the binary project types discussed in the previous section. In this context the most general (deterministic) delegation scheme involves the agent reporting a list of projects as summarized by the pair of integers (b, g) , where b is her reported number of bad projects and g is her reported number of good projects. The constraint that the agent must tell the truth, but not necessarily the whole truth, is captured by the requirement that the agent’s reports satisfy $b \leq B$ and $g \leq G$, where B and G are the actual numbers of bad and good projects. A delegation scheme in general is a choice function which maps a report (b, g) into a decision to implement (i) either a good project (provided $g \geq 1$), (ii) a bad project (provided $b \geq 1$), or (iii) no project at all. However, the principal need only consider a particular family of delegation schemes:

LEMMA 2: *The principal can restrict attention to the following family of delegation schemes: if the agent reports she has at least one good project, that project is implemented; if the agent reports she has b bad projects (and no good projects), a bad project is implemented provided that $b \geq m$ for some (possibly infinite) integer m .*

PROOF: Suppose, contrary to the claim, that for some report (b, g) with $g \geq 1$ the principal either implements a bad project (if $b \geq 1$) or nothing at all.

²⁵This information structure is akin to games of persuasion. For instance, consider the signalling model of [Shin \(2003\)](#) in which a number of projects are undertaken by a firm, the sum of whose outcomes determines the returns to shareholders. (Therefore, unlike our framework, the firm’s manager does not choose *which* project to pursue.) But the manager has interim information about the outcome of a random subset of the projects, and she can reveal a subset of those project outcomes she knows. Thus, the manager can only conceal poor outcomes, not make up good ones. One plausible equilibrium is where the manager reveals all the good news and conceals all the bad news.

²⁶An example of such a scheme concerns work-related travel plans. An employee may be able to get permission for an expensive flight more easily if she reveals a number of other expensive quotes than if she provides just one. As we discuss further below, such a scheme may give the principal some assurance that the employee is not concealing a cheap but personally inconvenient flight option.

If instead the principal modifies his policy so that with this report (b, g) he now chooses to implement a good project, then this modification weakly increases the principal's payoff. (If the modification does not alter the agent's report, it clearly boosts the principal's payoff. If it does induce the agent to change her report, the only way the principal's payoff could be lowered is if the original policy with report (b, g) was to implement a bad project, and now the agent switches to a report which causes no project to be implemented. However, it is clear that it cannot be in the agent's interest to switch to such a report, since she could obtain $u_L > 0$ from the new policy by making her original report.) Therefore, it remains to describe the principal's policy when the agent reports only a number of bad projects. But since the agent can always reduce the number of bad projects she reports, all that matters is the *minimum* number of bad projects reported, say m , which ensures that a bad project is implemented. This completes the proof. *Q.E.D.*

Thus, the principal need consider only the single number m , the minimal number of bad realizations needed to authorize a bad project, when choosing his preferred scheme.²⁷ When $m = 1$, the agent can implement any project she chooses and when $m = \infty$, there is a blanket ban on bad projects, and these are the two possible policies in the benchmark model when only the implemented project's characteristics could be verified. Note that if the maximum possible number of projects is $N < \infty$, a delegation scheme with $m = N$ is surely superior to a blanket ban on bad projects, since it allows a desirable project to be implemented when there are N projects and all are bad, and it achieves the same outcome otherwise. (Indeed, if the number of projects is known to be N for sure, setting $m = N$ yields the first-best outcome for the principal.)

We next calculate the optimal choice for m . Suppose a given project has probability P of being bad, and the PGF for the number of project draws is $\phi(\cdot)$. Then the number of bad projects has PGF $\phi(1 - P + Px)$.²⁸ Therefore, the probability there are exactly n bad projects is

$$\frac{P^n}{n!} \phi^{[n]}(1 - P),$$

²⁷Green and Laffont (1986) analyzed in general terms a principal–agent model where the agent's feasible reports depend on her private information. Our setup in which the agent can conceal but not fabricate projects is a special case of their model, and one which satisfies their “nested range condition.” Therefore, we can invoke their Proposition 1 to deduce that the principal can restrict attention to choice rules which induce the agent to report *all* her projects (in contrast to the rules in Lemma 2).

²⁸In general, if m is a discrete random variable with PGF ϕ_M , then the random variable generated by the sum of n independent realizations of m , where n is itself a discrete random variable with PGF ϕ_N , has PGF $\phi_N(\phi_X(\cdot))$.

where $\phi^{[n]}$ is the n th derivative of ϕ .²⁹ The probability there are exactly n bad projects and no good projects is $q_n P^n = (P^n/n!) \phi^{[n]}(0)$, and so the probability there are exactly n bad projects and at least one good project is $(P^n/n!)[\phi^{[n]}(1 - P) - \phi^{[n]}(0)]$. It follows that the principal's expected payoff with threshold $m \geq 1$, denoted W_m , is

$$\begin{aligned} W_m &= (v_H + \alpha u_L) \sum_{n=0}^{m-1} \frac{P^n}{n!} [\phi^{[n]}(1 - P) - \phi^{[n]}(0)] \\ &\quad + (v_L + \alpha u_H) \sum_{n=m}^{\infty} \frac{P^n}{n!} \phi^{[n]}(1 - P). \end{aligned}$$

To understand this expression, note that when the agent has fewer than m bad projects and at least one good project, she has no choice but to implement a good project, which accounts for the first sum above, whereas if she has at least m bad projects she will implement a bad project, yielding the second sum above.

Note that $W_{m+1} - W_m$ has the sign of

$$(21) \quad \frac{\Delta_v - \alpha \Delta_u}{v_H + \alpha u_L} - \frac{\phi^{[m]}(0)}{\phi^{[m]}(1 - P)},$$

and so W_m is single-peaked in m provided that $\phi^{[m]}(0)/\phi^{[m]}(1 - P)$ weakly increases with m , which in turn holds whenever $\phi^{[m+1]}(x)/\phi^{[m]}(x)$ weakly decreases with x for each $m \geq 1$. (This condition is a stronger version of assumption (8) used in the benchmark model.) When W_m is single-peaked, the principal's optimal policy is to choose the smallest m such that (21) is negative. To illustrate, consider the case where the number of projects follows a binomial distribution, that is, the sum of N Bernoulli variables with success probability a , so that $\phi(x) = (1 - a(1 - x))^N$. Then for $m \leq N$, expression (21) becomes

$$\frac{\Delta_v - \alpha \Delta_u}{v_H + \alpha u_L} - \left(\frac{1 - a}{1 - aP} \right)^{N-m},$$

which is indeed decreasing in m , and so the optimal m is the smallest m which makes the expression negative.³⁰ Notice that the optimal scheme is more permissive—in the sense that m is smaller—when α is larger and when N is smaller, which parallels the comparative statics for the benchmark model in Propositions 2 and 3.

²⁹Recall that for an arbitrary PGF $\psi(x)$, the probability of having realization n is equal to the coefficient of x^n in ψ , that is, is equal to $\psi^{[n]}(0)/n!$.

³⁰For instance, if $a = P = (\Delta_v - \alpha \Delta_u)/(v_H + \alpha u_L) = 1/2$, then $m = N - 1$ is optimal.

What about cases with an unbounded number of potential projects? When the number of projects comes from a Poisson distribution with mean μ , expression (21) becomes

$$\frac{\Delta_v - \alpha\Delta_u}{v_H + \alpha u_L} = e^{-\mu(1-P)},$$

which is independent of m . Thus, W_m is monotonically increasing in m if the above expression is positive, in which case a blanket ban on bad projects is optimal. Alternatively, W_m decreases with m if the above expression is negative, in which case the agent should have authority to implement any project she chooses. In either event, in the Poisson case the more complicated delegation schemes considered in this section cannot improve on a simple scheme in which the agent is just presented with a set of permitted projects.

The intuition for this result comes from noting that in the Poisson case the number of good projects and the number of bad projects are themselves *independent* Poisson random variables (with respective means $(1 - P)\mu$ and $P\mu$). Thus, even if the principal could costlessly observe the number of bad projects, the realized number of bad projects has no impact on his decision about whether or not to permit bad projects. Since the complicated schemes in this section are a *costly* way to gain information about the number of bad projects (since the agent then sometimes implements a bad project when she has a good project), any such scheme must strictly underperform relative to the best simple scheme. This has the additional implication that in the Poisson case the principal's optimal rule—say, to ban the bad project—is “renegotiation proof” in the following sense: even if the agent can credibly reveal a number of projects which are not permitted, the principal has no incentive to adjust his permission set.

4. CONCLUSIONS

Proceeding from the motivating example of welfare standards in merger policy, we have explored the nature of optimal discretion for a principal to give to an agent when the agent may have a choice of project. The principal's problem is to design the optimal set of permitted projects without knowing which projects are available to the agent, though being able to verify the characteristics of the project chosen by the agent. In other words, the problem is to set the optimal rule that the agent must obey, in circumstances where the principal can just check whether or not the rule has been met.

In the benchmark model the agent has a number (unknown to the principal) of projects from which to choose. The optimal permission set excludes some projects that are desirable for the principal because the loss from excluding marginally desirable projects is outweighed by the expected gain from thereby inducing the choice of better projects. We showed (i) the principal

permits more types of project when he puts more weight on the agent's welfare and (ii) the principal permits fewer types of project when the agent has more projects from which to choose.

In one variant of this model, we supposed that by incurring a private cost, the agent makes it more likely that a project emerges, and the optimal permission set was characterized by a linear relationship between the payoffs of principal and agent. To encourage agent initiative, the principal permits some projects which are undesirable *ex post*, in contrast to the bias induced in the benchmark model. In a second variant, the principal was able to offer a monetary reward to the agent for choosing a good project, but with liquidity constraints on the agent, it might nevertheless be preferable to ban mediocre projects than to reward good ones: the former policy has costs when all available projects are mediocre, while the latter involves payments whenever there is at least one good project. In a final variant, we considered a situation in which the principal can verify a *list* of reported projects. In some cases, the principal does better by using a more complex delegation scheme which, for instance, is more permissive toward mediocre projects when the agent has several such projects. When the number of projects comes from a Poisson distribution, however, there is no gain in fine-tuning schemes in this manner.

It would be useful to examine more systematically than we do here the relative benefits of offering financial inducements (including penalties as well as rewards) to choose good projects versus banning mediocre projects. Another way to develop the analysis could be to multiagent settings: it is after all a feature of many rules that they apply without discrimination to various agents in various situations.

APPENDIX

PROOF OF PROPOSITION 1: The principal's aim is to maximize (3) subject to the endpoint condition (5) and the equation of motion (4). We consider the control variable $r(\cdot)$ to be taken from the set of piecewise continuous functions defined on $[0, u_{\max}]$ which take values in $[v_{\min}, v_{\max}]$, in which case $x(\cdot)$ is continuous and piecewise differentiable.

Although this is already a well posed optimal control problem, it is more convenient to consider $s(u) \equiv \phi(x(u))$, rather than $x(u)$, as the state variable. In this case, the equation of motion (4) becomes

$$(22) \quad s'(u) = (1 - G(r(u), u))f(u)\tau(s(u)),$$

where $\tau(\cdot)$ is the function derived from $\phi(\cdot)$ such that $\phi'(x) \equiv \tau(\phi(x))$ for all $0 \leq x \leq 1$. (That is to say, $\tau(s) \equiv \phi'(\phi^{-1}(s))$.) Note that τ is an increasing

function, and it is weakly concave in s if and only if $\phi''(x)/\phi'(x) \equiv \tau'(\phi(x))$ weakly decreases with x .³¹ In sum, we wish to maximize

$$(23) \quad \int_0^{u_{\max}} [V(r(u), u) + \alpha u] [1 - G(r(u), u)] f(u) \tau(s(u)) du$$

subject to the endpoint condition $s(u_{\max}) = 1$ and the equation of motion (22). We proceed in three stages: (i) we show that an optimal solution exists; (ii) we derive necessary conditions for the optimal policy; and (iii) subject to a regularity condition, we show that a policy satisfying the necessary conditions is a globally optimal policy.

First, that a solution to problem (23) exists can be deduced from the Filippov–Cesari theorem (for instance, see Seierstad and Sydsæter (1987, Chap. 2, Theorem 8)). The only nontrivial requirement for this theorem to be invoked is that the set

$$\begin{aligned} N(s, u) = & \{ ([V(r, u) + \alpha u][1 - G(r, u)]f(u)\tau(s) - \gamma, \\ & [1 - G(r, u)]f(u)\tau(s)) : \gamma \geq 0, v_{\min} \leq r \leq v_{\max} \} \end{aligned}$$

be convex for each s and u . Write $\eta(p, u) \equiv \int_{r(p, u)}^{v_{\max}} vg(v, u) dv$, where $r(p, u)$ is defined implicitly by $G(r(p, u), u) \equiv 1 - p$. Thus $r(p, u)$ is the threshold such that a proportion p of projects lie above $r(p, u)$ for given u , and $\eta(p, u)$ is the integral of v above this threshold. Therefore, $pV(r(p, u), u) = \eta(p, u)$. Note that η is concave in p , and that the above set N is equal to

$$\begin{aligned} N(s, u) = & \{ ([\eta(p, u) + \alpha up]f(u)\tau(s) - \gamma, pf(u)\tau(s)) : \\ & \gamma \geq 0, 0 \leq p \leq 1 \}, \end{aligned}$$

which is convex since $\eta(p, u) + \alpha up$ is a concave function of p . Therefore, an optimal strategy exists.³²

Second, we describe the necessary conditions which must be satisfied by the optimal policy. *Pontryagin's maximum principle* (see Seierstad and Sydsæter (1987, Chap. 2, Theorem 1)) states that if a piecewise-continuous control variable $r(\cdot)$ solves problem (23), then there exists a continuous and piecewise-differentiable function $\lambda(\cdot)$ such that $\lambda(0) = 0$ and for all $0 \leq u \leq u_{\max}$,

$$(24) \quad r(u) \text{ maximizes } (V(r, u) + \alpha u - \lambda(u))(1 - G(r, u))$$

³¹For instance, in the Poisson case $\tau(s) = \mu s$, and if there are two projects for sure, then $\tau(s) = 2\sqrt{s}$. In general, $\tau(1)$ is equal to the expected number of projects.

³²Strictly speaking, the Filippov–Cesari theorem shows the existence of an optimal *measurable* control $r(u)$ rather than a piecewise-continuous control. However, in practice this is not an important limitation. (See Seierstad and Sydsæter (1987, Chap. 2, footnote 9).)

over $v_{\min} \leq r \leq v_{\max}$, and, except at points where r is discontinuous,

$$(25) \quad \lambda'(s) = (V(r(u), u) + \alpha u - \lambda(u))(1 - G(r(u), u))f(u)\tau'(s).$$

Note that (24) implies

$$(26) \quad r(u) + \alpha u - \lambda(u) = 0,$$

and so $\lambda(u)$ represents the gap between the optimal rule $r(u)$ and the naive rule $r_{\text{naive}}(u) = -\alpha u$. Since λ is continuous, it follows that r is itself continuous. Moreover, since $r(\cdot)$ is continuous it follows from the maximum principle that $\lambda(\cdot)$ is everywhere differentiable, in which case (26) implies that $r(\cdot)$ is itself everywhere differentiable. Since $\lambda(0) = 0$, it follows that $r(0) = 0$. Combining (25) and (26) yields

$$r'(u) + \alpha = (V(r(u), u) - r(u))(1 - G(r(u), u))f(u)\tau'(s(u)),$$

which is equation (7) in the text.

Finally, we discuss when a policy satisfying these necessary conditions is a global optimum. The *Arrow sufficiency theorem* (see Seierstad and Sydsæter (1987, Chap. 2, Theorem 5)) shows that the necessary conditions pick out a global optimum if

$$[V(r(u), u) - r(u)][1 - G(r(u), u)]f(u)\tau(s)$$

is concave in s for all u . However, since $[V(r(u), u) - r(u)][1 - G(r(u), u)]f(u)$ is positive, the result follows if τ is concave in s . This is so if and only if (8) holds. *Q.E.D.*

PROOF OF PROPOSITION 2: Condition (7) implies that at $u = 0$ and any other u such that $r_L(u) = r_H(u)$, we have

$$(27) \quad \frac{r'_L(u) + \alpha_L}{r'_H(u) + \alpha_H} = \frac{\zeta(x_L(u))}{\zeta(x_H(u))}.$$

If $x_L(0) < x_H(0)$, then by assumption (8), $\zeta(x_L(0)) \geq \zeta(x_H(0))$ and so (27) implies that $r'_L(0) > r'_H(0)$. In particular, $r_L(u) > r_H(u)$ for small $u > 0$. If $x_L(0) < x_H(0)$, then $r_L(\cdot)$ must cross $r_H(\cdot)$ at some point. (If r_L were uniformly above r_H , then clearly the fraction of prohibited projects with α_L would be greater than with α_H .) Let u^* be the first point above zero where the curves cross. In particular, we have $r'_L(u^*) \leq r'_H(u^*)$. In addition, we must have $x_H(u^*) > x_L(u^*)$ since $x_H(0) > x_L(0)$ and $r_H(u) \leq r_L(u)$ for $u \leq u^*$. But then (27) implies that

$$1 > \frac{r'_L(u^*) + \alpha_L}{r'_H(u^*) + \alpha_H} = \frac{\zeta(x_L(u^*))}{\zeta(x_H(u^*))} \geq 1,$$

a contradiction. We deduce that the curves can never cross, and so our initial assumption that $x_L(0) < x_H(0)$ cannot hold. Q.E.D.

PROOF OF PROPOSITION 3: First note that if $q_i^L \equiv q_i^H$ so that the two distributions coincide, then the result clearly holds. So from now on suppose that $q_i^L \neq q_i^H$ sometimes. Let $\zeta_i(x) = \phi_i''(x)/\phi_i'(x)$. We first show that MLRP implies that ϕ_H'/ϕ_L' strictly increases with x , that is, $\zeta_H(\cdot) > \zeta_L(\cdot)$. The derivative of ϕ_H'/ϕ_L' has the sign

$$\begin{aligned} & \left(\sum_{n=2}^{\infty} n(n-1)q_n^H x^{n-2} \right) \left(\sum_{n=1}^{\infty} nq_n^L x^{n-1} \right) \\ & - \left(\sum_{n=2}^{\infty} n(n-1)q_n^L x^{n-2} \right) \left(\sum_{n=1}^{\infty} nq_n^H x^{n-1} \right). \end{aligned}$$

We claim that the coefficient on each power x^N , for $N \geq 0$, in the above is nonnegative. Defining $a_k \equiv (k+2)q_{k+2}^H$ and $b_k \equiv (k+2)q_{k+2}^L$, the coefficient on x^N can be written as

$$\begin{aligned} (28) \quad & \sum_{k=0}^N (k+1)(a_k b_{N-1-k} - a_{N-1-k} b_k) \\ & = (N+1)(a_N b_{-1} - a_{-1} b_N) + \sum_{k=0}^{N-1} (k+1)(a_k b_{N-1-k} - a_{N-1-k} b_k) \\ & = (N+1)(a_N b_{-1} - a_{-1} b_N) \\ & + \sum_{k=0}^M [(N-k) - (k+1)](a_{N-1-k} b_k - a_k b_{N-1-k}), \end{aligned}$$

where M is the largest integer no greater than $\frac{N-1}{2}$. The final expression pairs together terms in $(N-1-k)$ with terms in k . Since a_k/b_k is increasing in k by MLRP, $a_{N-1-k}/b_{N-1-k} \geq a_k/b_k$ for all $k \leq M \leq \frac{N-1}{2}$. So every term in (28) is nonnegative. However, $q_i^L \neq q_i^H$ sometimes, the coefficient on at least some x^N must be strictly positive. It follows that ϕ_H'/ϕ_L' strictly increases with x , and hence that $\zeta_H(\cdot) > \zeta_L(\cdot)$.

If $x_L(0) > x_H(0)$, then we have

$$\frac{r'_L(0) + \alpha}{r'_H(0) + \alpha} = \frac{\zeta_L(x_L(0))}{\zeta_H(x_H(0))} \leq \frac{\zeta_L(x_H(0))}{\zeta_H(x_H(0))} < 1.$$

Here, the equality follows from (7), the first inequality follows since we assume that (8) holds for ϕ_L , and the final inequality follows from $\zeta_H > \zeta_L$. We deduce

that $r'_L(0) < r'_H(0)$. The rest of the proof follows the same lines (with L and H permuted) as that for Proposition 2. Q.E.D.

PROOF OF PROPOSITION 4: Consider first the impact of increasing μ , and let μ_L and $\mu_H > \mu_L$ be two values for μ . Let $r_L(\cdot)$ and $r_H(\cdot)$ be the corresponding optimal threshold rules. From (13) it follows that at $u = 0$ and any other u such that $r_L(u) = r_H(u)$, we have

$$\frac{r'_L(u) + \alpha}{r'_H(u) + \alpha} = \frac{\mu_L}{\mu_H} < 1,$$

so $r'_L(u) < r'_H(u)$ at all such u . So r_H can never cross r_L from above. We deduce that $r_H(u) > r_L(u)$ for all $u > 0$. The argument for the impact of α on $r(\cdot)$ is similar. Q.E.D.

PROOF OF PROPOSITION 5: Let $r(\cdot)$ be the candidate optimal threshold rule, and consider the impact on the principal's payoff in (19) of a small variation $r(\cdot) + t\eta(\cdot)$, where $\eta(\cdot)$ is an arbitrary piecewise-continuous function. Writing the principal's payoff (19) in terms of t , denoted $W(t)$, yields

$$\begin{aligned} W(t) = & \alpha \int_0^{u_{\max}} u [1 - G(r(u) + t\eta(u), u)] f(u) du \\ & + \sigma' \left(\int_0^{u_{\max}} u [1 - G(r(u) + t\eta(u), u)] f(u) du \right) \\ & \times \left(\int_0^{u_{\max}} \left(\int_{r(u)+t\eta(u)}^{v_{\max}} v g(v, u) dv \right) f(u) du \right), \end{aligned}$$

and so

$$\begin{aligned} W'(0) = & -(B\sigma''(A) + \alpha) \left(\int_0^{u_{\max}} \eta(u) u g(r(u), u) f(u) du \right) \\ & - \sigma'(A) \left(\int_0^{u_{\max}} \eta(u) r(u) g(r(u), u) f(u) du \right). \end{aligned}$$

Since $W'(0)$ must equal zero for all $\eta(\cdot)$, it follows that $r(\cdot)$ must satisfy (20). Q.E.D.

REFERENCES

- AGHION, P., AND J. TIROLE (1997): "Formal and Real Authority in Organizations," *Journal of Political Economy*, 105, 1–29. [216,218,231,233]
 ALONSO, R., AND N. MATOUSCHEK (2007): "Relational Delegation," *RAND Journal of Economics*, 38, 1070–1089. [219]
 ——— (2008): "Optimal Delegation," *Review of Economic Studies*, 75, 259–294. [217,233,234]

- AMADOR, M., I. WERNING, AND G.-M. ANGELETOS (2006): "Commitment vs. Flexibility," *Econometrica*, 74, 365–396. [217]
- BAKER, G., R. GIBBONS, AND K. MURPHY (1999): "Informal Authority in Organizations," *Journal of Law, Economics and Organization*, 15, 56–73. [216,219,231]
- BERKOVITCH, E., AND R. ISRAEL (2004): "Why the NPV Criterion Does Not Maximize NPV," *Review of Financial Studies*, 17, 239–255. [215,233,234]
- FARRELL, J., AND M. KATZ (2006): "The Economics of Welfare Standards in Antitrust," *Competition Policy International*, 2, 3–28. [213,214]
- FRIDOLFSSON, S.-O. (2007): "A Consumer Surplus Defense in Merger Control," in *The Political Economy of Antitrust*, ed. by V. Ghosal and J. Stennek. Amsterdam: Elsevier, 287–302. [213]
- GREEN, J., AND J.-J. LAFFONT (1986): "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 53, 447–456. [236]
- HOLMSTROM, B. (1984): "On the Theory of Delegation," in *Bayesian Models in Economic Theory*, ed. by M. Boyer and R. Kihlstrom. Amsterdam: Elsevier, 115–141. [216]
- HOLMSTROM, B., AND P. MILGROM (1991): "Multi-Task Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, 7, 24–52. [234]
- LYONS, B. (2002): "Could Politicians Be More Right Than Economists? A Theory of Merger Standards," Mimeo, University of East Anglia. [213,225]
- MARTIMORT, D., AND A. SEMENOV (2006): "Continuity in Mechanism Design Without Transfers," *Economics Letters*, 93, 182–189. [217]
- MELUMAD, N., AND T. SHIBANO (1991): "Communication in Settings With No Transfers," *RAND Journal of Economics*, 22, 173–198. [217]
- SEIERSTAD, A., AND K. SYDSÆTER (1987): *Optimal Control Theory With Economic Applications*. Amsterdam: North-Holland. [240,241]
- SHIN, H. S. (2003): "Disclosures and Asset Returns," *Econometrica*, 71, 105–133. [235]
- SZALAY, D. (2005): "The Economics of Clear Advice and Extreme Options," *Review of Economic Studies*, 72, 1173–1198. [217,231]

Dept. of Economics, University College London, London WC1E 6BT, United Kingdom; mark.armstrong@ucl.ac.uk

and

All Souls College, University of Oxford, Oxford OX1 4AL, United Kingdom; john.vickers@economics.ox.ac.uk.

Manuscript received June, 2008; final revision received August, 2009.

INSIDER TRADING WITH A RANDOM DEADLINE

BY RENÉ CALDENTEY AND ENNIO STACCHETTI¹

We consider a model of strategic trading with asymmetric information of an asset whose value follows a Brownian motion. An insider continuously observes a signal that tracks the evolution of the asset's fundamental value. The value of the asset is publicly revealed at a random time. The equilibrium has two regimes separated by an endogenously determined time T . In $[0, T)$, the insider gradually transfers her information to the market. By time T , all her information has been transferred and the price agrees with the market value of the asset. In the interval $[T, \infty)$, the insider trades large volumes and reveals her information immediately, so market prices track the market value perfectly. Despite this market efficiency, the insider is able to collect strictly positive rents after T .

KEYWORDS: Insider trading, Kyle model, market microstructure, asset pricing.

1. INTRODUCTION

THIS PAPER STUDIES a model of strategic trading with asymmetric information of an asset whose value follows a Brownian motion. An insider receives a flow of (noisy) signals that tracks the evolution of the asset value. Other traders receive no signals and can only observe the total volume of trade. Since there is uncertainty about the value of the asset before the game starts, the first signal generates a lumpy informational asymmetry between the insider and the rest of the market participants. Subsequently, the insider receives a sequence of updates regarding the fundamental valuation of the asset. At an unpredictable time, a public announcement reveals the current value of the asset to all the traders. In equilibrium, the insider releases all her private information by a finite time T and keeps the market fully informed thereafter. Thus, she does not find it profitable to maintain informational asymmetry indefinitely.

Kyle (1985) introduced a dynamic model of insider trading² where an insider receives only one signal and the fundamental asset value does not change over time. Through trade, the insider progressively releases her private information to the market as she exploits her informational advantage. The market is also populated by many liquidity traders who are uninformed and trade randomly. At time 0, the insider observes the value of an asset. The same information is publicly released later, at time 1, to all market participants. In each trading period in the time interval $[0, 1]$, traders submit order quantities to a risk-neutral market maker who sets prices competitively and trades in his own account to clear the market. The market maker cannot observe individual trades, but can

¹We gratefully acknowledge the feedback of David Pearce. We also thanks Markus Brunnermeier, Lasse Pedersen, and Debraj Ray and seminar participants at NYU.

²Glosten and Milgrom (1985) proposed an alternative formalization of Bagehot's (1971) informal model.

observe the total volume of trade in each trading period. The market maker also knows (in equilibrium) the strategy of the informed trader, and sets prices efficiently, conditional on past and present volumes of trade.

Kyle constructed a linear equilibrium where in each period the price adjustment is proportional to the volume of trade, and the insider's volume of trade is proportional to the gap between the asset value and the current market price. The market maker's estimate of the asset value, reflected in the current market price, improves over time. As the public announcement date approaches, this estimate converges to the value of the asset and the insider trades frantically in her desire to exploit any price differential.

Our model differs from Kyle's model in three important ways. First, the fundamental value of the asset follows a Brownian motion and, therefore, changes continuously over time. Second, in addition to the initial observation, the insider continuously receives a signal of the current fundamental value of the asset. Third, the public announcement date is unpredictable: it has an exponential distribution.

The first difference by itself is irrelevant. In Kyle's model it makes no difference whether at time 0 the insider observes the true value of the asset or just an unbiased signal. Moreover, the model where the insider observes the true value and the value of the asset follows a Brownian motion is formally equivalent to a model where the initial observation is an unbiased signal of the final value of the asset. But this feature of our model becomes important when it is combined with the second feature. Finally, the third feature removes the pressure in Kyle's model behind the trade frenzy that occurs as the announcement date approaches. In our model, where the announcement date is not deterministic, the insider has no urgency to exhaust all arbitrage opportunities, and release all her private information in the process, by a particular deadline. Thus, while it is evident that in Kyle's model the price will become efficient (in the sense that it incorporates all the available information) as time reaches the announcement date, it is unclear whether in our model the insider will ever fully reveal her private information.

It is exactly this feature of the equilibrium of the fixed horizon model that Back (1992) exploited to develop his elegant “backward programming” solution method. In a model with a random horizon, Back's method is not directly applicable.

Our model is not the first to introduce a public announcement with random time. Back and Baruch (2004) compared the models of Kyle (1985) and Glosten and Milgrom (1985). To facilitate the comparison, they adopted a Glosten and Milgrom model with a single long-lived insider (who times her transactions strategically) and a Kyle model with a random terminal time and a risky asset that takes only the values 0 or 1.

Kyle's original model is in discrete time. However, Kyle also showed that as the period length Δ converges to 0, the equilibrium converges to an equilibrium of the continuous-time limit model. He then interpreted the continuous-time

model as a good representation of a discrete-time model where the agents can trade frequently. We maintain this interpretation and view the continuous-time model as a mathematical convenience that affords us the powerful tools of stochastic calculus. The discrete-time version of our model has a unique equilibrium that converges to a well defined strategy profile as $\Delta \downarrow 0$. However, in our case, this limit strategy is not an equilibrium of the continuous-time model. The interpretation of the continuous-time model is therefore delicate and needs to be examined more carefully. The lack of “continuity” arises because in the limit the insider wants to trade at infinite rates after some time T . She still collects positive rents after T even though the price perfectly tracks the value of the asset. However, after T , her payoff function evaluated at the limit strategy is 0. Therefore, as we explain in Section 5, the limit strategy cannot be an equilibrium of the continuous-time model. Because our characterization of the equilibrium has a crisper form in the limit, our discussion below refers to the limit equilibrium.

Our model includes various special cases. The value of the asset remains constant over time if the variance of its Brownian motion is reduced to 0. Since in our model the insider observes the initial value without noise, the signals that track the value of the asset over time become superfluous. This version of our model is similar to Kyle’s model, where the insider is endowed only with an initial piece of private information, but with a random end time. Alternatively, we can specialize our model to give the insider no initial informational advantage. This is accomplished by informing *all* traders of the initial value of the asset. In this version of the model, the insider’s informational advantage arises exclusively from her ability to observe the evolution of the asset value. This is an important model in its own right. An interesting question in this model is how the insider “manages” the information asymmetry. For example, the insider could let the information asymmetry (the variance of the uninformed traders’ estimate of the current value) grow to reach asymptotically a certain limit or grow without bound. The larger is the information asymmetry, the more likely it is that the market will substantially misprice the asset and, therefore, the larger are the profitable arbitrage opportunities. Thus, in this model as well it is not evident how much of the insider’s information is incorporated in the market price and how quickly this happens. We study this special case in the process of constructing an equilibrium for our general model. It turns out that in equilibrium the insider fully reveals her information as soon as she receives it. Hence, the market price equals the asset value at all times. Yet, the insider makes strictly positive profits. In independent work, Chau and Vayanos (2008) reached the same conclusion (for this case without initial informational asymmetry) in a slightly different model. They assumed that the insider receives a flow of information, the asset pays a dividend, and there is no public announcement. In addition, they assumed that the market maker continuously observes a noisy signal of the value of the asset. In the absence of this noisy signal, their model would be formally equivalent to ours. Chau and Vayanos

(2008) limited attention to the steady state of their model and did not study how the equilibrium approaches the steady state. One implication of our results is that in the absence of an initial information asymmetry, the steady state is reached “immediately” (as the period length goes to 0), so although Chau and Vayanos (2008) assumed that trading had been taking place indefinitely, this is not needed.

The equilibrium of our general model has a remarkable feature. There is a time T , endogenously determined in equilibrium, by which the insider reveals all her information (if the public announcement has not yet occurred). Thus, even though there is no deterministic deadline, the price converges to the asset value at time T . Moreover, time T divides the equilibrium into two phases. As long as the public announcement does not occur, in the interval $[0, T)$ the insider gradually transfers her information to the market and the market’s uncertainty about the value of the asset decreases to 0 monotonically. In the interval $[T, \infty)$, the insider trades large volumes and reveals her information immediately, so market prices track the asset value perfectly. Nevertheless, as we explained above, after T , the insider collects strictly positive rents.

There is a vast literature on insider trading³ and many papers have extended Kyle’s model. Holden and Subrahmanyam (1992) and Foster and Viswanathan (1996) considered a market with multiple competing insiders. They showed that competition among insiders accelerates the release of their private information. In a one-period model with heterogeneous insiders, Spiegel and Subrahmanyam (1992) replaced Kyle’s uninformed liquidity traders with strategic utility-maximizing agents trading for hedging purposes. In a multiperiod setting, Mendelson and Tunca (2004) proposed an alternative endogenous liquidity trading model that allowed for various types of market information. Similar to our model, Back and Pedersen (1998) considered the case where the insider continuously observes private information. To prove that an equilibrium exists, they assumed that the insider’s initial amount of private information is relatively high compared to the flow of new information and that this flow decreases fast over time. We show that a similar condition is required in the continuous-time version of our model. Furthermore, we also show that it is precisely when this condition is violated—that is, when the insider’s initial private information is small compared to the inflow on new information—that our equilibrium reaches market efficiency at a fixed time T .

The rest of the paper is organized as follows. Section 2 introduces the discrete-time model and Section 3 constructs a Markovian equilibrium. In Section 4, we show that the Markovian equilibrium has a well defined limit equilibrium as the period length goes to 0 and we provide a full characterization of it. In Section 5, we formulate the continuous-time model and show that the

³For a comprehensive review of this literature and its connection to the broader market microstructure theory, we refer the reader to O’Hara (1997), Brunnermeier (2001), Biais, Glosten, and Spatt (2005), Amihud, Mendelson, and Pedersen (2006), and references therein.

limit equilibrium is *not* an equilibrium. Section 6 includes our concluding remarks.

2. MODEL DESCRIPTION

The market participants are the insider, the market maker, and a (large) number of liquidity traders. The market maker opens the floor for trading only at discrete times $\{t_n\}_{n \geq 0}$. These trading dates are evenly spaced over time (e.g., once a day) so that $t_n = n\Delta$ for some positive constant Δ . The interval of time $[t_n, t_{n+1})$ is called period n . During period n , the following sequence of events occurs. First, the insider (and only her) receives private information about the fundamental value V_n of the asset. Then the insider and the liquidity traders simultaneously place buy/sell orders x_n and y_n , respectively, for a quantity of the asset. An order is a binding contract to buy/sell a quantity of the asset (the “size of the order”) at a price determined by the market maker. Finally, after observing the total volume of trade $z_n = x_n + y_n$, the market maker sets the price p_n and trades the necessary quantity to close all orders. We assume that the market maker is not able to differentiate between insider and liquidity trading. He only observes the net volume of trade z_n .

This trading process continues until a random time τ , independent of the history of transactions and prices, when the fundamental value of the asset becomes public knowledge. At this time, the market price immediately matches the fundamental value and the insider loses her informational advantage. We assume that the public announcement occurs at the end of a period (after trading). That is, $\tau = \eta\Delta$, where $\eta \geq 0$ has a geometric distribution with probability of failure $\rho = e^{-\mu\Delta}$ for some fixed $\mu > 0$.

Liquidity traders are not strategic agents and they trade for idiosyncratic reasons. In particular, we assume that $\{y_n\}_{n \geq 0}$ is a sequence of independent and identically distributed (i.i.d.) normal random variables with mean 0 and variance $\Sigma_y = \sigma_y^2\Delta$. On the other hand, the insider trades strategically so as to maximize her expected net payoff during $[0, \tau]$. The insider’s payoff is driven by her informational advantage as she alone observes the evolution of the fundamental value of the asset V_n during $[0, \tau]$. We assume that V_n evolves as a random walk $\{V_n = V_{n-1} + v_n\}_{n \geq 1}$, where V_0 is normally distributed with mean \bar{V}_0 and variance Σ_0 , and $\{v_n\}_{n \geq 1}$ is a sequence of i.i.d. normal random variables with mean 0 and variance $\Sigma_v = \sigma_v^2\Delta$. The market maker and the rest of the market participants only know the distributions of V_0 and $\{v_n\}_{n \geq 1}$. Hence, V_0 represents a lumpy endowment of private information that the insider gets at time 0, while $\{v_n\}_{n \geq 1}$ represents the incremental private information that she receives over time.

At the beginning of each period n , before the fundamental value becomes public knowledge, the market maker commits to a pricing rule (that is legally binding). The rule specifies the price p_n for the current period’s transactions as a function of the total volume of trade z_n and the public history up to this time.

(For completeness, we define $p_{-1} = \mathbb{E}[V_0] = \bar{V}_0$.) The insider and the liquidity traders place their orders after the rule is announced. All orders are executed at the end of the period. The market maker observes the public history of prices and (total) volumes of trade. His information in period n is represented by the history $\mathcal{F}_n^M = (z_0, p_0, \dots, z_{n-1}, p_{n-1}, z_n)$. Similarly, the insider's information includes the public history of prices and trades, the private history of her orders, and the fundamental values she has observed. Her information in period n is represented by the history $\mathcal{F}_n^I = (v_0, x_0, z_0, p_0, \dots, x_{n-1}, z_{n-1}, p_{n-1}, v_n)$. The insider places her order at the beginning of the period, after observing the current value of the fundamental.

The insider and the market maker are risk-neutral agents. Given a trajectory $X = \{x_n\}$ for the insider's trading and $P = \{p_n\}$ for market prices, the insider's payoff is

$$\Pi(P, X) = \sum_{n=0}^{\eta} [V_n - p_n] x_n.$$

The insider maximizes the expected value of $\Pi(P, X)$. Since η has a geometric distribution,

$$\mathbb{E}[\Pi(P, X)] = \mathbb{E}\left[\sum_{n=0}^{\infty} \rho^n [V_n - p_n] x_n\right].$$

DEFINITION 1: A strategy for the market maker is an \mathcal{F}_n^M -adapted process $P = \{p_n\}_{0 \leq n \leq \eta}$, and a strategy for the insider is an \mathcal{F}_n^I -adapted process $X = \{x_n\}_{0 \leq n \leq \eta}$. The profile (P, X) is an equilibrium if (i) for any $n \geq 0$

$$p_n = \mathbb{E}[V_n | X, \mathcal{F}_n^M],$$

and (ii) given P, X maximizes $\mathbb{E}[\Pi(P, X)]$. Here $\mathbb{E}[V_n | X, \mathcal{F}_n^M]$ means the conditional expectation of V_n given the public history \mathcal{F}_n^M at time n and the insider's strategy X , which specifies how she trades every period as a function of her information.

We do not model explicitly competition among market makers, but we implicitly assume that our market maker competes in prices with other market makers.⁴ In equilibrium, this competition drives the market maker to set the

⁴The model is not exactly a game and our definition of an equilibrium differs from that of a Nash equilibrium. However, Kyle (1985) suggested that the two definitions would coincide in a game where two market makers simultaneously bid prices after observing the current volume of trade and the winner gets the right to clear the market at the winning price. To avoid collusion, we can assume that there is a large population of market makers and that each market maker participates in the bidding game only once.

price equal to the expected value of the asset's fundamental value given the history of information he has observed and the insider's trading strategy (this is condition (i) in Definition 1). The insider chooses her strategy so as to maximize her expected discounted profit, given that she knows how the market maker will choose prices (this is condition (ii) in Definition 1).

We will restrict attention to Markovian equilibria with a particular state space. At the beginning of period n , before the market maker observes the volume of trade, the state is (n, \bar{V}_n, Σ_n) , where $\bar{V}_n = \mathbb{E}[V_n | \mathcal{F}_{n-1}^M, X]$ is the market maker's estimate of V_n and $\Sigma_n = \mathbb{E}[(V_n - \bar{V}_n)^2 | \mathcal{F}_{n-1}^M, X]$ is the variance of this estimate, conditional on the insider's strategy X and market information available at the end of period $n-1$. Note that since the market maker's estimate of V_n depends on the strategy X of the insider, the state and corresponding Markovian strategy profile need to be specified simultaneously.

DEFINITION 2: A strategy profile (P, X) is Markovian if for each n , the insider's order x_n and the market maker's price p_n depend only on the current state (n, \bar{V}_n, Σ_n) and the signals they receive in period n : v_n for the insider and z_n for the market maker. In this case, we write $x_n = X(n, \bar{V}_n, \Sigma_n, V_n)$ and $p_n = P(n, \bar{V}_n, \Sigma_n, z_n)$. If (P, X) is a Markovian strategy profile, let

$$\Pi_n(\bar{V}_n, \Sigma_n, V_n) = \mathbb{E} \left[\sum_{k=n}^{\eta} (V_k - p_k) x_k \mid \bar{V}_n, \Sigma_n, V_n, (P, X) \right]$$

be the insider's expected payoff-to-go for the transactions made from period n until the fundamental value is publicly revealed, when the current state is (n, \bar{V}_n, Σ_n) and the insider observes V_n . When (P, X) is a Markovian equilibrium, $p_{n-1} = \bar{V}_n$ for all n .

3. MARKOVIAN EQUILIBRIUM

In this section, we construct a *linear Markovian equilibrium* (P, X) , that is, a Markovian equilibrium such that

$$(1) \quad \begin{aligned} P(n, \bar{V}_n, \Sigma_n, z_n) &= \bar{V}_n + \lambda_n(\Sigma_n) z_n, \\ X(n, \bar{V}_n, \Sigma_n, V_n) &= \beta_n(\Sigma_n)(V_n - \bar{V}_n), \end{aligned}$$

where $\{\lambda_n\}$ and $\{\beta_n\}$ are sequences of functions $\lambda_n, \beta_n : \mathbb{R}_{++} \rightarrow \mathbb{R}_+$. The construction exploits the key property that the trajectory $\{\Sigma_n\}$ is *deterministic* and independent of the history of trades. As a result, the sequences $\{\lambda_n\}$ and $\{\beta_n\}$ are also deterministic and hereafter we drop the arguments Σ_n (we also drop this argument in the function Π_n). Moreover, since in equilibrium $\bar{V}_n = p_{n-1}$ for all n , hereafter we do not differentiate these two variables and write, for

example, $\Pi_n(p_{n-1}, V_n)$ instead of $\Pi_n(\bar{V}_n, \Sigma_n, V_n)$.

THEOREM 1: *There exist unique sequences $\{\lambda_n\}, \{\beta_n\} \in \mathbb{R}_{++}$ such that the linear strategy profile (P, X) defined by (1) is a Markovian equilibrium. In equilibrium, $\{\Sigma_n\}$ is a deterministic trajectory that is not affected by the (stochastic) choices of the insider and the market maker. Furthermore, there exist deterministic sequences $\{\alpha_n\}, \{\gamma_n\} \subset \mathbb{R}_{++}$ such that the insider's expected payoff-to-go for (P, X) satisfies*

$$(2) \quad \Pi_n(p, V) = \alpha_n(V - p)^2 + \gamma_n \quad \text{for all } n \geq 0.$$

Given $\Sigma_0 > 0$, there is a unique nonnegative value β_0 that generates the equilibrium profile $\{(\Sigma_n, \beta_n, \lambda_n, \alpha_n, \gamma_n)\}_{n \geq 0}$ recursively through the systems of equations

$$(3) \quad \Sigma_{n+1} = \Sigma_v + \frac{\Sigma_n \Sigma_y}{\beta_n^2 \Sigma_n + \Sigma_y}, \quad \beta_{n+1} \Sigma_{n+1} = \rho \beta_n \Sigma_n \left[\frac{\Sigma_y^2}{\Sigma_y^2 - \beta_n^4 \Sigma_n^2} \right],$$

$$\lambda_n = \frac{\beta_n \Sigma_n}{\beta_n^2 \Sigma_n + \Sigma_y},$$

$$(4) \quad \alpha_n = \frac{1 - \lambda_n \beta_n}{2 \lambda_n}, \quad \rho \gamma_{n+1} = \gamma_n - \frac{1 - 2 \lambda_n \beta_n}{2 \lambda_n (1 - \lambda_n \beta_n)} (\Sigma_v + \lambda_n^2 \Sigma_y),$$

where

$$(5) \quad \gamma_0 = \sum_{k=0}^{\infty} \rho^k \left(\frac{1 - 2 \lambda_k \beta_k}{2 \lambda_k (1 - \lambda_k \beta_k)} \right) (\Sigma_v + \lambda_k^2 \Sigma_y).$$

PROOF: Let us consider first the market maker's equilibrium condition (i.e., condition (i) in Definition 1). If the insider uses the trading strategy $x_n = \beta_n(V_n - p_{n-1})$ for some deterministic sequence $\{\beta_n\}$, then the market price in period n satisfies

$$p_n = \mathbb{E}[V_n | z_n = y_n + \beta_n(V_n - p_{n-1}), \mathcal{F}_{n-1}^M].$$

Conditional on the available market information \mathcal{F}_{n-1}^M , the pair (V_n, z_n) is a normally distributed two-dimensional random vector. Hence, by the projection theorem for normal random variables, we get that

$$\begin{aligned} p_n &= \mathbb{E}[V_n | \mathcal{F}_{n-1}^M] + \frac{\mathbb{C}\text{ov}[V_n, z_n | \mathcal{F}_{n-1}^M]}{\mathbb{V}\text{ar}[z_n | \mathcal{F}_{n-1}^M]} (z_n - \mathbb{E}[z_n | \mathcal{F}_{n-1}^M]) \\ &= p_{n-1} + \frac{\beta_n \Sigma_n}{\beta_n^2 \Sigma_n + \Sigma_y} z_n. \end{aligned}$$

Note that this pricing rule satisfies condition (1) with λ_n as in equation (3). In addition,

$$\begin{aligned}\Sigma_{n+1} &= \mathbb{V}\text{ar}[V_{n+1}|z_n, \mathcal{F}_{n-1}^M] = \Sigma_v + \mathbb{V}\text{ar}[V_n|z_n, \mathcal{F}_{n-1}^M] \\ &= \Sigma_v + \mathbb{V}\text{ar}[V_n|\mathcal{F}_{n-1}^M] \left(1 - \frac{\mathbb{C}\text{ov}[V_n, z_n|\mathcal{F}_{n-1}^M]^2}{\mathbb{V}\text{ar}[V_n|\mathcal{F}_{n-1}^M]\mathbb{V}\text{ar}[z_n|\mathcal{F}_{n-1}^M]}\right) \\ &= \Sigma_v + \frac{\Sigma_n \Sigma_y}{\beta_n^2 \Sigma_n + \Sigma_y}.\end{aligned}$$

Since Σ_{n+1} is independent of z_n , it follows that the sequence $\{\Sigma_n\}$ is indeed deterministic.

Let us now turn to the insider optimization problem in period n . Assume that the market maker uses the pricing rule $p_n = p_{n-1} + \lambda_n z_n$ for some constant λ_n . Furthermore, suppose that there exist two constants α_{n+1} and γ_{n+1} such that $\Pi_{n+1}(p, V) = \alpha_{n+1}(V - p)^2 + \gamma_{n+1}$. Then, the insider's expected payoff-to-go in period n , $\Pi_n(p_{n-1}, V_n)$, is

$$\begin{aligned}\max_{x_n} \mathbb{E}[(V_n - p_{n-1} - \lambda_n(x_n + y_n))x_n \\ + \rho(\alpha_{n+1}[V_n + v_n - p_{n-1} - \lambda_n(x_n + y_n)]^2 + \gamma_{n+1})].\end{aligned}$$

Under the condition $\rho\lambda_n\alpha_{n+1} < 1$ (otherwise the insider's payoff would be unbounded), the optimization problem above is concave in x and the optimal solution is obtained from the first-order condition

$$(6) \quad x_n = \beta_n(V_n - p_{n-1}), \quad \text{where } \beta_n = \frac{1 - 2\rho\lambda_n\alpha_{n+1}}{2\lambda_n(1 - \rho\lambda_n\alpha_{n+1})}.$$

Thus X_n defined by (1) is indeed the insider's best reply function. Plugging back the optimal value of x_n into the optimization above, we get that

$$\Pi_n(p_{n-1}, V_n) = \frac{(V_n - p_{n-1})^2}{4\lambda_n(1 - \rho\lambda_n\alpha_{n+1})} + \rho(\alpha_{n+1}(\Sigma_v + \lambda_n^2 \Sigma_y) + \gamma_{n+1}).$$

That is, $\Pi_n(p_{n-1}, V_n)$ is a quadratic function of the price gap $V_n - p_{n-1}$, as required, and the coefficients of $\Pi_n(p, V)$ satisfy the recursive equations $\alpha_n = [4\lambda_n(1 - \rho\lambda_n\alpha_{n+1})]^{-1}$ and $\gamma_n = \rho(\gamma_{n+1} + \alpha_{n+1}(\Sigma_v + \lambda_n^2 \Sigma_y))$. Use (6) and then replace the expression for λ_n in (3) to obtain

$$\alpha_n = \frac{1}{4\lambda_n(1 - \rho\lambda_n\alpha_{n+1})} = \frac{1 - \lambda_n\beta_n}{2\lambda_n} = \frac{\Sigma_y}{2\beta_n\Sigma_n}.$$

Invert (6) and then replace the expression for λ_n in (3) to obtain

$$\alpha_{n+1} = \frac{1 - 2\lambda_n\beta_n}{2\rho\lambda_n(1 - \lambda_n\beta_n)} = \frac{\Sigma_y^2 - \beta_n^4 \Sigma_n^2}{2\rho\beta_n\Sigma_n\Sigma_y} = \frac{\Sigma_y}{2\beta_{n+1}\Sigma_{n+1}}.$$

The last equality produces the equation for $\beta_{n+1}\Sigma_{n+1}$ in (3).

Note that the first two equations in (3) and (4) form an independent difference equation in (Σ_n, β_n) alone. For any initial value vector (Σ_0, β_0) , this difference equation has a unique solution. The value of Σ_0 is given, but the value of β_0 is “free.” The next two equations are static: the variables (λ_n, α_n) can be computed independently once the sequence $\{(\Sigma_k, \beta_k)\}$ has been constructed. Similarly, given $\{(\Sigma_k, \beta_k)\}$, $\{\gamma_k\}$ is uniquely defined by the initial value γ_0 and its linear dynamic equation in (4).

To complete the proof, we need to show that there exists a unique value for β_0 that generates—through the recursions (3) and (4)—a sequence $\{(\Sigma_n, \beta_n, \lambda_n, \alpha_n, \gamma_n)\}_{n \geq 0}$ that specifies an equilibrium. In the Appendix, we show that there exists a function $\Psi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that (i) $\beta_0 > \Psi(\Sigma_0)$ leads to bluffing and (ii) $\beta_0 < \Psi(\Sigma_0)$ leads to unbounded payoffs. As we discuss below, both bluffing and unbounded payoffs are not consistent with equilibrium. Hence, for each $\Sigma_0 > 0$, $\beta_0 = \Psi(\Sigma_0)$ is the only feasible choice.

Bluffing: If $\beta_0 > \Psi(\Sigma_0)$, then eventually $\beta_n < 0$ for some n (see the Appendix). That is, the insider bluffs trading on the wrong side of the spread. Moreover, when $\beta_n < 0$, (3) and (4) imply that $\lambda_n < 0$ and $\alpha_n < 0$, which is a contradiction.

Unbounded Payoffs: If $\beta_0 < \Psi(\Sigma_0)$, the sequence $\{(\Sigma_n, \beta_n)\}$ converges to $(\infty, 0)$. Then the market maker’s strategy specified by the corresponding sequence $\{\lambda_n\}$ allows the insider to extract unbounded payoffs (see Lemmas 1 and 3 in the Appendix). But the sequence $\{\beta_n\}$ generated by (3) and (4) generates a bounded payoff (see Lemma 2). Therefore, $\{\beta_n\}$ is not optimal for the insider, despite the fact that it was constructed to satisfy local optimality conditions. Thus, in this case, $(\{\lambda_n\}, \{\beta_n\})$ is not an equilibrium.

Iterate the recursive equation for γ_n in (4) to get

$$\begin{aligned} \rho^n \gamma_n &= \gamma_0 - \sum_{k=0}^{n-1} \rho^k \left(\frac{1 - 2\lambda_k \beta_k}{2\lambda_k(1 - \lambda_k \beta_k)} \right) (\Sigma_v + \lambda_k^2 \Sigma_y) \\ &= \gamma_0 - \sum_{k=1}^n \rho^k \alpha_k (\Sigma_v + \lambda_{k-1}^2 \Sigma_y). \end{aligned}$$

When $\beta_0 = \Psi(\Sigma_0)$, the sequence $\{(\Sigma_k, \beta_k, \lambda_k, \alpha_k)\}$ converges (and is independent of γ_0). Therefore, if we choose γ_0 as in (5), then $\lim_{n \rightarrow \infty} \rho^n \gamma_n = 0$. This implies that

$$\lim_{n \rightarrow \infty} \rho^n \mathbb{E}[\Pi_n(p_{n-1}, V_n)] = \lim_{n \rightarrow \infty} \rho^n (\alpha_n \Sigma_n + \gamma_n) = 0.$$

That is, the transversality condition is satisfied. Thus, the sequence $\{\beta_n\}$ is optimal against $\{\lambda_n\}$ and generates the continuation value functions $\Pi_n(p, V) = \alpha_n(V - p)^2 + \gamma_n$.

The optimality conditions included in (3) assume that the value function is quadratic. Lemmas 1 and 2 in the Appendix imply that for *any* deterministic equilibrium, the value function is quadratic. Thus, we conclude that there is a unique deterministic equilibrium $(\{\lambda_n\}, \{\beta_n\})$, determined by equations (3) and (4) and the initial condition $\beta_0 = \Psi(\Sigma_0)$. $Q.E.D.$

In general, we can characterize the set of linear Markovian equilibria as those sequences $\{(\Sigma_n, \beta_n, \lambda_n, \alpha_n, \gamma_n)\}_{n \geq 0}$ that satisfy the recursions in (3) and (4) and converge to the (unique) stationary equilibrium of the game $(\hat{\Sigma}, \hat{\beta}, \hat{\lambda}, \hat{\alpha}, \hat{\gamma})$ given by

$$(7) \quad \begin{aligned} \hat{\Sigma} &= \frac{1 + \sqrt{1 - \rho}}{\sqrt{1 - \rho}} \Sigma_v, \quad \hat{\beta} = \left(\frac{\Sigma_y(1 - \rho)}{\Sigma_v(1 + \sqrt{1 - \rho})} \right)^{1/2}, \\ \hat{\lambda} &= \left(\frac{\Sigma_v(1 - \sqrt{1 - \rho})}{\Sigma_y \rho} \right)^{1/2}, \\ \hat{\alpha} &= \frac{1}{2} \left(\frac{\Sigma_y}{\Sigma_v(1 + \sqrt{1 - \rho})} \right)^{1/2}, \quad \hat{\gamma} = \frac{\rho \hat{\alpha} (\Sigma_v + \hat{\lambda}^2 \Sigma_y)}{1 - \rho}. \end{aligned}$$

The following proposition highlights some additional properties of an equilibrium profile $\{(\Sigma_n, \beta_n)\}_{n \geq 0}$ and its proof follows directly from the properties of Ψ discussed in the proof of Theorem 1 in the Appendix. Some of these properties are used in the proof of Theorem 2 that characterizes the limiting continuous-time profile as Δ goes to 0.

PROPOSITION 1: *Let $\{(\Sigma_n, \beta_n)\}_{n \geq 0}$ be the discrete-time equilibrium of Theorem 1. Suppose Σ_0 is greater (less) than or equal to $\hat{\Sigma}$. Then the sequences $\{\Sigma_n\}$ and $\{\beta_n \Sigma_n\}$ are decreasing (increasing) in n .*

According to Proposition 1, despite the fact that the insider's trades are informative and reduce the market uncertainty, when the initial variance $\Sigma_0 < \hat{\Sigma}$, Σ_n increases with n . In this case, the variance reduction induced by insider trading is insufficient to compensate for the additional uncertainty generated by the evolution of $\{V_n\}$, and market prices are less informative over time.

4. CONTINUOUS-TIME APPROXIMATION

In this section, we analyze the limit of the discrete-time linear equilibrium of Theorem 1 as Δ goes to 0. In particular, we will show that the discrete-time linear Markovian equilibrium $\{(\Sigma_n, \beta_n, \lambda_n, \alpha_n, \gamma_n) : n \geq 0\}$ converges point-

wise to a continuous-time profile $\{(\Sigma_t, \beta_t, \lambda_t, \alpha_t, \gamma_t) : t \in \mathbb{R}_+\}$ in an appropriate sense.

First, let us explicitly rewrite the discrete-time model in terms of the calendar time t . For any time $t \geq 0$, the corresponding trading period is denoted by $n_t = \lfloor t/\Delta \rfloor$. We would like to express the insider's strategy $x_n = \beta_n(V_n - p_{n-1})$ in terms of her trading rate per unit of time. For this, we define $\beta^\Delta(t) = \beta_{n_t}/\Delta$. For any $t \geq 0$, define the continuous time extensions

$$\begin{aligned} p^\Delta(t) &= p_{n_t-1}, & V^\Delta(t) &= V_{n_t}, & \Sigma^\Delta(t) &= \Sigma_{n_t}, \\ \lambda^\Delta(t) &= \lambda_{n_t}, & \alpha^\Delta(t) &= \alpha_{n_t}, & \gamma^\Delta(t) &= \gamma_{n_t}, \\ \Pi^\Delta(t) &= \Pi_{n_t}(p^\Delta(t), V^\Delta(t)) = \alpha^\Delta(t)(V^\Delta(t) - p^\Delta(t))^2 + \gamma^\Delta(t), \end{aligned}$$

and the cumulative trading processes

$$X^\Delta(t) = \sum_{k=0}^{n_t} x_k, \quad Y^\Delta(t) = \sum_{k=0}^{n_t} y_k, \quad Z^\Delta(t) = X^\Delta(t) + Y^\Delta(t).$$

For ease of exposition, we assume that there exist two independent Brownian motions B_t^y and B_t^v such that $y_n = \sigma_y(B_{(n+1)\Delta}^y - B_{n\Delta}^y)$ and $v_n = \sigma_v(B_{(n+1)\Delta}^v - B_{n\Delta}^v)$. It follows that $Y^\Delta(t)$ and $V^\Delta(t)$ converge uniformly over compact sets to $Y_t = \sigma_y B_t^y$ and $V_t = \sigma_v B_t^v$, respectively. Also, in the limit, as $\Delta \downarrow 0$, τ is exponentially distributed with rate μ . Finally, recall that $\Sigma_y = \sigma_y^2 \Delta$, $\Sigma_v = \sigma_v^2 \Delta$, and $\rho = e^{-\mu\Delta}$.

THEOREM 2: *Let T be the unique nonnegative root of the equation*

$$\Sigma_0 + \sigma_v^2 T = \sigma_v^2 \left[\frac{e^{2\mu T} - 1}{2\mu} \right]$$

and define, for all $t \geq 0$,

$$(8) \quad \begin{aligned} \Sigma_t &= \frac{\sigma_v^2}{2\mu} [e^{2\mu(T-t)^+} - 2\mu(T-t)^+ - 1], & \beta_t &= \frac{\sigma_v \sigma_y e^{\mu(T-t)^+}}{\Sigma_t}, \\ \lambda_t &= \frac{\sigma_v e^{\mu(T-t)^+}}{\sigma_y}, & \alpha_t &= \frac{\sigma_y e^{-\mu(T-t)^+}}{2\sigma_v}, \\ \gamma_t &= \frac{\sigma_y \sigma_v e^{-\mu(T-t)^+}}{4\mu} [e^{2\mu(T-t)^+} + 2\mu(T-t)^+ + 3]. \end{aligned}$$

Then $(\Sigma^\Delta(t), \beta^\Delta(t), \lambda^\Delta(t), \alpha^\Delta(t), \gamma^\Delta(t))$ converges pointwise to $(\Sigma_t, \beta_t, \lambda_t, \alpha_t, \gamma_t)$ as $\Delta \downarrow 0$ for all $t \geq 0$. The market price $P^\Delta(t)$, the insider cumulative trading process $X^\Delta(t)$, and the market trading process $Z^\Delta(t)$ converge weakly to P_t ,

X_t , and Z_t , respectively, solutions of the system of stochastic differential equations (SDEs)

$$\begin{aligned} dZ_t &= dX_t + dY_t, \quad dP_t = \lambda_t dZ_t, \\ dX_t &= \begin{cases} \beta_t(V_t - P_t) dt, & \text{if } t < T, \\ \sigma_y dB_t^y + \sigma_y dB_t^v, & \text{if } t \geq T, \end{cases} \end{aligned}$$

with border conditions $Z_0 = X_0 = Y_0 = 0$ and $P_0 = \mathbb{E}[V_0]$. The insider expected payoff converges to $\Pi_t = \alpha_t(V_t - p_t)^2 + \gamma_t$.

As in Kyle's (1985) model, we could be tempted to argue that the limiting profile $(\Sigma_t, \beta_t, \lambda_t, \alpha_t, \gamma_t)$ is an equilibrium of a continuous-time model in which trades and prices change continuously. In Section 5, however, we will show that this (continuity) property does not hold in our model. Hence, we can only interpret the continuous-time profile $(\Sigma_t, \beta_t, \lambda_t, \alpha_t, \gamma_t)$ as an asymptotically good approximation of the discrete-time equilibrium of Theorem 1 when agents trade frequently. With this interpretation, we will refer to $(\Sigma_t, \beta_t, \lambda_t, \alpha_t, \gamma_t)$ as the *limit equilibrium*.

Theorem 2 reveals a number of important features of the limit equilibrium. A remarkable property is the existence of a finite time T , endogenously determined, such that $\Sigma_t = 0$ for $t \geq T$. That is, for Δ sufficiently small, the insider essentially reveals all her private information by time T . After T , the price always matches the fundamental value of the asset. Despite this market efficiency the insider is still able to collect positive rents ($\Pi(t, 0) = \gamma_T > 0$) in $t \in [T, \infty)$. The source of these rents is the continuous inflow of new information that the insiders gets by privately observing the evolution of the fundamental value. Indeed, one can show (see Theorem 3 below) that in the absence of these rents, either because V_t is constant or because the insider loses her capacity to track V_t , the insider would have no incentive to speed up her trading and market efficiency would only be reached asymptotically ($T = \infty$).

To get a sense of how likely it is that market efficiency is reached in the limit equilibrium, let us compare T and the average time $1/\mu$ at which the announcement date occurs. From the definition of T in Theorem 2, we can show that

$$T \leq \frac{1}{\mu} \quad \text{if} \quad \Sigma_0 \leq \left(\frac{e^2 - 3}{2} \right) \frac{\sigma_v^2}{\mu} \sim 2 \frac{\sigma_v^2}{\mu}.$$

Roughly speaking, the previous inequalities suggest that, on average, market efficiency is reached when the insider's initial (lumpy) private information Σ_0 is less than twice her average cumulative inflow of new private information σ_v^2/μ . Furthermore, one can show that as $\sigma_v \rightarrow \infty$, the switching time T converges to 0 and market efficiency is reach instantaneously. On the other hand, as $\sigma_v \downarrow 0$, the switching time T diverges to $+\infty$ and efficiency is only reached

asymptotically. The volatility coefficient σ_v determines the amount of information asymmetry. The following proposition shows that the higher is σ_v , the faster the insider reveals her information, but also the larger is her profit. Let $\mathbb{E}[\Pi_t]$ be the market's best estimate of the insider's expected continuation payoff from time t on, that is, $\mathbb{E}[\Pi_t] = \alpha_t \Sigma_t + \gamma_t$. Because of the deterministic evolution of Σ_t , α_t , and γ_t , $\mathbb{E}[\Pi_t]$ is also the insider's ex ante (at time 0, before observing any signals) expected payoff-to-go from t onward.

PROPOSITION 2: *In the limit equilibrium, the value of Σ_t weakly decreases with σ_v for all $t \geq 0$. On the other hand, $\mathbb{E}[\Pi_t]$ is equal to*

$$\mathbb{E}[\Pi_t] = \frac{\sigma_v \sigma_y}{\mu} \cosh(\mu(T-t)^+),$$

which is increasing in σ_v for all $t \geq 0$.

The more volatile is the fundamental value, the faster the price adjusts to the current intrinsic value. However, this efficiency comes at a cost. Indeed, the insider is willing to trade away her private information faster because the market maker compensates her for doing so. Hence, we expect market prices to be more informative when the volatility of the fundamental value is higher. For example, when there is no volatility ($\sigma_v = 0$), market efficiency ($\Sigma_t = 0$) is reached only asymptotically as $t \rightarrow \infty$ and the insider's ex ante payoff is minimized.

In a discrete-time equilibrium, the market maker's expected payoff is 0. This property is preserved in the limit equilibrium of Theorem 2. Thus, the liquidity traders' expected loss must equal the insider's expected profit, $\mathbb{E}[\Pi_t]$, which according to Proposition 2 decreases monotonically with time in $[0, T)$ and stays constant after T . Thus, liquidity traders who place their orders late in the game expect to make smaller losses.

Theorem 2 also shows that the market maker fulfills his obligation in a rather strong sense after T . He is concerned with setting prices so that $p_t = \mathbb{E}[V_t | \mathcal{F}_t^M]$. Theorem 2 implies that p_t converges uniformly on compact sets to V_t in $[T, \infty)$.⁵ As a result, after T , the insider trading volume X_t behaves as a Brownian motion and has unbounded variation. It is also interesting to note that $X_t - X_T$ is independent of σ_v .

Finally, we note that the limit equilibrium satisfies the smooth-pasting condition

$$\lim_{t \uparrow T} \dot{\Sigma}_t = 0.$$

This is in contrast to the equilibria obtained in models that assume a fixed announcement date (e.g., Kyle (1985)), where Σ_t does not approach 0 smoothly.

⁵This follows from the Skorohod representation theorem and the fact that $M_t = V_t - p_t$ converges weakly to (the continuous process) 0 for $t \geq T$.

5. CONTINUOUS-TIME EQUILIBRIUM

In this section, we formulate the continuous-time counterpart of the discrete-time model of Section 3 and show that the limit equilibrium of Theorem 2 is not an equilibrium of this model. However, we also show that in a modified continuous-time model where the insider's flow of new information is bounded, the limit equilibrium is an equilibrium.

Similar to the discrete-time model, we denote by V_t the fundamental value of the asset at time t which evolves as an arithmetic Brownian motion, $V_t = \sigma_v B_t^v$.

A strategy profile is a pair of processes (X, P) , where $X_t \in \mathcal{F}_t^I$ is the insider's cumulative trading volume up to time t , and $P_t \in \mathcal{F}_t^M$ is the price set by the market maker at time t . Following the formulation of the continuous-time model in Back (1992), we restrict the trading process X to the class \mathcal{S} of continuous, \mathcal{F}_t^I -adapted square-integrable semimartingales. This is a technical requirement that allows us to write the insider payoff as a stochastic integral of the market price with respect to her trading strategy. More precisely, for a given profile (X, P) , the insider's expected discounted payoff, $\mathbb{E}[\Pi(P, X)]$, is defined as

$$\mathbb{E}[\Pi(P, X)] = \mathbb{E}\left[V_\tau X_\tau - \int_0^\tau P_t dX_t - [X, P]_\tau\right],$$

where $[X, P]_t$ is the quadratic covariation between X_t and P_t .⁶ A continuous-time equilibrium is a profile (X, P) with the following properties: (i) given P , $X \in \mathcal{S}$ maximizes $\mathbb{E}[\Pi(X, P)]$, and (ii) the price process P satisfies the equilibrium condition

$$P_t = \mathbb{E}[V_t | \mathcal{F}_t^M, X], \quad 0 \leq t < \tau.$$

For the analysis that follows, we find it convenient to rewrite the insider's payoff using the following identity

$$V_\tau X_\tau = \int_0^\tau V_t dX_t + \int_0^\tau X_t dV_t + \int_0^\tau d[X, V]_t,$$

where $[X, V]_t$ is the quadratic covariation between X_t and V_t . Plugging this identity back into Π , taking expectation, and canceling the stochastic integral with respect to the martingale V_t , we get

$$\begin{aligned} \mathbb{E}[\Pi(P, X)] &= \mathbb{E}\left[\int_0^\infty e^{-\mu t} (V_t - P_t) dX_t\right. \\ &\quad \left. + \int_0^\infty e^{-\mu t} d[X, V]_t - \int_0^\infty e^{-\mu t} d[X, P]_t\right], \end{aligned}$$

⁶Intuitively, this term arises because the price paid by the insider is computed at the end of the period, and, therefore, it includes the effect of the insider's last trade dX_t . For a formal derivation, see equation (11) in Back (1992).

since τ is exponentially distributed with rate μ and is independent of \mathcal{F}_t^I .

Now we show that the strategy profile (P, X) associated to the limit equilibrium of Theorem 2 cannot be an equilibrium of the continuous-time model. Consider the insider's expected payoff-to-go from time T onward,

$$\begin{aligned}\Pi_T(P, X) = \mathbb{E} & \left[\int_T^\infty e^{-\mu(t-T)} (V_t - P_t) dX_t \right. \\ & \left. + \int_T^\infty e^{-\mu(t-T)} d[X, V]_t - \int_T^\infty e^{-\mu(t-T)} d[X, P]_t \right].\end{aligned}$$

After time T , the market maker's pricing strategy P is given by $dP_t = \lambda_T dZ_t$, where $\lambda_T = \sigma_v/\sigma_y$, and the insider's cumulative volume of trade is a martingale process such that $dX_t = \sigma_y [dB_t^v - dB_t^y]$. Thus, $V_t - P_t \equiv 0$, the first stochastic integral with respect to X_t has 0 expectation, and the quadratic covariations between X_t and V_t and between X_t and P_t satisfy $d[X, V]_t = \sigma_y \sigma_v dt$ and $d[X, P]_t = \lambda_T \sigma_y^2 dt = \sigma_y \sigma_v dt$, respectively. It follows that $\Pi_T(P, X) = 0$ and so X cannot be a best reply to P . This shows that there is a discontinuity in the insider's payoff function as we move from discrete time to continuous time. Indeed, recall that the insider's payoff Π^A of the discrete-time equilibrium of Theorem 2 satisfies $\lim_{\Delta \downarrow 0} \Pi_t^A = (\sigma_y \sigma_v)/\mu > 0$ for all $t \geq T$. This discontinuity is the result of the divergence of the insider's trading rate β_t^A to infinity for $t \geq T$ as $\Delta \downarrow 0$.⁷ In turn, this divergence is due to the existence of an unbounded flow of future private information. When the inflow of new information is small (for example, when $\sigma_v = 0$ because V_t is constant or the insider cannot track V_t after $t = 0$), the insider would collect small rents after the market reaches full efficiency. Therefore, the insider instead spends her private information slowly and market efficiency is reached only asymptotically ($T = \infty$). In this case the limit equilibrium of Theorem 2 is effectively an equilibrium of the continuous-time game. This and other properties of the equilibrium are summarized in the following theorem.

We now assume that the insider's strategy belongs to the space \mathcal{B} of trade rates β such that

$$(9) \quad \mathbb{E} \left[\int_0^\infty e^{-\mu t} |\beta_t| M_t^2 dt \right] < \infty.$$

Condition (9) rules out some bluffing schemes where the insider trades in the "wrong" direction and accumulates unbounded losses before accumulating unbounded gains.

⁷In a modified model with quadratic transactional costs, the insider's trading strategy would be bounded (as a referee suggested) and we expect the limit of discrete-time equilibria to be itself an equilibrium of the continuous-time model.

THEOREM 3: Suppose the asset's volatility $\sigma_v(t)$ is a function of time, and let Γ_t be the insider's cumulative inflow of private information from time t onward, that is,

$$\Gamma_t = \int_t^\infty \sigma_v^2(t) dt.$$

Assume that $\Gamma_0 < \infty$ and $(\Sigma_0 + \Gamma_0)e^{-2\mu t} > \Gamma_t$ for all t . When the insider's strategy space is constrained by (9), there exists a continuous-time linear Markovian equilibrium that satisfies

$$(10) \quad \Sigma_t = (\Sigma_0 + \Gamma_0)e^{-2\mu t} - \Gamma_t, \quad \lambda_t = \sqrt{\frac{2\mu(\Sigma_0 + \Gamma_0)}{\sigma_y^2}} e^{-\mu t}, \quad \beta_t = \frac{\sigma_y^2 \lambda_t}{\Sigma_t},$$

$$(11) \quad \alpha_t = \frac{e^{\mu t}}{2} \sqrt{\frac{\sigma_y^2}{2\mu(\Sigma_0 + \Gamma_0)}}, \quad \gamma_t = \alpha_t \Gamma_t + \frac{\sigma_y^2 \lambda_t}{4\mu}.$$

Under the conditions of Theorem 3, in equilibrium $\Sigma_t \downarrow 0$ as $t \rightarrow \infty$, but $\Sigma_t > 0$ for all $t \geq 0$. More importantly, the trading rate β_t remains bounded for all $t \geq 0$, so the insider's strategy is a process of bounded variation. When the flow of new information is substantial, the insider is happy to trade intensely to exploit current arbitrage opportunities. Even though in the process she "informs" the market about what she knows now, new arbitrage opportunities will develop soon. In the limit equilibrium, she transfers all her information (initial + flow) by time T , but when this flow is relatively low, she is not willing to trade that fast.

6. CONCLUSIONS

The paper introduces a model that combines a random announcement time with an insider who receives a flow of information. The new model produces a (limit) equilibrium with novel features. Two distinct regimes emerge. Before the endogenous time T , the insider is indifferent about how to consume her information stock, which includes the initial signal and the flow information she receives in the interval $(0, T]$. Nevertheless, in equilibrium she exhausts all this stock by time T , so that the market reaches full efficiency at time T . After T , she is eager to exhaust any additional piece of information immediately. As she does, she keeps the market fully informed until the public announcement, which reveals no further information.

The flow of new information, that in principle exacerbates the informational asymmetry, in equilibrium induces the insider to release her information faster. Interestingly, the market is uniformly better informed and reaches full efficiency earlier when this source of informational asymmetry (the variation of

the innovation process) is larger. However, the larger the asymmetry, the larger are the rents extracted by the insider.

The analysis also exposes a potential difficulty with continuous-time models. The natural discrete-time model has an equilibrium that, albeit difficult to construct explicitly, has a well defined limit as the period length decreases to 0. However, this limit equilibrium is *not* an equilibrium of the corresponding continuous-time model.

APPENDIX

DEFINITION OF Ψ : To characterize the function $\Psi(z)$, we find it convenient to introduce the change of variables

$$A_n = \frac{\Sigma_n}{\Sigma_v}, \quad B_n = \frac{\beta_n \Sigma_n}{\sqrt{\Sigma_y \Sigma_v}}.$$

Then equation (3) implies that $(A_{n+1}, B_{n+1}) = F(A_n, B_n)$, where

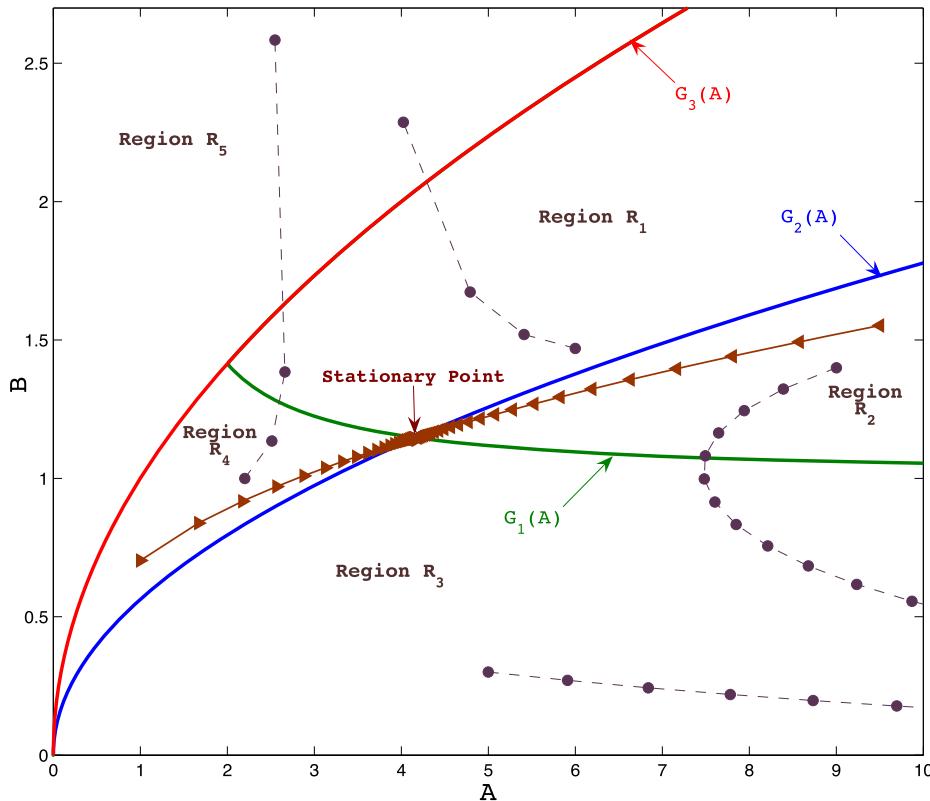
$$F_A(A_n, B_n) = 1 + \frac{A_n^2}{A_n + B_n^2}, \quad F_B(A_n, B_n) = \rho \left[\frac{A_n^2 B_n}{A_n^2 - B_n^4} \right].$$

Let

$$G_1(A) = \sqrt{\frac{A}{A-1}}, \quad G_2(A) = \sqrt{A}[1-\rho]^{1/4}, \quad G_3(A) = \sqrt{A}.$$

Since in equilibrium β_n must be positive for all n , a point (A, B) is *feasible* only if $F_B(A, B) \geq 0$, that is, only if $B \leq G_3(A)$. The function G_1 is defined so that $F_A(A, G_1(A)) = A$. If $B > G_1(A)$, then $F_A(A, B) < A$, and if $B < G_1(A)$, then $F_A(A, B) > A$. Similarly, the function G_2 is defined so that $F_B(A, G_2(A)) = B$. If $B > G_2(A)$, then $F_B(A, B) > B$, and if $B < G_2(A)$, then $F_B(A, B) < B$. As Figure 1 shows, the graphs of these functions partition the (A, B) space into five regions. In R_1 , $F(A, B)$ is always to the left and higher than (A, B) , and any sequence $\{(A_n, B_n)\}$ with initial point (A_0, B_0) in this region eventually crosses the graph of G_3 and becomes infeasible. In R_2 , $F(A, B)$ is always to the left and lower than (A, B) . In R_3 , $F(A, B)$ is always to the right and lower than (A, B) . In R_4 , $F(A, B)$ is always to the right and higher than (A, B) . R_5 is the region of infeasible points. In Figure 1 we have also plotted four sequences, each starting in a different region. A sequence that remains feasible must start in R_2 , R_3 , or R_4 , and any sequence that starts in R_3 always remain feasible, but not all sequences that start in R_2 or R_4 remain feasible. Sequences that start in R_1 always become infeasible.

By definition, the intersection of the graphs of G_1 and G_2 defines a stationary point (\hat{A}, \hat{B}) such that $(\hat{A}, \hat{B}) = F(\hat{A}, \hat{B})$ (see equation (7) for details of this stationary point in terms of the original state variables).

FIGURE 1.—Partition induced by the functions G_1 , G_2 , and G_3 .

Now, by continuity of the vector field F , there exists a curve \mathcal{C} , contained in $R_2 \cup R_4$ and passing through (\hat{A}, \hat{B}) , such that $F(A, B) \in \mathcal{C}$ for all $(A, B) \in \mathcal{C}$. That is, \mathcal{C} is the largest subset of \mathbb{R}^2 such that $F(\mathcal{C}) \subset \mathcal{C}$ and $(\hat{A}, \hat{B}) \in \mathcal{C}$. We do not have an analytical representation for \mathcal{C} , but we can approximate it numerically. This curve is strictly increasing and it approaches the origin to the left (but it does not contain it). Therefore, there exists a strictly increasing function $\psi : (0, \infty) \rightarrow (0, \infty)$, such that $(A, B) \in \mathcal{C}$ if and only if $B = \psi(A)$. For any initial $A_0 > 0$, let $B_0 = \psi(A_0)$. Then the sequence $\{(A_n, B_n)\}$, where $(A_{n+1}, B_{n+1}) = F(A_n, B_n)$ for each n , is contained in \mathcal{C} (that is, $B_n = \psi(A_n)$ for all $n \geq 0$) and, therefore, remains feasible forever. Moreover, $(A_n, B_n) \rightarrow (\hat{A}, \hat{B})$ as $n \rightarrow \infty$. When $A_0 < \hat{A}$ (respectively, $A_0 > \hat{A}$), $B_0 < \hat{B}$ ($B_0 > \hat{B}$) and $\{(A_n, B_n)\}$ is monotonically increasing (decreasing). In summary, for any given $\Sigma_0 > 0$, if we initialize

$$\beta_0 = \Psi(\Sigma_0), \quad \text{where} \quad \Psi(\Sigma_0) = \frac{\sqrt{\Sigma_y \Sigma_v}}{\Sigma_0} \psi\left(\frac{\Sigma_0}{\Sigma_v}\right),$$

we obtain a feasible sequence $\{(\Sigma_n, \beta_n, \lambda_n, \alpha_n)\}$. Moreover, this sequence converges. In particular, $\{\lambda_n\}$ is decreasing and converges to $\hat{\lambda} > 0$. Therefore, there exists $M > 0$ such that $\lambda_n \leq M$ for all n and

$$\sum_{n=1}^{\infty} \frac{\rho^n}{\lambda_n} < +\infty.$$

LEMMA 1: *Assume that the market maker's strategy $\{P_n\}$ is specified by a deterministic sequence $\{\lambda_n\} \subset \mathbb{R}_{++}$. Let*

$$S = \sum_{n=1}^{\infty} \frac{\rho^n}{\lambda_n}.$$

If $S = \infty$, then $\Pi_n(p, V) = \infty$ for all $n \geq 0$ and $(p, V) \in \mathbb{R}^2$. If $S < \infty$ and there is $M > 0$ such that $\lambda_n < M$ and $\rho \lambda_n / \lambda_{n+1} \leq 1$ for all $n \geq 0$, then there exist positive sequences $\{\alpha_n\}$ and $\{\gamma_n\}$ such that $\lambda_n \alpha_{n+1} \leq 1/2$ and $\rho \Pi_n(p, V) = \alpha_n(p - V)^2 + \gamma_n$ for all $n \geq 0$.

PROOF: For each $n \geq 0$ and each $k \geq 0$, consider the finite horizon problem for the insider where the fundamental value is made public at the end of period $n+k$ if it has not been publicly revealed before. Let $\Pi_{k,n}(p, V)$ be the insider's optimal discounted value from period n onward in this problem, when the price and fundamental value in period $n-1$ are (p, V) . Obviously, $\Pi_{k,n}(p, V) \leq \Pi_n(p, V)$ (because the insider can always choose $x_s = 0$ for all $s > n+k$) and $\lim_{k \rightarrow \infty} \Pi_{k,n}(p, V) = \Pi_n(p, V)$ for all $n \geq 0$ and all $(p, V) \in \mathbb{R}^2$.

We first show inductively in k that for each n , either

$$(12) \quad \Pi_{k,n}(p, V) = \frac{a_{k,n}}{\lambda_n} (V - p)^2 + \frac{b_{k,n}}{\lambda_n} \Sigma_v + c_{k,n} \lambda_n \Sigma_y$$

for some constants $(a_{k,n}, b_{k,n}, c_{k,n})$ or $\Pi_{k,n} \equiv \infty$. When $k = 0$,

$$\begin{aligned} \Pi_{0,n}(p, V) &= \max(V - p - \lambda_n x) x \\ &= \frac{(V - p)^2}{4\lambda_n}, \end{aligned}$$

so $\Pi_{0,n}$ satisfies (12) with $a_{0,n} = 1/4$ and $b_{0,n} = c_{0,n} = 0$ for all $n \geq 0$. By induction, assume first that $\Pi_{k,n+1}$ satisfies (12) for a given (k, n) . We then show that either $\Pi_{k+1,n}$ also satisfies (12) or $\Pi_{k+1,n} \equiv \infty$. We have that

$$\begin{aligned} \Pi_{k+1,n}(p, V) &= \max_{x \in \mathbb{R}} (V - p - \lambda_n x) x \\ &\quad + \rho \mathbb{E}[\Pi_{k,n+1}(V + W_n, p + \lambda_n(x + Y_n))] \end{aligned}$$

$$\begin{aligned}
&= \max_{x \in \mathbb{R}} (V - p - \lambda_n x) x \\
&\quad + \rho \left[\frac{a_{k,n+1}}{\lambda_{n+1}} [(V - p - \lambda_n x)^2 + \Sigma_v + \lambda_n^2 \Sigma_y] \right. \\
&\quad \left. + \frac{b_{k,n+1}}{\lambda_{n+1}} \Sigma_v + c_{k,n+1} \lambda_{n+1} \Sigma_y \right].
\end{aligned}$$

When $\rho a_{k,n+1} \lambda_n / \lambda_{n+1} < 1$, the quadratic objective function is concave and $\Pi_{k+1,n}$ satisfies (12) with

$$(13) \quad a_{k+1,n} = \frac{1}{4} \left[1 - a_{k,n+1} \frac{\rho \lambda_n}{\lambda_{n+1}} \right]^{-1},$$

$$(14) \quad b_{k+1,n} = \frac{\rho \lambda_n}{\lambda_{n+1}} [a_{k,n+1} + b_{k,n+1}], \quad c_{k+1,n} = \rho \left[a_{k,n+1} \frac{\lambda_n}{\lambda_{n+1}} + c_{k,n+1} \frac{\lambda_{n+1}}{\lambda_n} \right].$$

When $\rho a_{k,n+1} \lambda_n / \lambda_{n+1} \geq 1$, the quadratic objective function is convex and $\Pi_{k+1,n} \equiv \infty$. By induction, now assume instead that $\Pi_{k,n+1} \equiv \infty$. Then $\Pi_{k+n+1-s,s} \equiv \infty$ for all $s = 0, \dots, n$. This concludes the proof by induction.

Let us now assume that $\sum \rho^n / \lambda_n = \infty$. In this case, we will show that $\Pi_{k,n}(p, v) \rightarrow \infty$ as $k \rightarrow \infty$, for all n and (p, v) .

Since $a_{0,n} = 1/4$ and $\rho \lambda_n / \lambda_{n+1} > 0$ for all $n \geq 0$, it is easy to see (by induction) that (13) implies $a_{k,n} > 1/4$ for all $k \geq 1$ and $n \geq 0$. Fix $n \geq 0$. For any $k \geq 1$, if there exists $j \in \{1, \dots, k\}$ such that $1 \leq a_{k-j,t+j} \rho \lambda_{n+j} / \lambda_{n+j+1}$, then $\Pi_{k-j+1,n+j-1} \equiv \infty$, which implies that $\Pi_{k,n} \equiv \infty$ and $\Pi_n \equiv \infty$. Conversely, if $1 > a_{k-j,t+j} \rho \lambda_{n+j} / \lambda_{n+j+1}$ for all $j \in \{1, \dots, k\}$, then (14) implies that

$$\begin{aligned}
b_{k,n} &\geq \frac{\rho \lambda_n}{\lambda_{n+1}} \left[\frac{1}{4} + b_{k-1,n+1} \right] \geq \frac{\rho \lambda_n}{\lambda_{n+1}} \left[\frac{1}{4} + \frac{\rho \lambda_{n+1}}{\lambda_{n+2}} \left[\frac{1}{4} + b_{k-2,n+2} \right] \right] \geq \dots \\
&\geq \frac{\lambda_n}{4} \left[\frac{\rho}{\lambda_{n+1}} + \dots + \frac{\rho^k}{\lambda_{n+k}} \right].
\end{aligned}$$

Note that

$$\sum_{j=1}^{\infty} \frac{\rho^j}{\lambda_{n+j}} = \frac{1}{\rho^n} \sum_{j=t+1}^{\infty} \frac{\rho^j}{\lambda_j} = \frac{1}{\rho^n} \left[\sum_{j=1}^{\infty} \frac{\rho^j}{\lambda_j} - \sum_{j=1}^t \frac{\rho^j}{\lambda_j} \right] = \infty.$$

Therefore,

$$\Pi_n(p, v) \geq \Pi_{k,n}(p, v) \geq \frac{\Sigma_v}{\lambda_n} b_{k,n} \geq \frac{\Sigma_v}{4} \sum_{j=1}^k \frac{\rho^j}{\lambda_{n+j}} \quad \text{for all } k \geq 1,$$

which implies again that $\Pi_n \equiv \infty$ since the last term converges to ∞ as $k \rightarrow \infty$. Thus, either way, $\Pi_n \equiv \infty$.

Finally, assume that $\sum \rho^n / \lambda_n < \infty$ and $\rho \lambda_n / \lambda_{n+1} \leq 1$ for all $n \geq 1$. In this case, we show that each $\Pi_n(p, V)$ is a quadratic function of $(V - p)$.

Since $a_{0,n} = 1/4$, it is easy to show by induction that $1/4 < a_{k,n} < 1/2$ for all $k \geq 1$ and $n \geq 0$. The function $f(a, d) = [4(1 - ad)]^{-1}$ is increasing in a and d when $ad < 1$. Since $a_{1,n+1} > 1/4 = a_{0,n+1}$ for all $n \geq 0$, $a_{2,n} = f(a_{1,n+1}, d_n) > f(a_{0,n+1}, d_n) = a_{1,n}$ for all $n \geq 0$. Repeating this argument forward, we conclude that $\{a_{k,n}\}_{k=1}^{\infty}$ is an increasing sequence and it must converge. Let $\alpha_n = \lim_{k \rightarrow \infty} a_{k,n} / \lambda_n$. Now $a_{k,n+1} < 1/2$ for all k and $\rho \lambda_n / \lambda_{n+1} \leq 1$ imply that $\lambda_n \alpha_{n+1} \leq 1/2$.

Again, $a_{k,n} < 1/2$ for all $k \geq 0$ and $n \geq 0$ imply that

$$\begin{aligned} b_{k,n} &\leq \frac{\rho \lambda_n}{\lambda_{n+1}} \left[\frac{1}{2} + b_{k-1,n+1} \right] \leq \dots \leq \frac{\lambda_n}{2} \left[\frac{\rho}{\lambda_{n+1}} + \dots + \frac{\rho^k}{\lambda_{n+k}} \right] \\ &< \frac{\lambda_n}{2\rho^n} \sum_{j=t+1}^{\infty} \frac{\rho^j}{\lambda_j} < \infty. \end{aligned}$$

By induction in k , we now show that $b_{k,n} < b_{k+1,n}$ for all $k \geq 0$ and $n \geq 1$. Clearly $b_{0,n} = 0 < b_{1,n}$ for all $n \geq 0$. Since $a_{k,n+1} < a_{k+1,n+1}$, if the inequality holds for (k, n) , then

$$b_{k+1,n} = d_n[a_{k,n+1} + b_{k,n+1}] < d_n[a_{k+1,n+1} + b_{k+1,n+1}] = b_{k+2,n}.$$

That is, for each $n \geq 0$, the sequence $\{b_{k,n}\}_{k=0}^{\infty}$ is increasing and hence it must converge. Solving (14), we obtain

$$c_{k,n} = \frac{1}{\lambda_n} \sum_{j=1}^k \rho^j \frac{\lambda_{n+j-1}^2}{\lambda_{n+j}} a_{k-j,n+j}.$$

One can show that for each $n \geq 0$, the sequence $\{c_{k,n}\}_{k=1}^{\infty}$ is increasing, and since $\lambda_s \leq M$ and $a_{j,s} < 1/2$ for all j and s ,

$$c_{k,n} \leq \frac{M^2}{2\lambda_n} \sum_{j=1}^k \frac{\rho^j}{\lambda_{n+j}} < \frac{M^2}{2\lambda_n \rho^n} \sum_{j=t+1}^{\infty} \frac{\rho^j}{\lambda_j} < \infty$$

and the sequence must converge. Let $\gamma_n = \lim_{k \rightarrow \infty} [b_{k,n} \Sigma_v / \lambda_n + c_{k,n} \lambda_n \Sigma_y]$. Then $\Pi_n(p, v) = \alpha_n(v - p)^2 + \gamma_n$. Q.E.D.

LEMMA 2: *Let $\{\beta_n\}$ be an arbitrary deterministic strategy for the insider. Assume that $\{\lambda_n\}$ satisfies the equilibrium condition $p_n = \mathbb{E}[V | \{\beta_n\}, \mathcal{F}_n^M]$. Then the insider's expected payoff when she follows $\{\beta_n\}$ is finite.*

PROOF: The insider's payoff satisfies

$$\Pi = \sum_{n=0}^{\eta} (V_{\eta} - p_n) x_n = \sum_{n=0}^{\eta} \beta_n (V_{\eta} - p_n) (V_n - p_{n-1}).$$

Let us write p_n and V_n in terms of the primitive stochastic sequences $\{v_n\}$ and $\{y_n\}$ with $v_0 = V_0$. We have that $V_n = \sum_{n=0}^{\eta} v_n$. In addition,

$$\begin{aligned} p_n &= p_{n-1} + \lambda_n (\beta_n (V_n - p_{n-1}) + y_n) \\ &= (1 - \lambda_n \beta_n) p_{n-1} + \lambda_n \beta_n \sum_{n=0}^{\eta} v_n + \lambda_n y_n. \end{aligned}$$

Suppose that there exist sequences $A(n)$, $B(k, n)$, and $C(k, n)$ such that $A(0) = 1$, $B(0, 0) = C(0, 0) = B(k, n) = C(k, n) = 0$ for $k > n$, and

$$p_n = A(n) p_{-1} + \sum_{k=0}^n [B(k, n) v_k + C(k, n) y_k].$$

It follows that

$$\begin{aligned} A(n) &= (1 - \lambda_n \beta_n) A(n-1), \\ B(k, n) &= \lambda_n \beta_n + (1 - \lambda_n \beta_n) B(k, n-1) \quad \text{for } 0 \leq k \leq n, \\ C(k, n) &= (1 - \lambda_n \beta_n) C(k, n-1) \quad \text{for } 0 \leq k < n \quad \text{and} \\ C(n, n) &= \lambda_n. \end{aligned}$$

For $j \geq i$, let $\psi(i, j) := \prod_{k=i}^j (1 - \lambda_k \beta_k)$. Iterating the recursions above, we get

$$\begin{aligned} A(n) &= \psi(0, n), \quad B(k, n) = \sum_{j=k}^n \psi(j+1, n) \lambda_j \beta_j, \\ C(k, n) &= \psi(k+1, n) \lambda_k. \end{aligned}$$

The insider's conditional expected payoff given η is

$$\begin{aligned} \mathbb{E}[\Pi | \eta] &= \sum_{n=0}^{\eta} \beta_n \mathbb{E}[(V_{\eta} - p_n)(V_n - p_{n-1}) | \eta] \\ &= \sum_{n=0}^{\eta} \beta_n \mathbb{E}[(V_n - p_n)(V_n - p_{n-1})] = \sum_{n=0}^{\eta} [D_n^1 + D_n^2 + D_n^3], \end{aligned}$$

where

$$\begin{aligned} D_n^1 &= \beta_n A(n) A(n-1) p_{-1}, \\ D_n^2 &= \beta_n \sum_{k=0}^n (1 - B(k, n-1))(1 - B(k, n)) \Sigma_v(n), \\ D_n^3 &= \beta_n \sum_{k=0}^n C(k, n-1) C(k, n) \Sigma_y, \end{aligned}$$

$\Sigma_v(0) = \Sigma_0$, and $\Sigma_v(n) = \Sigma_v$ for $n > 0$. We now bound these terms. We have that

$$\begin{aligned} \lambda_n &= \frac{\beta_n \Sigma_n}{\beta_n^2 \Sigma_n + \Sigma_y} \\ \Rightarrow 1 - \lambda_n \beta_n &= \frac{\Sigma_y}{\beta_n^2 \Sigma_n + \Sigma_y} \in (0, 1) \quad \text{for all } n \geq 0. \end{aligned}$$

Hence, $A(n)$ is a nonnegative decreasing sequence with $0 \leq A(n) \leq A(0) \leq 1$. Moreover, $0 \leq B(k, n) \leq B(0, n)$ for all $0 \leq k \leq n$. The sequence $B(0, n)$ satisfies the recursion

$$B(0, n) = \lambda_n \beta_n + (1 - \lambda_n \beta_n) B(0, n-1), \quad B(0, 0) = 0.$$

Therefore, $0 \leq B(k, n) \leq B(0, n) \leq 1$. Also, the function $f(x) = x/(ax^2 + 1)$ reaches its global maximum at $x = \pm\sqrt{1/a}$. Therefore,

$$|\lambda_n| \leq \frac{\sqrt{\Sigma_n}}{1 + \Sigma_y} \leq \frac{1}{2} \sqrt{\frac{\Sigma_n}{\Sigma_y}}, \quad |\beta_n|(1 - \lambda_n \beta_n) = \lambda_n \frac{\Sigma_y}{\Sigma_n} \leq \frac{1}{2} \sqrt{\frac{\Sigma_y}{\Sigma_n}}.$$

Thus we obtain the bounds

$$|D_n^1| = |\beta_n|(1 - \lambda_n \beta_n) \prod_{k=0}^{n-1} (1 - \lambda_k \beta_k)^2 \leq \frac{1}{2} \sqrt{\frac{\Sigma_y}{\Sigma_n}} \leq \frac{1}{2} \sqrt{\frac{\Sigma_y}{\Sigma_v}},$$

since $1 - \lambda_k \beta_k \in (0, 1)$ and $\Sigma_n \geq \Sigma_v$ for all $n \geq 1$. For the second term we have that

$$\begin{aligned} |D_n^2| &= |\beta_n|(1 - \lambda_n \beta_n) \sum_{k=0}^n (1 - B(k, n-1))^2 \Sigma_v(n) \\ &\leq \frac{n+1}{2} \sqrt{\frac{\Sigma_y}{\Sigma_v}} \max\{\Sigma_0, \Sigma_v\}. \end{aligned}$$

Finally, for the third term we have that

$$|D_n^3| = |\beta_n|(1 - \lambda_n \beta_n) \sum_{k=0}^n C(k, n-1)^2 \Sigma_y \leq \frac{\Sigma_y}{2} \sqrt{\frac{\Sigma_y}{\Sigma_v}} \sum_{k=0}^n \lambda_k^2,$$

where the last inequality uses the fact that $C(k, n-1) \leq \lambda_k$. The recursion for Σ_{n+1} implies that $\Sigma_{n+1} \leq \Sigma_v + \Sigma_n$. Therefore, $\Sigma_n \leq n\Sigma_v + \Sigma_0$. We conclude that

$$|D_n^3| \leq \frac{\Sigma_y}{2} \sqrt{\frac{\Sigma_y}{\Sigma_v}} \sum_{k=0}^n \frac{\Sigma_n}{4\Sigma_y} \leq \frac{1}{8} \sqrt{\frac{\Sigma_y}{\Sigma_v}} \frac{(n+1)(n+2)}{2} \max\{\Sigma_0, \Sigma_v\}.$$

Combining all the pieces together we get that

$$|D_n| \leq \frac{1}{2} \sqrt{\frac{\Sigma_y}{\Sigma_v}} \left[1 + \left(1 + \frac{n+2}{8} \right) (n+1) \max\{\Sigma_0, \Sigma_v\} \right] \leq Kn^2$$

for some constant K . Therefore,

$$\mathbb{E}[\Pi] = \mathbb{E}[\mathbb{E}[\Pi|\eta]] = \sum_{n=0}^{\infty} \rho^n D_n < \infty. \quad Q.E.D.$$

LEMMA 3: Choose $\beta_0 < \Psi(\Sigma_0)$, and let $\{\lambda_n\}$ and $\{\beta_n\}$ be the corresponding strategies generated by (3) and (4) for the market maker and the insider. Then $\sum \rho^n / \lambda_n = \infty$ and the insider can make infinite profits. Moreover, $\{\beta_n\}$ is not a best reply against $\{\lambda_n\}$.

PROOF: We show that if $\beta_0 < \Psi(\Sigma_0)$, then $\sum \rho^n / \lambda_n = \infty$. Lemma 1 above then implies that the insider's expected payoff is unbounded. However, by Lemma 2, $\{\beta_n\}$ generates finite profits. Therefore, $\{\beta_n\}$ is not optimal against $\{\lambda_n\}$.

When $\beta_0 < \Psi(\Sigma_0)$, the sequence $\{(A_n, B_n)\}$ lies below \mathcal{C} and remains feasible forever. Moreover, for some finite N , $(A_n, B_n) \in R_3$ for all $n \geq N$. Therefore, $A_n < A_{n+1}$ for all $n \geq N$ and $A_n \rightarrow \infty$. Recall that the graphs of G_1 and G_2 intersect at (\hat{A}, \hat{B}) , and that $(A, B) \in R_3$ and $A \geq \hat{A}$ imply that $B \leq G_1(A)$. The function $h(A) = (A-1)^2/[A(A-2)]$ is decreasing for all $A > 2$, and $h(A) \rightarrow 1$ as $A \rightarrow \infty$. Let $\omega \in (\rho, 1)$. Without loss of generality, assume that N is such that $A_N \geq \hat{A}$ and $h(A_N) \leq \omega/\rho$. Then $B_n \leq G_1(A_n)$ for all $n \geq N$ and, therefore, for all $n \geq N$,

$$(15) \quad \begin{aligned} B_{n+1} &= F_B(A_n, B_n) = \rho \left[\frac{A_n^2 B_n}{A_n^2 - B_n^4} \right] \\ &\leq \rho \left[\frac{A_n^2 B_n}{A_n^2 - [G_1(A_n)]^4} \right] = \rho h(A_n) B_n \leq \omega B_n. \end{aligned}$$

Since $B_N \leq \hat{B}$, this implies that $B_n \leq \hat{B}\omega^{n-N}$ for all $n \geq N$. From equation (3),

$$\lambda_n = \frac{\beta_n \Sigma_n}{\beta^2 \Sigma_n + \Sigma_y} = \frac{A_n B_n}{A_n + B_n^2} \sqrt{\frac{\Sigma_v}{\Sigma_y}} < B_n \sqrt{\frac{\Sigma_v}{\Sigma_y}}.$$

Since we would like to show that $\sum \rho^n / \lambda_n = \infty$, we need a tighter upper bound on B_n . Note, however, that

$$B_{n+1} = F_B(A_n, B_n) = \rho \left[\frac{A_n^2 B_n}{A_n^2 - B_n^4} \right] \geq \rho B_n \quad \text{for all } n \geq 0,$$

so there is not a lot of slack in the previous upper bound (15) for B_{n+1} .

For any $\varepsilon > 0$, let $N^* > N$ be such that $\hat{B}\omega^{N^*-N} < \varepsilon$. Then, for all $n \geq N^*$,

$$\begin{aligned} A_{n+1} = F_A(A_n, B_n) &= 1 + \frac{A_n^2}{A_n + B_n^2} \geq 1 + \frac{A_n}{1 + \varepsilon^2/A_n} \\ &\geq 1 + A_n \left[1 - \frac{\varepsilon^2}{A_n} \right] = A_n + (1 - \varepsilon^2). \end{aligned}$$

Let $e = 1 - \varepsilon^2$. Then $A_{N^*+n} \geq A_{N^*} + ne > ne$ for all $n \geq N^*$. Feeding this bound back into (15), we obtain that

$$\begin{aligned} B_{N^*+n+3} &\leq \rho h((n+2)e) B_{N^*+2+n} \leq \dots \\ &\leq \rho^n h((n+2)e) h((n+1)e) \dots h(3e) B_{N^*+3}. \end{aligned}$$

Choose $\varepsilon < 1/4$ so that $\varepsilon^2 < 1/16$. Then, for all $k \geq 3$,

$$\begin{aligned} h(k e) &= \frac{[k-1-k\varepsilon^2]^2}{[k-k\varepsilon^2][k-2-k\varepsilon^2]} \\ &= 1 + \frac{1}{k(k-2)-2k(k-1)\varepsilon^2+k^2\varepsilon^4} \\ &< 1 + \frac{1}{k[k-2-2(k-1)\varepsilon^2]} < 1 + \frac{8}{k[7k-15]} \leq 1 + \frac{4}{k^2}. \end{aligned}$$

Let

$$\begin{aligned} H_n &= \left[1 + \frac{4}{1^2} \right] \left[1 + \frac{4}{2^2} \right] \dots \left[1 + \frac{4}{n^2} \right], \\ a_n &= \frac{1}{H_n} = \left[\frac{1^2}{1^2+4} \right] \dots \left[\frac{n^2}{n^2+4} \right]. \end{aligned}$$

Note that $[1 + 4/1^2][1 + 4/2^2] = 10$. Hence, $B_{N^*+n+3} < \rho^n B_{N^*+3} H_{n+2}/10$. Therefore,

$$\sqrt{\frac{\Sigma_v}{\Sigma_y}} \sum_{n \geq 1} \frac{\rho^n}{\lambda_n} > \sum_{n \geq 1} \frac{\rho^n}{B_n} > \sum_{n \geq 1} \frac{10\rho^{N^*+3+n}}{\rho^n H_{n+2} B_{N^*+3}} = \frac{10}{B_{N^*+3}} \rho^{N^*+3} \sum_{n \geq 3} a_n.$$

Gauss's test (see, for example, Knopp (1990)) states that if

$$\frac{a_{n+1}}{a_n} = 1 - \frac{c}{n} - \frac{g_n}{n^\varepsilon},$$

where $\varepsilon > 1$ and $\{g_n\}$ is bounded, then $\sum a_n$ converges when $c > 1$ and diverges when $c \leq 1$. In our case,

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^2}{(n+1)^2 + 4} = 1 - \left[\frac{4n^2}{(n+1)^2 + 4} \right] \frac{1}{n^2},$$

so $c = 0$ and $\varepsilon = 2$. Therefore $\sum a_n = \infty$, so $\sum \rho^n / \lambda_n = \infty$ and the insider makes infinite profits. *Q.E.D.*

The following lemma is used in the proof of Theorem 2.

LEMMA 4: *Let $\{f_n\}$ be a sequence of convex functions on $[T, \infty)$ (where $T \in \mathbb{R}$ is arbitrary). Assume that f_n converges pointwise to 0. That is, for all $t \geq T$, $f_n(t) \rightarrow 0$ as $n \rightarrow \infty$. Then, for each $t > T$, $\partial f_n(t) \rightarrow \{0\}$.*

PROOF: By contradiction, assume that there exists $t^* > T$ and a subgradient $s_n \in \partial f_n(t^*)$ for each n , such that $\{s_n\}$ does not converge to 0. Without loss of generality, assume that $s_n \rightarrow \bar{s} < 0$. Then, for each $t \in [T, t^*)$, $f_n(t) \geq f_n(t^*) + s_n(t - t^*)$, and taking limits as $n \rightarrow \infty$, we obtain $0 \geq -\bar{s}(t^* - t) > 0$, which is a contradiction. *Q.E.D.*

PROOF OF THEOREM 2: To emphasize the dependence of the discrete-time equilibrium on the length of a period, let us denote by $\{(\Sigma_n^\Delta, \beta_n^\Delta, \lambda_n^\Delta, \alpha_n^\Delta, \gamma_n^\Delta)\}$ the discrete-time equilibrium of Theorem 1 and denote by $(\hat{\Sigma}^\Delta, \hat{\beta}^\Delta, \hat{\lambda}^\Delta, \hat{\alpha}^\Delta, \hat{\gamma}^\Delta)$ the corresponding stationary equilibrium for an arbitrary $\Delta > 0$. Since $\Sigma_0 > 0$, it follows that $\Sigma_0 > \hat{\Sigma}^\Delta \sim O(\sqrt{\Delta})$ for Δ sufficiently small. As a result, the sequences $\{\Sigma_n^\Delta\}$ and $\{\lambda_n^\Delta\}$ are monotonically decreasing, while the sequences $\{\beta_n^\Delta\}$ and $\{\alpha_n^\Delta\}$ are monotonically increasing in n for Δ sufficiently small. Also, recall that the profile $(\Sigma^\Delta(t), \beta^\Delta(t), \lambda^\Delta(t), \alpha^\Delta(t), \gamma^\Delta(t))$ is a continuous-time piecewise-linear approximation of the discrete-time equilibrium such that $(\Sigma^\Delta(t), \beta^\Delta(t), \lambda^\Delta(t), \alpha^\Delta(t), \gamma^\Delta(t)) = (\Sigma_n^\Delta, \beta_n^\Delta/\Delta, \lambda_n^\Delta, \alpha_n^\Delta, \gamma_n^\Delta)$ for all $t \in [n\Delta, (n+1)\Delta)$.

The remaining proof is divided into two parts. In Part I, we show the pointwise convergence of $(\Sigma^\Delta(t), \beta^\Delta(t), \lambda^\Delta(t), \alpha^\Delta(t), \gamma^\Delta(t))$ to $(\Sigma_t, \beta_t, \lambda_t, \alpha_t, \gamma_t)$ as $\Delta \downarrow 0$. In Part II, we prove the weak convergence of $(P^\Delta(t), X^\Delta(t), Z^\Delta(t))$ to (P_t, X_t, Z_t) .

Part I. The proof of this part is organized as follows. First, we show that equations (3) and (4), which characterize the evolution of $\{(\Sigma_n^\Delta, \beta_n^\Delta, \lambda_n^\Delta, \alpha_n^\Delta, \gamma_n^\Delta)\}$, converge to a system of ordinary differential equations as $\Delta \downarrow 0$. Then we show that the solution of these ordinary differential equations (ODEs) defines a continuous-time profile $(\Sigma(t), \beta(t), \lambda(t), \alpha(t), \gamma(t))$ that is arbitrarily closed (as $\Delta \downarrow 0$) to $(\Sigma^\Delta(t), \beta^\Delta(t), \lambda^\Delta(t), \alpha^\Delta(t), \gamma^\Delta(t))$ for all $t < T$, where $T = \sup\{t > 0 : \Sigma(s) > 0, \forall s < t\}$. As in the discrete-time case, $\beta(0)$ is a free parameter for this continuous-time profile. Finally, we show that $\beta(0)$ is uniquely determined using two properties of the discrete-time equilibrium: (i) Σ_n^Δ is decreasing in n , which provides a lower bound on $\beta(0)$, and (ii) $\beta_n^\Delta \Sigma_n^\Delta$ is decreasing in n , which provides an upper bound on $\beta(0)$. We conclude the first part of the proof, showing that these upper and lower bounds coincide.

For notational convenience, let us define $q^\Delta(t) := \beta^\Delta(t)\Sigma^\Delta(t)$ and $r^\Delta(t) := \Delta/\Sigma^\Delta(t)$. The recursive equations (3) and (4) imply that

$$(16) \quad \begin{aligned} \frac{\Sigma^\Delta(t + \Delta) - \Sigma^\Delta(t)}{\Delta} &= \sigma_v^2 - \frac{(q^\Delta(t))^2}{\sigma_y^2 + (q^\Delta(t))^2 r^\Delta(t)}, \\ \frac{q^\Delta(t + \Delta) - q^\Delta(t)}{\Delta} &= \left[\frac{\sigma_y^4(e^{-\mu\Delta} - 1)/\Delta + (q^\Delta(t))^4 r^\Delta(t)/\Sigma^\Delta(t)}{\sigma_y^4 - (q^\Delta(t))^4(r^\Delta(t))^2} \right] q^\Delta(t), \\ \lambda^\Delta(t) &= \frac{q^\Delta(t)}{\sigma_v^2 + (q^\Delta(t))^2 r^\Delta(t)}, \quad \alpha^\Delta(t) = \frac{1 - \lambda^\Delta(t)\beta^\Delta(t)\Delta}{2\lambda^\Delta(t)}, \\ \frac{\gamma^\Delta(t + \Delta) - \gamma^\Delta(t)}{\Delta} &= \frac{e^{\mu\Delta} - 1}{\Delta} \gamma^\Delta(t) - \frac{1 - 2\lambda^\Delta(t)\beta^\Delta(t)\Delta}{2\lambda^\Delta(t)(1 - \lambda^\Delta(t)\beta^\Delta(t)\Delta)} (\sigma_v^2 + (\lambda^\Delta(t))^2 \sigma_y^2) e^{\mu\Delta}. \end{aligned}$$

For a given t , suppose that $\limsup_{\Delta \downarrow 0} (\beta^\Delta(t))^2 \Sigma^\Delta(t) < \infty$. Then $(q^\Delta(t))^2 r^\Delta(t)/\sqrt{\Delta}$ is negligible for Δ sufficiently small and, as $\Delta \downarrow 0$, the system of equations (16) converges to

$$(17) \quad \dot{\Sigma}(t) = \sigma_v^2 - \frac{q(t)^2}{\sigma_y^2}, \quad \dot{q}(t) = -\mu q(t),$$

$$(18) \quad \lambda(t) = \frac{1}{2\alpha(t)} = \frac{q(t)}{\sigma_y^2}, \quad \dot{\gamma}(t) = \mu \gamma(t) - \frac{\sigma_v^2 + \lambda(t)^2 \sigma_y^2}{2\lambda(t)}.$$

Since $\Sigma_0 > 0$, the condition $\limsup_{\Delta \downarrow 0} (q^\Delta(t))^2 r^\Delta(t) = 0$ is satisfied at $t = 0$ (otherwise $\liminf_{\Delta \downarrow 0} \lambda_\Delta^A = \infty$). It follows (by continuity) that the convergence above

holds for all $t \in [0, T)$ for some positive $T > 0$. Then by integrating (17) in this range, we obtain a continuous-time profile $(\Sigma(t), \beta(t), \lambda(t), \alpha(t), \gamma(t))$ given by

$$(19) \quad \begin{aligned} \Sigma(t) &= \Sigma_0 + \sigma_v^2 t - \frac{(\beta(0)\Sigma_0)^2}{2\mu\sigma_y^2}(1 - e^{-2\mu t}), \\ q(t) &= \beta(0)\Sigma_0 e^{-\mu t} \quad \text{for } t < T, \end{aligned}$$

for some constant of integration $\beta(0)$. Note that for this continuous-time solution, the condition $\limsup_{\Delta \downarrow 0} (q^\Delta(t))^2 r^\Delta(t) = 0$ reduces to $\Sigma(t) > 0$. Hence, given $\beta(0)$, T is uniquely determined as the smallest (positive) solution of the equation $\Sigma(t) = 0$. We denote by $T(\beta(0))$ this value which solves

$$0 = \Sigma_0 + \sigma_v^2 T - \frac{(\beta(0)\Sigma_0)^2}{2\mu\sigma_y^2}(1 - e^{-2\mu T}).$$

Suppose $\beta(0)$ is small enough so that $T(\beta(0)) = \infty$. Then $\Sigma(t) > 0$ for all $t \geq 0$ and $\lim_{\Delta \downarrow 0} (\Sigma^\Delta(t), \beta^\Delta(t)) = (\Sigma(t), \beta(t))$ for all $t \geq 0$. But in this case $\lim_{t \rightarrow \infty} \Sigma(t) = \infty$, which implies that $\lim_{t \rightarrow \infty} \Sigma^\Delta(t) = \infty$ for Δ sufficiently small, contradicting the monotonicity of Σ_n^Δ . As a result, $T(\beta(0)) < \infty$ and so $\beta(0)$ is bounded below by $\beta^L(0)$ such that $\dot{\Sigma}(t) = \dot{\Sigma}(t) = 0$ at $t = T(\beta^L(0))$. That is, $\beta^L(0)$ satisfies

$$\sigma_v^2 - \frac{(\beta^L(0)\Sigma_0)^2}{\sigma_y^2} e^{-2\mu T} = 0$$

where T solves

$$0 = \Sigma_0 + \sigma_v^2 T - \sigma_v^2 \left(\frac{e^{2\mu T} - 1}{2\mu} \right).$$

Suppose now that $\beta(0) > \beta^L(0)$ so that $T(\beta(0)) < \infty$. Then the fact that Σ_n^Δ is monotonically decreasing in n for all Δ and equation (19) lead to

$$\begin{aligned} \Sigma_t &= \lim_{\Delta \downarrow 0} \Sigma^\Delta(t) \\ &= \begin{cases} \Sigma_0 + \sigma_v^2 t - \frac{(\beta(0)\Sigma_0)^2}{2\mu\sigma_y^2}(1 - e^{-2\mu t}), & \text{if } t < T(\beta(0)), \\ 0, & \text{if } t \geq T(\beta(0)). \end{cases} \end{aligned}$$

In what follows, we show that for $t \geq T(\beta(0))$,

$$(20) \quad \lim_{\Delta \downarrow 0} \frac{\Sigma^\Delta(t + \Delta) - \Sigma^\Delta(t)}{\Delta} = 0.$$

Equation (16) and the fact that $\Sigma^\Delta(t)$ and $q^\Delta(t)$ are decreasing functions of t imply that $\Sigma^\Delta(t)$ is convex for all $t \geq T(\beta(0))$. Hence, by Lemma 4, $\Sigma^\Delta(t)$ satisfies equation (20). We can use this result together with the first equation in (16) to show that for $t \geq T(\beta(0))$,

$$\lim_{\Delta \downarrow 0} q^\Delta(t) = \lim_{\Delta \downarrow 0} \widehat{\beta}^\Delta \widehat{\Sigma}^\Delta = \sigma_v \sigma_y \quad \text{a.e.}$$

This follows from the fact that $q^\Delta(t)$ is decreasing in t for all Δ and, therefore, is bounded, which implies

$$\lim_{\Delta \downarrow 0} (q^\Delta(t))^2 r^\Delta(t) = \lim_{\Delta \downarrow 0} (\beta^\Delta(t) \Sigma^\Delta(t))^2 \frac{\Delta}{\Sigma^\Delta(t)} = 0$$

since $\Sigma^\Delta(t) \geq \widehat{\Sigma}^\Delta \geq \sigma_v(\Delta + \sqrt{\Delta}/\mu)$. On the other hand, from equation (19) we get that

$$\lim_{t \uparrow T(\beta(0))} \lim_{\Delta \downarrow 0} q^\Delta(t) = \beta(0) \Sigma_0 e^{-\mu T(\beta(0))}.$$

Hence, unless $\beta(0) \Sigma_0 e^{-\mu T(\beta(0))} = \sigma_v \sigma_y$ (or equivalently $\beta(0) = \beta^L(0)$), the limit function $q(t)$ would have a discontinuity at $t = T(\beta(0))$. But from the second equation in (16), such a discontinuity is not possible because the term

$$\left[\frac{\sigma_y^4 (e^{-\mu \Delta} - 1)/\Delta + (q^\Delta(t))^4 r^\Delta(t)/\Sigma^\Delta(t)}{\sigma_y^4 - (q^\Delta(t))^4 (r^\Delta(t))^2} \right] q^\Delta(t)$$

is uniformly bounded in t and Δ ⁸, which by the Arzelà–Ascoli theorem implies that the limit function $q(t)$ is continuous.

Part II. To prove the weak convergence of $(P^\Delta(t), X^\Delta(t), Z^\Delta(t))$ to (P_t, X_t, Z_t) , we introduce the price gap processes $M_t^\Delta := V_t^\Delta - P_t^\Delta$ and $M_t := V_t - P_t$, and show that M_t^Δ converges weakly to M_t . Specifically, we will invoke Theorem 2.1 in Prokhorov (1956) and prove the convergence of the finite-dimensional distributions of M_t^Δ to those of M_t , and then show the compactness of M_t^Δ in Δ .

For any $\Delta > 0$, the corresponding discrete-time equilibrium characterizes the values of M_t^Δ , β_t^Δ , and λ_t^Δ only at the discrete sequence of times $\{i\Delta\}_{i \geq 0}$. To extend these functions to \mathbb{R}_+ , we introduce the following notation: for any $t > 0$, we define, $n_t^\Delta := \lim_{s \uparrow t} \lfloor s/\Delta \rfloor$ and $\underline{t}^\Delta := n_t^\Delta \Delta$, and for any function f_t^Δ , we define $f_t^\Delta := f_{\underline{t}^\Delta}^\Delta$. Using a slight abuse of notation, we redefine the continuous piecewise linear version of M_t^Δ for any $t > 0$ as

$$M_t^\Delta = M_{\underline{t}^\Delta}^\Delta (1 - \lambda_{\underline{t}^\Delta}^\Delta \beta_{\underline{t}^\Delta}^\Delta (t - \underline{t}^\Delta)) + \sigma_v (B_t^v - B_{\underline{t}^\Delta}^v) - \lambda_{\underline{t}^\Delta}^\Delta \sigma_y (B_t^y - B_{\underline{t}^\Delta}^y),$$

⁸This follows from the fact that in equilibrium, $q^\Delta(t)$ is nonnegative and decreasing in t , and $(q^\Delta(t))^4 (r^\Delta(t))^2 \leq \sigma_y^4 (1 - \rho)$ (see Figure 1).

with border condition $M_0^\Delta = V_0 - \mathbb{E}[V_0]$. Since we are only concerned with the weak convergence of M_t^Δ , we will simplify the notation, replacing the term $\sigma_v(B_t^v - B_{t^\Delta}^v) - \lambda_{\underline{t}}^\Delta \sigma_y(B_t^y - B_{t^\Delta}^y)$ by $\sigma_{\underline{t}}^\Delta(B_t - B_{t^\Delta})$, where B_t is a Wiener process and $(\sigma_{\underline{t}}^\Delta)^2 = \sigma_v^2 + \sigma_y^2(\lambda_{\underline{t}}^\Delta)^2$. Iterating the recursion for M_t^Δ above, we get that

$$M_t^\Delta = M_0^\Delta A(0, n_t^\Delta) + \sum_{k=0}^{n_t^\Delta} A(k+1, n_t^\Delta) \sigma_{k\Delta}^\Delta (B_{(k+1)\Delta \wedge t^\Delta} - B_{k\Delta}),$$

where

$$A(j, n_t^\Delta) := \prod_{k=j}^{n_t^\Delta} (1 - \lambda_{k\Delta}^\Delta \beta_{k\Delta}^\Delta (\min\{(k+1)\Delta, t^\Delta\} - k\Delta)).$$

Since both λ_t^Δ and β_t^Δ are deterministic processes, it follows that M_t^Δ is a Gaussian process. Hence, its finite-dimensional distribution is fully characterized by its mean and variance–covariance processes. For $t > 0$, we have

$$\mu_t^\Delta := \mathbb{E}[M_t^\Delta] = M_0^\Delta A(0, n_t^\Delta).$$

From Part I, we know that (i) λ_t^Δ converges pointwise to a smooth, strictly positive, and bounded function λ_t for all $t \geq 0$, and (ii) the function β_t^Δ is nondecreasing in t for all $t \geq 0$ and converges pointwise to a smooth nondecreasing function β_t in $[0, T)$ and to infinity in $t \geq T$. We conclude that as $\Delta \downarrow 0$, then

$$\lim_{\Delta \downarrow 0} \mu_t^\Delta = \mathbb{1}(t < T) e^{-\int_0^t \lambda(s) \beta(s) ds},$$

where $\mathbb{1}(t < T)$ is the indicator function equal to 1 if $t < T$ and equal to 0 otherwise. Similarly, if we define $\Gamma^\Delta(t, t') := \mathbb{E}[(M_t^\Delta - \mu_t^\Delta)(M_{t'}^\Delta - \mu_{t'}^\Delta)]$ to be the variance–covariance process of M_t^Δ , then as $\Delta \downarrow 0$, we get (for $t < t'$)

$$\lim_{\Delta \downarrow 0} \Gamma^\Delta(t, t') = \mathbb{1}(t' < T) e^{-\int_t^{t'} \lambda_s \beta_s ds} \int_0^t e^{-2 \int_s^t \lambda_u \beta_u du} (\sigma_v^2 + \lambda_s^2 \sigma_y^2) ds.$$

Consider now the limiting gap process $M_t = V_t - P_t$. It follows from the system of SDEs in Theorem 2 that M_t satisfies the SDE

$$dM_t = -\lambda_t \beta_t M_t + \sigma_y dB_t^y - \lambda_t \sigma_y dB_t^y \quad \text{for all } t < T$$

and $M_t = 0$ for $t \geq T$. As a result, M_t is also a Gaussian process. Furthermore, for $t < T$, we can integrate the SDE above to get

$$M_t = M_0 e^{-\int_0^t \lambda_s \beta_s ds} + \int_0^t e^{-\int_s^t \lambda_u \beta_u du} (\sigma_v dB_s^v - \lambda_s \sigma_y dB_s^y).$$

It is a matter of simple calculations to show that the mean process $\mathbb{E}[M_t]$ and the variance–covariance process $\Gamma(t, t') = \mathbb{E}[(M_t - \mathbb{E}[M_t])(M_{t'} - \mathbb{E}[M_{t'}])]$ coincide with $\lim_{\Delta \downarrow 0} \mu_t^\Delta$ and $\lim_{\Delta \downarrow 0} \Gamma^\Delta(t, t')$ computed above. We conclude that the finite-dimensional distribution of M_t^Δ converges to the finite-dimensional distribution of M_t for all $t \geq 0$. In particular, it is worth noticing that M_T^Δ converges weakly to 0 as $\Delta \downarrow 0$.

We now prove that $\{M_t^\Delta : \Delta > 0\}$ is tight in $[0, T]$ for an arbitrary $T > 0$. For this we show that for every $\varepsilon > 0$

$$\lim_{\delta \downarrow 0} \limsup_{\Delta \downarrow 0} \mathbb{P}\left(\sup_{|t-s| \leq \delta} |M_t^\Delta - M_s^\Delta| \geq \varepsilon\right) = 0.$$

For $0 \leq s < t \leq T$ such that $i\Delta \leq s \leq (i+1)\Delta$ and $j\Delta \leq t \leq (j+1)\Delta$, it follows that

$$|M_t^\Delta - M_s^\Delta| \leq |M_{j\Delta}^\Delta - M_{i\Delta}^\Delta| + |M_{(i+1)\Delta}^\Delta - M_{i\Delta}^\Delta| + |M_{(j+1)\Delta}^\Delta - M_{j\Delta}^\Delta|.$$

For notational convenience, let us introduce the notation $M_i^\Delta = M_{i\Delta}^\Delta$ (a similar notation is used for $\beta_i^\Delta, \lambda_i^\Delta, \sigma_i^\Delta, \Sigma_i^\Delta$, and B_i). Defining $\delta^\Delta := \lfloor \delta/\Delta \rfloor + 1$, $T^\Delta := \lfloor T/\Delta \rfloor$, and $\mathcal{T}^\Delta := \lfloor T/\Delta \rfloor$, for Δ sufficiently small, the previous inequality implies that

$$\mathbb{P}\left(\sup_{|t-s| \leq \delta} |M_t^\Delta - M_s^\Delta| \geq \varepsilon\right) \leq 3\mathbb{P}\left(\sup_{|j-i| \leq \delta^\Delta} |M_j^\Delta - M_i^\Delta| \geq \varepsilon/3\right).$$

Using the recursion for M_i^Δ , we get that

$$|M_j^\Delta - M_i^\Delta| \leq \sum_{k=i}^{j-1} \lambda_k^\Delta \beta_k^\Delta \Delta |M_k^\Delta| + \left| \sum_{k=i}^{j-1} \sigma_k^\Delta (B_{k+1} - B_k) \right|$$

and so

$$\begin{aligned} & \mathbb{P}\left(\sup_{|j-i| \leq \delta^\Delta} |M_j^\Delta - M_i^\Delta| \geq \varepsilon\right) \\ & \leq \mathbb{P}\left(\sup_{|j-i| \leq \delta^\Delta} \sum_{k=i}^{j-1} \lambda_k^\Delta \beta_k^\Delta \Delta |M_k^\Delta| \geq \varepsilon/2\right) \\ & \quad + \mathbb{P}\left(\sup_{|j-i| \leq \delta^\Delta} \left| \sum_{k=i}^{j-1} \sigma_k^\Delta (B_{k+1} - B_k) \right| \geq \varepsilon/2\right). \end{aligned}$$

From Part I we know that $\lambda^\Delta(t)$ is uniformly bounded in Δ and t and strictly positive. So, for the purpose of the result that we need to prove, we can conve-

niently assume that without loss of generality (w.l.o.g.), $\lambda_k^\Delta = 1$ (and $\sigma_k^\Delta = 1$). It follows that for the last term on the right that

$$\lim_{\delta \downarrow 0} \limsup_{\Delta \downarrow 0} \mathbb{P} \left(\sup_{|j-i| \leq \delta^\Delta} \left| \sum_{k=i}^{j-1} \sigma_k^\Delta (B_{k+1} - B_k) \right| \geq \varepsilon/2 \right) = 0$$

(e.g., by invoking Lévy's theorem on the modulus of continuity for Brownian motion). Hence, to complete the proof, it is now sufficient to show that

$$(21) \quad \lim_{\delta \downarrow 0} \limsup_{\Delta \downarrow 0} \mathbb{P} \left(\sup_{|j-i| \leq \delta^\Delta} \sum_{k=i}^{j-1} \beta_k^\Delta |M_k^\Delta| \Delta \geq \varepsilon \right) = 0.$$

From the definition of M_k^Δ (and the assumption $\lambda_k^\Delta = \sigma_k^\Delta = 1$), we get that

$$\begin{aligned} \beta_k^\Delta |M_k^\Delta| &\leq |M_0^\Delta| \beta_k^\Delta \prod_{j=0}^{k-1} (1 - \beta_j^\Delta \Delta) \\ &+ \left| \sum_{j=0}^{k-1} \beta_k^\Delta \left(\prod_{n=j+1}^{k-1} (1 - \beta_n^\Delta \Delta) \right) (B_{j+1} - B_j) \right|. \end{aligned}$$

Suppose $\mathcal{T} < T$. Then in the region $t \in [0, \mathcal{T}]$, the function β_t^Δ is uniformly bounded in Δ and t , and the condition in equation (21) will follow. So, let us assume that $\mathcal{T} \geq T$. From Part I, it follows that there exist positive constants

$$\frac{K_1}{\Sigma_t^\Delta} \leq \beta_t^\Delta \leq \frac{K_2}{\Sigma_t^\Delta} \quad \text{for } t \in [0, \mathcal{T}].$$

This follows from the fact that $\beta_t^\Delta \Sigma_t^\Delta$ converges to a positive bounded function as Δ goes to zero. Furthermore, given the limiting behavior of Σ_t^Δ as $\Delta \downarrow 0$, one can show that for any $t \in [T, \mathcal{T}]$, there exists a positive constant K_t (independent of Δ) such that $\Sigma_s^\Delta \leq \Sigma_t^\Delta + K_t(t-s)^2$ for all $s \in [0, t]$. As a result, in $[T, \mathcal{T}]$ we get that

$$\begin{aligned} &\sup_{|j-i| \leq \delta^\Delta} |M_0^\Delta| \Delta \sum_{k=i}^{j-1} \beta_k^\Delta \prod_{j=0}^{k-1} (1 - \beta_j^\Delta \Delta) \\ &\leq |M_0^\Delta| (\delta + \Delta) \max_{T^\Delta \leq k \leq \mathcal{T}^\Delta} \left\{ \beta_k^\Delta \exp \left(- \sum_{j=0}^{k-1} \beta_j^\Delta \Delta \right) \right\} \\ &\leq |M_0^\Delta| (\delta + \Delta) \max_{T^\Delta \leq k \leq \mathcal{T}^\Delta} \left\{ \frac{K_2}{\Sigma_k^\Delta} \exp \left(- \int_0^{k\Delta} \frac{K_1 dt}{\Sigma_k^\Delta + K_{k\Delta}(k\Delta - s)^2} \right) \right\} \end{aligned}$$

$$\begin{aligned}
&= |M_0^\Delta|(\delta + \Delta) \max_{T^\Delta \leq k \leq T^\Delta} \left\{ \frac{K_2}{\Sigma_k^\Delta} \exp \left(-\frac{K_1}{\sqrt{K_{k\Delta} \Sigma_k^\Delta}} \arctan \left(\frac{k\Delta}{\sqrt{K_{k\Delta} \Sigma_k^\Delta}} \right) \right) \right\} \\
&\xrightarrow{\Delta \downarrow 0} 0,
\end{aligned}$$

where the convergence follows from the fact that $\Sigma_t^\Delta \rightarrow 0$ as $\Delta \downarrow 0$ for any $t \geq T$. It follows from the previous derivation that there exists a constant \bar{K} independent of Δ such that

$$\max_{T^\Delta \leq k \leq T^\Delta} \left\{ \beta_k^\Delta \prod_{n=0}^{k-1} (1 - \beta_n^\Delta \Delta) \right\} \leq \bar{K}.$$

Finally, we have that

$$\begin{aligned}
&\mathbb{P} \left(\sup_{|j-i| \leq \delta^\Delta} \sum_{k=i}^{j-1} \Delta \left| \sum_{j=0}^{k-1} \beta_k^\Delta \left(\prod_{n=j+1}^{k-1} (1 - \beta_n^\Delta \Delta) \right) (B_{j+1} - B_j) \right| \geq \varepsilon \right) \\
&\leq \mathbb{P} \left((\delta + \Delta) \times \max_{T^\Delta \leq k \leq T^\Delta} \left\{ \left| \sum_{j=0}^{k-1} \beta_k^\Delta \left(\prod_{n=j+1}^{k-1} (1 - \beta_n^\Delta \Delta) \right) (B_{j+1} - B_j) \right| \right\} \geq \varepsilon \right) \\
&\leq \mathbb{P} \left((\delta + \Delta) \bar{K} \max_{T^\Delta \leq k \leq T^\Delta} \left\{ \left| \sum_{j=0}^{k-1} \prod_{n=0}^j (1 - \beta_n^\Delta \Delta)^{-1} (B_{j+1} - B_j) \right| \right\} \geq \varepsilon \right) \\
&\leq \frac{(\delta + \Delta)^2 \bar{K}^2}{\varepsilon^2} \mathbb{E} \left[\sum_{j=0}^{T^\Delta-1} \prod_{n=0}^j (1 - \beta_n^\Delta \Delta)^{-2} \Delta \right] \\
&\leq \frac{(\delta + \Delta)^2 \bar{K}^2 T^\Delta}{\varepsilon^2} \max_{0 \leq j \leq T^\Delta} \prod_{n=0}^j (1 - \beta_n^\Delta \Delta)^{-2}.
\end{aligned}$$

The third inequality uses Doob's inequality. From Part I we know that β_n^Δ is an increasing function of n and that it converges to $\beta_\infty^\Delta = \hat{K}/\sqrt{\Delta}$ for a fixed constant \hat{K} independent of Δ . As a result, $\max_{0 \leq j \leq T^\Delta} \prod_{n=0}^j (1 - \beta_n^\Delta \Delta)^{-2}$ is uniformly bounded. We conclude that the $\lim_{\delta \downarrow 0} \limsup_{\Delta \downarrow 0}$ of the probability above is equal to zero as required. $Q.E.D.$

PROOF OF PROPOSITION 2: Recall from Theorem 2 that Σ_t satisfies

$$\Sigma_t = \Sigma_0 + \sigma_v^2 t - \sigma_v^2 e^{2\mu t} \left[\frac{1 - e^{-2\mu t}}{2\mu} \right] \quad \text{for } t < T$$

and $\Sigma_T = 0$ for all $t \geq T$, where $T \geq 0$ is the unique solution to

$$\Sigma_0 + \sigma_v^2 T = \sigma_v^2 \left[\frac{e^{2\mu T} - 1}{2\mu} \right].$$

Since T decreases with σ_v , it suffices to prove that Σ_t decreases with σ_v for $t < T$.

In what follows, and without loss of generality, we will normalize the value of μ such that $2\mu = 1$ (this is equivalent to rescaling time). With this normalization, the derivative of Σ_t ($t < T$) with respect to σ_v^2 is equal to

$$\frac{\partial \Sigma_t}{\partial \sigma_v^2} = t - e^T (1 - e^{-t}) - \sigma_v^2 e^T (1 - e^{-t}) \frac{\partial T}{\partial \sigma_v^2} \quad \text{for } t < T.$$

In addition, from the definition of T , it follows that

$$\frac{\partial T}{\partial \sigma_v^2} = \frac{1}{\sigma_v^2} \left[\frac{1 + T - e^T}{e^T - 1} \right].$$

Plugging this value back onto $\partial \Sigma_t / \partial \sigma_v^2$, we get that for $t < T$,

$$\frac{\partial \Sigma_t}{\partial \sigma_v^2} = t - (1 - e^{-t}) \left[\frac{T}{1 - e^{-T}} \right] \leq 0.$$

The inequality follows from the fact that $t/(1 - e^{-t})$ is an increasing function of t .

Let us now prove the monotonicity of the insider's ex ante expected payoff. First of all, from the expressions for Σ_t , α_t , and γ_t in Theorem 2, it follows that $\mathbb{E}[\Pi_t] = \alpha_t \Sigma_t + \gamma_t$ is equal to (under the normalization $2\mu = 1$)

$$\mathbb{E}[\Pi_t] = 2\sigma_y \sigma_v \cosh\left(\frac{1}{2}(T - t)^+\right) \quad \text{for } t \geq 0.$$

Note that to prove the monotonicity of $\mathbb{E}[\Pi_t]$ with respect to σ_v , it is enough to focus on the case $t \leq T$. The derivative with respect to σ_v is given by

$$\begin{aligned} \frac{\partial \mathbb{E}[\Pi_t]}{\partial \sigma_v} &= 2\sigma_y \cosh\left(\frac{1}{2}(T - t)\right) + \sigma_y \sigma_v \sinh\left(\frac{1}{2}(T - t)\right) \frac{\partial T}{\partial \sigma_v} \\ &= 2\sigma_y \cosh\left(\frac{1}{2}(T - t)\right) + 2\sigma_y \sinh\left(\frac{1}{2}(T - t)\right) \left[\frac{1 + T - e^T}{e^T - 1} \right] \end{aligned}$$

$$\begin{aligned}
&= 2\sigma_y \sinh\left(\frac{1}{2}(T-t)\right) \left[\frac{T}{e^T - 1} \right] + 2\sigma_y \exp\left(\frac{T-t}{2}\right) \\
&\geq 0. \quad Q.E.D.
\end{aligned}$$

PROOF OF THEOREM 3: The insider's Hamilton–Jacobi–Bellman (HJB) optimality condition are given by

$$\begin{aligned}
0 &= \max_{\beta} \left\{ -\lambda_t \beta M \Pi_M + \frac{1}{2} \lambda_t^2 \sigma_y^2 \Pi_{MM} + \Pi_t - \mu \Pi + M^2 \beta \right\} \\
\text{for } t &\in [0, \infty).
\end{aligned}$$

Suppose, we guess a quadratic value function of the form $\Pi(t, M) = \alpha_t M^2 + \gamma_t$ for deterministic functions α_t and γ_t . The HJB equation is satisfied if and only if $\dot{\alpha}_t - \mu \alpha_t = 0$, $1 - 2\lambda_t \alpha_t = 0$, and $\alpha_t (\sigma_v^2(t) + \lambda_t^2 \sigma_y^2) + \dot{\gamma}_t - \mu \gamma_t = 0$. The first two conditions lead to $\lambda_t = \lambda_0 e^{-\mu t}$ and $\alpha_t = e^{\mu t} / [2\lambda_0]$ for some constant $\lambda_0 > 0$. Replacing these two functions, the solution of the last differential equation is

$$\gamma_t = \frac{1}{2\lambda_0} (C + \Gamma_t) e^{\mu t} + \frac{\sigma_y^2 \lambda_t}{4\mu}$$

for some constant $C \geq 0$ (since $\Gamma_t \downarrow 0$ as $t \rightarrow \infty$, $C \geq 0$ is required to ensure that $\gamma_t \geq 0$ for all t).

Note that the HJB condition does not provide any information about how to select the insider's strategy β_t . (Effectively, we have solved the HJB equation using the fact that the insider is indeed indifferent.) To determine the value of β_t , we must turn to the market maker's filtering conditions. The condition $P_t = \mathbb{E}[V_t | \mathcal{F}_t^M]$ implies that P_t is the orthogonal projection V_t on \mathcal{F}_t^M in L^2 , and we can interpret the equilibrium market price as the solution to a classical Kalman–Bucy filtering problem. Let the signal process be the value of the fundamental V_t , with dynamics $dV_t = \sigma_v dB_t^v$ and the observation process be the price process P_t , with dynamics $dP_t = \lambda_t dZ_t = \beta_t \lambda_t (V_t - P_t) dt + \sigma_y \lambda_t dB_t^y$. Let v_t be the corresponding optimal (in mean square sense) filtering estimate of V_t and let Σ_t be the filtering error. Then the equilibrium condition is $P_t = v_t$. The generalized Kalman filter conditions for the pair (V_t, P_t) are given by

$$dv_t = \frac{\Sigma_t \beta_t}{\lambda_t \sigma_y^2} [dP_t - \lambda_t \beta_t (v_t - P_t) dt], \quad \dot{\Sigma}_t = \sigma_v^2 - \frac{(\Sigma_t \beta_t)^2}{\sigma_y^2}.$$

To recover the identity $P_t = v_t$, we need to impose that $\Sigma_t \beta_t = \lambda_t \sigma_y^2$. This equality together with the border condition $v_0 = P_0$ imply that $v_t = P_t$ for all $t > 0$. This equality also implies that $(\Sigma_t \beta_t)^2 = \lambda_t^2 \sigma_y^4$. Therefore, the market maker's filtering conditions are

$$\Sigma_t \beta_t = \lambda_t \sigma_y^2, \quad \dot{\Sigma}_t = \sigma_v^2(t) - \sigma_y^2 \lambda_t^2,$$

which guarantee that the market maker equilibrium condition $P_t = \mathbb{E}[V_t | \mathcal{F}_t^M]$ is satisfied. Since $\lambda_t = \lambda_0 e^{-\mu t}$, it follows that

$$\Sigma_t = \Sigma_0 + \Gamma_0 - \Gamma_t - \frac{\sigma_y^2 \lambda_0^2}{2\mu} (1 - e^{-\mu t}), \quad \beta_t = \frac{\sigma_y^2 \lambda_0 e^{-\mu t}}{\Sigma_t}.$$

To complete the proof, we need to specify the values of the two constant λ_0 and C and verify that the proposed value function $\Pi(t, M) = \alpha_t M^2 + \gamma_t$ and trading strategy β_t effectively solve the insider's problem. This final step is achieved by imposing the transversality condition $\lim_{t \rightarrow \infty} e^{-\mu t} \mathbb{E}[\Pi(t, M_t)] = 0$ for β_t .

To avoid confusion, we now use β_t to denote the trading strategy in equation (10) and $\tilde{\beta}_t$ to denote an arbitrary policy in \mathcal{B} . To explicate the dependence of M_t on a trading strategy $\{\tilde{\beta}_s : 0 \leq s \leq t\}$, we will use the notation $M_t(\tilde{\beta})$.

Since $\Pi(t, M) = \alpha_t M^2 + \gamma_t$ satisfies the HJB equation for any strategy $\tilde{\beta} \in \mathcal{B}$, it follows that

$$(22) \quad \begin{aligned} \Pi(0, M_0) &= \mathbb{E}\left[\int_0^t e^{-\mu s} \tilde{\beta}_s M_s^2(\tilde{\beta}) ds + e^{-\mu t} \Pi(t, M_t(\tilde{\beta}))\right] \\ &\geq \mathbb{E}\left[\int_0^t e^{-\mu s} \tilde{\beta}_s M_s^2(\tilde{\beta}) ds\right]. \end{aligned}$$

In addition,

$$e^{-\mu t} \mathbb{E}[\Pi(t, M_t(\beta_t))] = \frac{1}{2\lambda_0} \left(\mathbb{E}[M_t^2(\beta)] + C + \Gamma_t + \frac{\sigma_y^2 \lambda_0^2 e^{-2\mu t}}{2\mu} \right).$$

Thus, the transversality condition holds only if $C = 0$ and $\lim_{t \rightarrow \infty} \mathbb{E}[M_t^2(\beta)] = 0$. We can show that

$$\begin{aligned} \mathbb{E}[M_t^2(\beta)] &= M_0^2 e^{-2\lambda_0 \int_0^t e^{-\mu s} \beta_s ds} \\ &\quad + \int_0^t e^{-2\lambda_0 \int_s^t e^{-\mu u} \beta_u du} (\sigma_v^2(s) + \sigma_y^2 \lambda_0^2 e^{-2\mu s}) ds. \end{aligned}$$

Hence, $\lim_{t \rightarrow \infty} \mathbb{E}[M_t^2(\beta)] = 0$ if $\int_0^\infty e^{-\mu s} \beta_s ds = \infty$. This last requirement together with the fact that $\beta_t = \sigma_y^2 \lambda_0 e^{-\mu t} / \Sigma_t$ imply that $\lim_{t \rightarrow \infty} \Sigma_t = 0$. Therefore,

$$\lambda_0 = \sqrt{\frac{2\mu(\Sigma_0 + \Gamma_0)}{\sigma_y^2}}.$$

With these choices of λ_0 and C , the transversality condition is satisfied for β_t , and taking limits in equation (22), we get that

$$\begin{aligned}\Pi(0, M_0) &= \mathbb{E} \left[\int_0^\infty e^{-\mu s} \beta_s M_s^2(\beta) ds \right] \\ &\geq \mathbb{E} \left[\int_0^\infty e^{-\mu s} \tilde{\beta}_s M_s^2(\tilde{\beta}) ds \right] \quad \text{for all } \tilde{\beta} \in \mathcal{B}.\end{aligned}$$

In the last step, we used (9) and $\beta_t > 0$ for all t , and invoked the Lebesgue convergence theorem to interchange limits and expectations.

We conclude the proof by showing that the equilibrium strategy satisfies condition (9). Given the expression for $\mathbb{E}[M_t]$ above, this condition is equivalent to

$$M_0^2 \int_0^\infty \dot{f}_t e^{-f_t} dt + \int_0^\infty f_t e^{-f_t} \left[\int_0^t (\sigma_v^2(s) + \sigma_y^2 \lambda_0^2 e^{-2\mu s}) e^{f_s} ds \right] dt < \infty,$$

where $f_t := 2\lambda_0 \int_0^t e^{-\mu s} \beta_s ds$ and \dot{f}_t is its first derivative with respect to t . Note that the first integral is equal to 1. Using the Fubini theorem to reverse the order of integration, the second integral is equal to

$$\begin{aligned}& \int_0^\infty (\sigma_v^2(s) + \sigma_y^2 \lambda_0^2 e^{-2\mu s}) e^{f_s} \left[\int_s^\infty f_t e^{-f_t} dt \right] ds \\ &= \int_0^\infty (\sigma_v^2(s) + \sigma_y^2 \lambda_0^2 e^{-2\mu s}) ds = \Gamma_0 + \sigma_y^2 \lambda_0^2 / (2\mu) < \infty. \quad Q.E.D.\end{aligned}$$

REFERENCES

- AMIHUD, Y., H. MENDELSON, AND L. PEDERSEN (2006): "Liquidity and Asset Prices," *Foundations and Trends in Finance*, 1, 269–364. [248]
- BACK, K. (1992): "Insider Trading in Continuous Time," *Review of Financial Studies*, 5, 387–409. [246,259]
- BACK, K., AND S. BARUCH (2004): "Information in Securities Markets: Kyle Meets Glosten and Milgrom," *Econometrica*, 72, 433–465. [246]
- BACK, K., AND H. PEDERSEN (1998): "Long-Lived Information and Intraday Patterns," *Journal of Financial Markets*, 1, 385–402. [248]
- BAGEHOT, W. (1971): "The Only Game in Town," *Financial Analyst Journal*, 22, 12–14. [245]
- BIAIS, B., L. GLOSTEN, AND C. SPATT (2005): "Market Microstructure: A Survey of Microfoundations, Empirical Results and Policy Implications," *Journal of Financial Markets*, 8, 217–264. [248]
- BRUNNERMEIER, M. (2001): *Asset Pricing Under Asymmetric Information*. New York: Oxford University Press. [248]
- CHAU, M., AND D. VAYANOS (2008): "Strong-Form Efficiency With Monopolistic Insiders," *The Review of Financial Studies*, 21, 2275–2306. [247,248]
- FOSTER, F., AND S. VISWANATHAN (1996): "Strategic Trading When Agents Forecast the Forecasts of Others," *The Journal of Finance*, 51, 1437–1478. [248]

- GLOSTEN, L., AND P. MILGROM (1985): "Bid, Ask and Transaction Prices in a Specialist Market With Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71–100. [245, 246]
- HOLDEN, G., AND A. SUBRAHMANYAM (1992): "Long-Lived Private Information and Imperfect Competition," *The Journal of Finance*, 47, 247–270. [248]
- KNOPP, K. (1990): *Theory and Applications of Infinite Series*. New York: Dover. [271]
- KYLE, A. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53, 1315–1335. [245, 246, 250, 257, 258]
- MENDELSON, H., AND T. TUNCA (2004): "Strategic Trading, Liquidity, and Information Acquisition," *Review of Financial Studies*, 17, 295–337. [248]
- O'HARA, M. (1997): *Market Microstructure Theory*. MA: Blackwell Publishing. [248]
- PROKHOROV, Y. (1956): "Convergence of Random Processes and Limit Theorems in Probability Theory," *Theory of Probability and It's Applications*, 1, 157–214. [274]
- SPIEGEL, M., AND A. SUBRAHMANYAM (1992): "Informed Speculation and Hedging in a Non-competitive Securities Market," *Review of Financial Studies*, 5, 307–329. [248]

*Stern School of Business, New York University, 44 West Fourth Street, Suite 8-77,
New York, NY 10012, U.S.A.; rcaldent@stern.nyu.edu*

and

*Dept. of Economics, New York University, 19 West Fourth Street, 6th Floor, New
York, NY 10012, U.S.A.; ennio@nyu.edu.*

Manuscript received April, 2008; final revision received July, 2009.

MEDIATED PARTNERSHIPS

BY DAVID RAHMAN AND ICHIRO OBARA¹

This paper studies partnerships that employ a mediator to improve their contractual ability. Intuitively, profitable deviations must be attributable, that is, there must be some group behavior such that an individual can be statistically identified as innocent, to provide incentives in partnerships. Mediated partnerships add value by effectively using different behavior to attribute different deviations. As a result, mediated partnerships are necessary to provide the right incentives in a wide range of economic environments.

KEYWORDS: Mediated contracts, partnerships, private monitoring.

1. INTRODUCTION

PROVIDING INCENTIVES IN PARTNERSHIPS is a classic topic of economic theory.² Although it is well known that communication is a basic facet of incentive provision (Aumann (1974), Forges (1986), Myerson (1986)), this insight has not been systematically applied to partnership problems. This paper adds to the literature by asking the following question. Consider a group of individuals whose behavior is subject to moral hazard, but who have rich communication and contractual protocols: (i) a disinterested mediator who can make confidential, verifiable but nonbinding recommendations to agents, and (ii) budget-balanced payment schemes³ that may depend on both the mediator's recommendations and individual reports. What outcomes can this group enforce?

Our main result (Theorem 1) shows that identifying obedient agents (IOA) is both necessary and sufficient for every outcome to be virtually enforceable⁴ in this mediated environment, regardless of preferences. IOA means that for any profile of deviations, there is some behavior by the agents that statistically identifies an innocent individual after any unilateral deviation in the profile. IOA enjoys the following crucial property: different behavior may be used to attribute innocence after different deviations.

Let us intuitively explain this result. On the one hand, providing incentives with budget balance requires punishing some agents and rewarding others si-

¹Many thanks are owed to Harold Demsetz, Larry Jones, Michihiro Kandori, Narayana Kocherlakota, David Levine, Roger Myerson, Itai Sher, Joe Ostroy, Phil Reny, Joel Sobel, Bill Zame, a co-editor, and four anonymous referees for help with previous drafts. We are also grateful to numerous seminar audiences. D. Rahman gratefully acknowledges financial support from the Spanish Ministry of Education Grant SEJ 2004-07861 while at Universidad Carlos III de Madrid and the National Science Foundation Grant SES 09-22253.

²See Alchian and Demsetz (1972), Holmström (1982), Radner, Myerson, and Maskin (1986), Legros and Matsushima (1991), Legros and Matthews (1993), d'Aspremont and Gérard-Varet (1998), and others.

³Budget balance means that the sum of payments across individuals always equals zero.

⁴An outcome is “virtually enforceable” if there is an enforceable outcome arbitrarily close to it.

multaneously. If, after a unilateral deviation, an innocent party cannot be identified, then the deviator could have been anyone, so the only way to discourage the deviation is to punish everyone. However, this violates budget balance. On the other hand, IOA implies that budget-balanced incentives can be provided by rewarding the innocent and punishing all others. To prove this, we establish and take advantage of the following observation. Rich contractual protocols enable the use of payments that differ after different recommended actions. We show that effectively, to reward the innocent after a given deviation profile, different behavior may be used to find such innocent parties. But this is just the definition of IOA. Without rich contractual protocols, the same payments must be made after every recommendation, and we show that as a result, the same behavior must be used to identify the innocent.

The value of mediated partnerships over ordinary ones (Theorems 2 and 4) now follows. Without payment schemes contingent on recommendations, it is possible to provide incentives by rewarding the innocent only if the same behavior is used to attribute innocence after every deviation. The difference between this requirement and the clearly less stringent IOA characterizes the value of mediated partnerships. As it turns out, mediated partnerships provide incentives in many natural environments where incentives would otherwise fail. For instance, for generic distributions of output, mediated partnerships can provide incentives⁵ even without production complementarities,⁶ yet ordinary ones cannot (Example 1).⁷

This paper adds to the literature (Section 6) in two basic ways. First, it extends the work of Legros and Matthews (1993), who derived nearly efficient partnerships in restricted environments with output-contingent contracts. Although they noted that identifying the innocent is important for budget-balanced incentives, they did not address statistical identification and did not use different behavior to identify the innocent after different deviations. Second, being necessary for Theorem 1, IOA exhausts the informational economies from identifying the innocent rather than the guilty.⁸ This contrasts with the literature on repeated games, where restricted communication protocols were used by Kandori (2003) and others to prove the Folk theorem.⁹ Such papers typically require a version of pairwise full rank (Fudenberg, Levine, and Maskin (1994)), which intuitively means identifying the deviator after every

⁵See the working paper version of this paper for a proof of genericity.

⁶See Legros and Matthews (1993, Example B) to enforce partnerships with complementarities.

⁷For example, we do not require that the distribution of output has a “moving support,” that is, the support of the distribution depends on individual behavior. This assumption, made by Legros and Matthews (1993), is not generic, so an arbitrarily small change in probabilities leads to its failure.

⁸Heuristically, knowing who deviated implies knowing someone who did not deviate, but knowing someone who did not deviate does not necessarily imply knowing who did.

⁹See Section 6 for a more detailed discussion of this literature.

deviation. This is clearly more restrictive than IOA, which only requires identifying a nondeviator.

The paper is organized as follows. Section 2 presents a motivating example where a mediated partnership is virtually enforced, yet none of the papers above applies. Section 3 presents the model and main definitions. Section 4 states our main results, discussed above. Section 5 refines our main assumptions in the specific context of public monitoring and studies participation as well as liability constraints. Section 6 reviews the literature on contract theory and repeated games, and compares it to this paper. Finally, Section 7 concludes. Proofs appear in the Appendix.

2. EXAMPLE

We begin our analysis of mediated partnerships with an example to capture the intuition behind our main result, Theorem 1. The example suggests the following intuitive way to attain a “nearly efficient” partnership: appoint a *secret principal*.

EXAMPLE 1: Consider a fixed group of n individuals. Each agent i can either work ($a_i = 1$) or shirk ($a_i = 0$). Let $c > 0$ be each individual’s cost of effort. Effort is not observable. Output is publicly verifiable and can be either good (g) or bad (b). The probability of g equals $P(\sum_i a_i)$, where P is a strictly increasing function of the sum of efforts. Finally, assume that each individual i ’s utility function equals $z_i - ca_i$, where z_i is the amount of money received by i .

Radner, Myerson, and Maskin (1986) introduced this partnership in the context of repeated games. They considered the problem of providing incentives for everyone to work—if not all the time, at least most of the time—without needing to inject or withdraw resources from the group as a whole. They effectively showed that in this environment there do not exist output-contingent rewards that both (i) balance the group’s budget, that is, the sum of individual payments always equals zero, and (ii) induce everyone to work most of the time, let alone all of the time. Indeed, for everyone to work at all, they must be rewarded when output is good. However, this arrangement violates budget balance, since everyone being rewarded when output is good clearly implies that the sum of payments across agents is greater when output is good than when it is bad.

An arrangement that still does not solve the partnership problem, but nevertheless induces most people to work, is appointing an agent to play the role of Holmström’s principal. Call this agent 1 and define output-contingent payments to individuals as follows. For $i = 2, \dots, n$, let $\zeta_i(g) = \bar{z}$ and $\zeta_i(b) = 0$ be

agent i 's output-contingent money payment for some $\bar{z} \geq 0$. To satisfy budget balance, agent 1's transfer equals

$$\zeta_1 = -\sum_{i=2}^n \zeta_i.$$

By construction, the budget is balanced. It is easy to see that everyone but agent 1 will work if \bar{z} is sufficiently large. However, agent 1 has the incentive to shirk.¹⁰

With mediated contracts, it is possible to induce everyone to work most of the time. Indeed, consider the following incentive scheme. For any small $\varepsilon > 0$, a mediator or machine asks every individual to work (call this event $\mathbf{1}$) with probability $1 - \varepsilon$. With probability ε/n , agent i is picked (assume everyone is picked with equal probability) and secretly asked to shirk, while all others are asked to work (call this event $\mathbf{1}_{-i}$). For $i = 1, \dots, n$, let $\zeta_i(g|\mathbf{1}) = \zeta_i(b|\mathbf{1}) = 0$ be agent i 's contingent transfer if the mediator asked everyone to work. Otherwise, if agent i was secretly asked to shirk, for $j \neq i$, let $\zeta_j(g|\mathbf{1}_{-i}) = \bar{z}$ and $\zeta_j(b|\mathbf{1}_{-i}) = 0$ be agent j 's transfer. For agent i , let

$$\zeta_i(\mathbf{1}_{-i}) = -\sum_{j \neq i} \zeta_j(\mathbf{1}_{-i}).$$

By construction, this contract is budget-balanced. It is also incentive compatible. Indeed, it is clear that asking an agent to shirk is always incentive compatible. If agent i is recommended to work, incentive compatibility requires that

$$\frac{\varepsilon(n-1)}{n} P(n-1)\bar{z} - \left[\frac{\varepsilon(n-1)}{n} + (1-\varepsilon) \right] c \geq \frac{\varepsilon(n-1)}{n} P(n-2)\bar{z},$$

which is satisfied if \bar{z} is sufficiently large because P is strictly increasing.¹¹ Under this contract, everyone works with probability $1 - \varepsilon$, for any $\varepsilon > 0$, by choosing \bar{z} appropriately, so everyone working is approximated with budget-balanced transfers.

The arrangement above solves the partnership problem of Radner, Myerson, and Maskin (1986) by occasionally appointing a secret principal. To induce everyone to work, this contract effectively appoints a different principal for different workers. Appointing the principals secretly allows for them to be used simultaneously. Finally, they are chosen only seldom to reduce the inherent loss from having a principal in the first place.

¹⁰This contract follows Holmström's suggestion to the letter: agent 1 is a "fixed" principal who absorbs the incentive payments to all others by "breaking" everyone else's budget constraint.

¹¹Here, $\frac{\varepsilon(n-1)}{n} + (1-\varepsilon)$ is the probability that an agent is asked to work and $\frac{\varepsilon(n-1)}{n}$ is the probability that, in addition, someone else was appointed the secret principal.

Example 1 reveals the logic behind our main result, Theorem 1. If a worker deviates (i.e., shirks), then he will decrease the probability of g not only when everyone else is asked to work, but also when a principal is appointed. In this latter case, innocence can be attributed to the principal, so the deviator can be punished by having every worker pay the principal. In other words, for each worker and any deviation by the worker there is a profile of actions by others such that his deviation can be statistically distinguished from someone else's (in this case, a principal, since the principal's deviation would raise the probability of g). This turns out to be not only necessary but also sufficient for solving any partnership problem.

3. MODEL

This section develops our model of mediated partnerships. It describes the environment, the timing of agents' interaction, notions of enforcement, and attribution.

Let $I = \{1, \dots, n\}$ be a finite set of agents, let A_i be a finite set of actions available to any agent $i \in I$, and let $A = \prod_i A_i$ be the (nonempty) space of action profiles. Write $v: I \times A \rightarrow \mathbb{R}$ for the profile of agents' utility functions, where $v_i(a)$ denotes the utility to any agent $i \in I$ from any action profile $a \in A$. A *correlated strategy* is any probability measure $\sigma \in \Delta(A)$.¹² Let S_i be a finite set of *private signals* observable only by agent $i \in I$ and let S_0 be a finite set of *publicly verifiable* signals. Let $S := \prod_{j=0}^n S_j$ be the (nonempty) space of all signal profiles. A *monitoring technology* is a measure-valued map $\Pr: A \rightarrow \Delta(S)$, where $\Pr(s|a)$ denotes the conditional probability that signal profile s was observed given that action profile a was played.

We model rich communication protocols by introducing a disinterested *mediator* who fulfills two roles: (i) making confidential recommendations to agents over what action to take and (ii) revealing the entire profile of recommendations publicly at the end of the game. This mediator may be seen as a proxy for any preplay communication among the players (Aumann (1987)).

Incentives are provided to agents with linear transfers. An *incentive scheme* is any map $\zeta: I \times A \times S \rightarrow \mathbb{R}$ that assigns monetary payments contingent on individuals, recommended actions, and reported signals, all of which are assumed verifiable.

DEFINITION 1: A *contract* is any pair (σ, ζ) , where σ is a correlated strategy and ζ is an incentive scheme. It is called *standard* if $\zeta_i(a, s)$ is not a function of a , that is, payments do not depend on recommendations; otherwise, the contract is called *mediated*.

¹²If X is a finite set, $\Delta(X) = \{\mu \in \mathbb{R}_+^X : \sum_x \mu(x) = 1\}$ is the set of probability vectors on X .

Standard contracts are a special case of mediated ones, but not otherwise. For instance, the secret principal of Section 2 is a nonstandard mediated contract, since payments depend on recommendations. The literature has mostly focused on standard contracts to study incentives, whereas this paper concentrates on mediated ones.

It is important to emphasize that a standard contract does not do away with the mediator altogether—only as regards payments. Indeed, as will be seen below and was suggested in Example 1 above, we emphasize using the mediator not so much to correlate behavior, but rather to correlate payoffs so as to provide incentives.

The timing of agents' interaction unfolds as follows. First, agents agree on some contract (σ, ζ) . A profile of recommendations is drawn according to σ and made to agents confidentially by some mediator. Agents then simultaneously take some action, which is neither verifiable nor directly observable. Next, agents observe unverifiable private signals and submit a verifiable report of their observations before observing the public signal (the timing of signals is not essential, just simplifying). Finally, recommendation- and report-contingent transfers are made according to ζ .

Specifically, we assume that agents report their private signals simultaneously, and consider contracts where agents willingly behave honestly (report truthfully) and obediently (follow recommendations). In other words, strategic behavior is assumed to constitute a *communication equilibrium*, as in Myerson (1986) and Forges (1986), of the game induced by a given contract (σ, ζ) .

If every agent is honest and obedient, agent i 's expected utility from (σ, ζ) is

$$\sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a) \zeta_i(a, s) \Pr(s|a).$$

Of course, agent i may disobey his recommendation a_i to play some other action b_i and lie about his privately observed signal. A *reporting strategy* is a map $\rho_i: S_i \rightarrow S_i$, where $\rho_i(s_i)$ is the reported signal when i privately observes s_i . For instance, the truthful reporting strategy is the identity map $\tau_i: S_i \rightarrow S_i$ with $\tau_i(s_i) = s_i$. Let R_i be the set of all reporting strategies for agent i . For every agent i and every pair $(b_i, \rho_i) \in A_i \times R_i$, the conditional probability that $s \in S$ will be reported when everyone else is honest and plays $a_{-i} \in A_{-i}$ equals¹³

$$\Pr(s|a_{-i}, b_i, \rho_i) := \sum_{t_i \in \rho_i^{-1}(s_i)} \Pr(s_{-i}, t_i | a_{-i}, b_i).$$

A contract (σ, ζ) is *incentive compatible* if obeying recommendations and reporting honestly is optimal for every agent when everyone else is honest and

¹³We use the notation $s = (s_{-i}, s_i)$ for $s_i \in S_i$ and $s_{-i} \in S_{-i} = \prod_{j \neq i} S_j$; similarly for $a = (a_{-i}, a_i)$.

obedient, that is, $\forall i \in I, a_i \in A_i, (b_i, \rho_i) \in A_i \times R_i$,

$$(*) \quad \sum_{a_{-i}} \sigma(a)(v_i(a_{-i}, b_i) - v_i(a)) \\ \leq \sum_{(a_{-i}, s)} \sigma(a)\zeta_i(a, s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)).$$

The left-hand side of $(*)$ reflects the *utility gain*¹⁴ for an agent i from playing b_i when asked to play a_i . The right-hand side reflects his *monetary loss* from playing (b_i, ρ_i) relative to honesty and obedience. Such a loss originates from two sources. On the one hand, playing b_i instead of a_i may change conditional probabilities over signals. On the other, reporting according to ρ_i may affect conditional payments.

DEFINITION 2: A correlated strategy σ is *exactly enforceable* (or simply *enforceable*) if there is an incentive scheme $\zeta : I \times A \times S \rightarrow \mathbb{R}$ to satisfy $(*)$ for all (i, a_i, b_i, ρ_i) and

$$(**) \quad \forall (a, s), \quad \sum_{i \in I} \zeta_i(a, s) = 0.$$

Call σ *virtually enforceable* if there exists a sequence $\{\sigma^m\}$ of enforceable correlated strategies such that $\sigma^m \rightarrow \sigma$.

A correlated strategy is enforceable if there is a budget-balanced¹⁵ incentive scheme that makes it incentive compatible. It is virtually enforceable if it is the limit of enforceable correlated strategies. This requires budget balance along the way, not just asymptotically. For instance, in Example 1, everybody working is virtually enforceable, but not exactly enforceable.

We end this section by defining a key condition called identifying obedient players, which will be shown to characterize enforcement. We begin with some preliminaries.

A *strategy* for any agent i is a map $\alpha_i : A_i \rightarrow \Delta(A_i \times R_i)$, where $\alpha_i(b_i, \rho_i | a_i)$ stands for the probability that i reacts by playing (b_i, ρ_i) when recommended to play a_i . For any σ and any α_i , let $\Pr(\sigma, \alpha_i) \in \Delta(S)$, defined pointwise by

$$\Pr(s|\sigma, \alpha_i) = \sum_{a \in A} \sigma(a) \sum_{(b_i, \rho_i)} \Pr(s|a_{-i}, b_i, \rho_i) \alpha_i(b_i, \rho_i | a_i),$$

be the vector of report probabilities if agent i deviates from σ according to α_i .

¹⁴Specifically, utility gain is probability-weighted, weighted by $\sigma(a_i) = \sum_{a_{-i}} \sigma(a)$, the probability of a_i .

¹⁵Budget balance means here that the sum of payments across individuals always equals zero. Some authors use budget balance to mean that payments add up to the value of some output. On the other hand, our model may be interpreted as using utilities that are net of profit shares.

DEFINITION 3: A strategy profile $\alpha = (\alpha_1, \dots, \alpha_n)$ is *unattributable* if

$$\forall a \in A, \quad \Pr(a, \alpha_1) = \dots = \Pr(a, \alpha_n).^{16}$$

Call α *attributable* if it is not unattributable, that is, there exist agents i and j such that $\Pr(a, \alpha_i) \neq \Pr(a, \alpha_j)$ for some $a \in A$.

Intuitively, a strategy profile α is unattributable if a unilateral deviation from honesty and obedience by any agent i to a strategy α_i in the profile would lead to the same conditional distribution over reports. Heuristically, after a deviation (from honesty and obedience) belonging to some unattributable profile, even if the fact that someone deviated was detected, anyone could have been the culprit.

Call α_i *disobedient* if $\alpha_i(b_i, \rho_i | a_i) > 0$ for some $a_i \neq b_i$, that is, it disobeys some recommendation with positive probability. A disobedient strategy may be “honest,” that is, ρ_i may equal τ_i . However, dishonesty by itself (obeying recommendations but choosing $\rho_i \neq \tau_i$) is not labeled as disobedience. A *disobedient strategy profile* is any $\alpha = (\alpha_1, \dots, \alpha_n)$ such that α_i is disobedient for at least one agent i .

DEFINITION 4: A monitoring technology *identifies obedient agents* (IOA) if every disobedient strategy profile is attributable.

IOA means that for every disobedience by some arbitrary agent i and every profile of others’ strategies, an action profile exists such that i ’s unilateral deviation has a different effect on report probabilities from at least one other agent. For instance, the monitoring technology of Example 1 identifies obedient agents. There, if a worker shirks, then good news becomes *less* likely, whereas if a principal works, then good news becomes *more* likely. Hence, a strategy profile with i disobeying is attributable by just having another agent behave differently from i . This implies IOA. Intuitively, IOA holds by using different principals for different workers.

4. RESULTS

This section presents the paper’s main results, characterizing enforceable outcomes in terms of the monitoring technology, with and without mediated contracts. We begin with a key lemma that provides a dual characterization of IOA.

LEMMA 1: *A monitoring technology identifies obedient agents if and only if there exists a function $\xi: I \times A \times S \rightarrow \mathbb{R}$ such that $\sum_i \xi_i(a, s) = 0$ for every*

¹⁶We slightly abuse notation by identifying action profiles with pure correlated strategies.

(a, s) and

$$\forall(i, a_i, b_i, \rho_i), \quad 0 \leq \sum_{(a_{-i}, s)} \xi_i(a, s) (\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a))$$

with a strict inequality whenever $a_i \neq b_i$.

Intuitively, Lemma 1 shows that IOA is equivalent to the existence of budget-balanced “probability-weighted” transfers ξ such that (i) the budget is balanced, (ii) no deviation is profitable, and (iii) every disobedience incurs a strictly positive monetary cost. If every action profile is recommended with positive probability, that is, if $\sigma \in \Delta^0(A) := \{\sigma \in \Delta(A) : \sigma(a) > 0 \ \forall a \in A\}$ is any completely mixed correlated strategy, then there is an incentive scheme ζ with $\xi_i(a, s) = \sigma(a)\zeta_i(a, s)$ for all (i, a, s) . Therefore, IOA implies that given $\sigma \in \Delta^0(A)$ and ξ satisfying Lemma 1, for any profile v of agents’ utility functions, we may scale ζ appropriately to overcome all incentive constraints simultaneously. Hence, the second half of Lemma 1 implies that every completely mixed correlated strategy is exactly enforceable, regardless of the utility profile. Approximating each correlated strategy with completely mixed ones establishes half of our main result, Theorem 1 below. The other half argues that if IOA fails, then there exist a profile of utility functions and a correlated strategy that is not virtually enforceable. In this sense, IOA is the weakest condition on a monitoring technology that—Independently of preferences—guarantees virtual enforcement.

THEOREM 1: *A monitoring technology identifies obedient agents if and only if for any profile of utility functions, every correlated strategy is virtually enforceable.*

Theorem 1 characterizes monitoring technologies such that “everything” is virtually enforceable, regardless of preferences. It says that identifying obedient agents in a weak sense is not only necessary, but also sufficient for virtual enforcement. Intuitively, if, after a disobedience, some innocent agent can be statistically identified then that agent can be rewarded at the expense of everyone else, thereby punishing the deviator. Heuristically, if a strategy profile can be attributed, then there is an incentive scheme that discourages every strategy in that profile. Theorem 1 says that for every disobedient strategy profile there is a scheme that discourages it if and only if there is a scheme that discourages all disobedient strategy profiles simultaneously.

To put Theorem 1 in perspective, consider the scope of enforcement with standard contracts. By Example 1, IOA is generally not enough for enforcement with standard contracts, but the following strengthening is. Given a subset $B \subset A$ of action profiles and an agent i , let $B_i := \{b_i \in A_i : \exists b_{-i} \in A_{-i} \text{ s.t. } b \in B\}$ be the projection of B on A_i . Call a strategy α_i *B-disobedient* if it is disobedient at some $a_i \in B_i$, that is, if $\alpha_i(b_i, \rho_i | a_i) > 0$ for some $b_i \neq a_i \in B_i$. A *B-disobedient strategy profile* is any $\alpha = (\alpha_1, \dots, \alpha_n)$ such that α_i is *B-disobedient* for some agent i . Given $\sigma \in \Delta(A)$, α is *attributable at* σ if there exist agents i

and j such that $\Pr(\sigma, \alpha_i) \neq \Pr(\sigma, \alpha_j)$, and say \Pr identifies obedient agents at σ (IOA- σ) if every supp σ -disobedient¹⁶ strategy profile is attributable at σ . Intuitively, IOA- σ differs from IOA in that IOA allows for different α 's to be attributed at different σ 's, whereas IOA- σ does not.

THEOREM 2: *A monitoring technology identifies obedient agents at σ if and only if for any profile of utility functions, σ is exactly enforceable with a standard contract.*

Theorem 2 characterizes enforceability with standard contracts of any correlated strategy σ in terms of IOA- σ . Intuitively, it says that enforcement with standard contracts requires that every α be attributable at the same σ .¹⁷ Theorem 2 also sheds light onto the value of mediated contracts. Indeed, the proof of Theorem 1 shows that enforcing a completely mixed correlated strategy (i.e., such that $\sigma(a) > 0$ for all a) only requires IOA, by allowing for different strategy profiles to be attributable at different action profiles. This condition is clearly weaker than IOA- σ . On the other hand, IOA is generally not enough to enforce a given pure-strategy profile a , as Example 1 shows with $a = \mathbf{1}$ there. Since agents receive only one recommendation under a , there is no use for mediated contracts, so by Theorem 2, IOA- a characterizes exact enforcement of a with both standard and mediated contracts.¹⁸

Now consider the intermediate case where σ has arbitrary support. Fix a subset of action profiles $B \subset A$. A strategy profile $\alpha = (\alpha_1, \dots, \alpha_n)$ is *B-attributable* if there exist agents i and j such that $\Pr(a, \alpha_i) \neq \Pr(a, \alpha_j)$ for some $a \in B$. Otherwise, α is called *B-unattributable*. For instance, *A*-attribution is just attribution. Say \Pr *B*-identifies obedient agents (*B*-IOA) if every *B*-disobedient strategy profile is *B*-attributable. For instance, *A*-IOA is just IOA and $\{a\}$ -IOA equals IOA- a .

THEOREM 3: *For any subset $B \subset A$, the following statements are equivalent:*

- (i) *The monitoring technology B -identifies obedient agents.*
- (ii) *Every correlated strategy with support equal to B is enforceable for any profile of utility functions.*
- (iii) *Some fixed correlated strategy with support equal to B is enforceable for any profile of utility functions.*

Theorem 3 characterizes enforcement with mediated contracts of any correlated strategy σ with supp σ -IOA. Hence, only the support of a correlated strategy matters for its enforcement for all preferences. Moreover, any other correlated strategy with support contained in supp σ becomes virtually en-

¹⁶By definition, supp $\sigma = \{a \in A : \sigma(a) > 0\}$ is the support of σ .

¹⁷Even for virtual enforcement with standard contracts, the same σ must attribute all α 's. For example, in Example 1 there is no sequence $\{\sigma^m\}$ with $\sigma^m \rightarrow \mathbf{1}$ and \Pr satisfying IOA- σ^m for all m .

¹⁸Again, we abuse notation by labeling a as both an action profile and a pure correlated strategy.

forceable, just as with Theorem 1. Intuitively, mediated contracts allow for different actions in the support of a correlated strategy to attribute different strategy profiles, unlike standard contracts, as shown above. Therefore, clearly IOA- σ is more restrictive than $\text{supp } \sigma$ -IOA.

Although the results above focused on enforcement for all utility profiles, restricting attention to fixed preferences does not introduce additional complications and yields similar results. Indeed, fix a profile $v: I \times A \rightarrow \mathbb{R}$ of utility functions. A natural weakening of IOA involves allowing unprofitable strategy profiles to be unattributable. A strategy profile α is called σ -*profitable* if

$$\sum_{(i,a,b_i,\rho_i)} \sigma(a) \alpha_i(b_i, \rho_i | a_i) (v_i(a_{-i}, b_i) - v_i(a)) > 0.$$

Intuitively, the profile α is σ -profitable if the sum of each agent's utility gains from a unilateral deviation in the profile is positive. Enforcement now amounts to the following declarations.

THEOREM 4: (i) *Every σ -profitable strategy profile is $\text{supp } \sigma$ -attributable if and only if σ is enforceable.* (ii) *Every σ -profitable strategy profile is attributable at σ if and only if σ is enforceable with a standard contract.*

Theorem 4 characterizes enforceability with and without mediated contracts. It describes how mediated contracts add value by relaxing the burden of attribution: Every profile α that is attributable at σ is $\text{supp } \sigma$ -attributable, but not conversely. For instance, in Example 1, let $\sigma(S)$ be the probability that $S \subset I$ are asked to work, and suppose that $\sigma(I) > 0$. Let α be the strategy profile where every agent i shirks with probability p_i if asked to work (and obeys if asked to shirk), with $p_i = \sigma(I)[P(n-1) - P(n)] / \sum_{S \ni i} \sigma(S)[P(|S| - 1) - P(|S|)] \in (0, 1]$. By construction, the probability of good output equals $\sigma(I)P(n-1) + \sum_{S \neq I} \sigma(S)P(|S|)$, which is independent of i . Therefore, α is not attributable at any σ with $\sigma(I) > 0$. However, α is attributable, since the monitoring technology identifies obedient agents.

5. DISCUSSION

In this section we decompose IOA to understand it better under the assumption of public monitoring. We also consider participation and liability constraints.

5.1. Public Monitoring

To help understand IOA, let us temporarily restrict attention to *publicly verifiable* monitoring technologies, that is, such that $|S_i| = 1$ for all $i \neq 0$. In this case, IOA can be naturally decomposed into two parts. We formalize this decomposition next.

A strategy α_i for any agent i is *detectable* if $\Pr(a, \alpha_i) \neq \Pr(a)$ at some $a \in A$. Say Pr *detects unilateral disobedience* (DUD) if every disobedient strategy is detectable,¹⁹ where different action profiles may be used to detect different strategies. Say *detection implies attribution* (DIA) if for every detectable strategy α_i and every strategy profile α_{-i} , $\alpha = (\alpha_{-i}, \alpha_i)$ is attributable. Intuitively, DIA says that if a strategy is detected, someone can be (statistically) ruled out as innocent.

THEOREM 5: *A publicly verifiable monitoring technology identifies obedient agents if and only if (i) it detects unilateral disobedience and (ii) detection implies attribution.*

An immediate example of DIA is Holmström's (1982) principal, that is, an individual i_0 with no actions to take or signals to observe (both A_{i_0} and S_{i_0} are singletons). The principal is automatically obedient, so every detectable strategy can be discouraged with budget balance by rewarding him and punishing everyone else. DIA isolates this idea and finds when the principal's role can be fulfilled internally. It helps to provide budget-balanced incentives by identifying innocent individuals to be rewarded and punishing all others (if necessary) when a disobedience is detected.

Next, we give a dual characterization of DIA that sheds light onto the role it plays in Theorem 1. A publicly verifiable monitoring technology Pr satisfies *incentive compatibility implies enforcement* (ICE) if for every $K: A \times S \rightarrow \mathbb{R}$, there exists $\xi: I \times A \times S \rightarrow \mathbb{R}$ such that

$$\forall(a, s), \quad \sum_{i \in I} \xi_i(a, s) = K(a, s),$$

$$\forall(i, a_i, b_i), \quad 0 \leq \sum_{(a_{-i}, s)} \xi_i(a, s) (\Pr(s|a_{-i}, b_i) - \Pr(s|a)).$$

The function $K(a, s)$ may be regarded as a budgetary surplus or deficit for each combination of recommended action and realized signal. Intuitively, ICE means that any budget can be attained by some payment scheme that avoids disrupting any incentive compatibility constraints. As it turns out, this is equivalent to DIA.

THEOREM 6: *Given any publicly verifiable monitoring technology, detection implies attribution if and only if incentive compatibility implies enforcement.*

This result helps to clarify the roles of DUD and DIA in Theorem 1. Rahman (2008) showed that DUD characterizes virtual enforcement without budget

¹⁹This condition on a monitoring technology was introduced and analyzed by Rahman (2008).

balance of any correlated strategy σ , regardless of preferences. ICE guarantees the existence of a further contract to absorb any budgetary deficit or surplus of the original contract without violating any incentive constraints. Therefore, the original contract plus this further contract can now virtually enforce σ with a balanced budget.²⁰

If the monitoring technology is not publicly verifiable, DUD plus DIA is sufficient but unnecessary for IOA. Necessity fails in general because there may exist dishonest but obedient strategies that IOA allows to remain unattributable even if detectable, as the next example shows.²¹

EXAMPLE 2: There are three agents and A_i is a singleton for every agent i , so IOA is automatically satisfied. There are no public signals and each agent observes a binary private signal: $S_i = \{0, 1\}$ for all i . The monitoring technology is

$$\Pr(s) := \begin{cases} \frac{6}{25}, & \text{if } \sum_i s_i = 3, \\ \frac{3}{25}, & \text{if } \sum_i s_i = 1 \text{ or } 2, \\ \frac{1}{25}, & \text{if } \sum_i s_i = 0. \end{cases}$$

The following equation is a profile of (trivially obedient) unattributable strategies that are also detectable, violating DIA. Suppose that agent i deviates by lying with probability $2/5$ after observing $s_i = 1$ and lying with probability $3/5$ after observing $s_i = 0$. For every agent i , the joint distribution of reported private signals becomes

$$\Pr(s) = \begin{cases} \frac{27}{125}, & \text{if } \sum_i s_i = 3, \\ \frac{18}{125}, & \text{if } \sum_i s_i = 2, \\ \frac{12}{125}, & \text{if } \sum_i s_i = 1, \\ \frac{8}{125}, & \text{if } \sum_i s_i = 0. \end{cases}$$

²⁰A comparable argument was provided by d'Aspremont, Cremer, and Gérard-Varet (2004) for Bayesian mechanisms.

²¹Without a publicly verifiable monitoring technology, IOA is equivalent to DUD plus “disobedient detection implies attribution,” that is, every disobedient and detectable strategy is attributable. However, this latter condition lacks an easily interpreted dual version as in Theorem 6.

5.2. Participation and Liability

Individual rationality—or participation—constraints are easily incorporated into the present study of incentives by imposing the family of inequalities

$$\forall i \in I, \quad \sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a) \zeta_i(a, s) \Pr(s|a) \geq 0.$$

THEOREM 7: *Participation is not a binding constraint if $\sum_i v_i(a) \geq 0$ for all $a \in A$.*

Theorem 7 generalizes standard results (e.g., d'Aspremont and Gérard-Varet (1998, Lemma 1)) to our setting.

Next, we study limited liability given $z \in \mathbb{R}_+^I$, by imposing constraints of the form $\zeta_i(a, s) \geq -z_i$. Intuitively, an agent can never pay any more than z_i . Call z_i agent i 's *liability*, and call z the *distribution of liability*. A group's *total liability* is defined by $\widehat{z} = \sum_i z_i$. Without participation constraints, Theorem 5 of Legros and Matsushima (1991) and Theorem 4 of Legros and Matthews (1993) easily generalize to this setting.

THEOREM 8: *In the absence of participation constraints, only total liability affects the set of enforceable outcomes, not the distribution of liability.*

Including participation constraints leads to the following characterization.

THEOREM 9: *The correlated strategy σ is enforceable with individual rationality and liability limited by z if and only if*

$$\begin{aligned} & \sum_{(a,i,b_i,\rho_i)} \sigma(a) \alpha_i(b_i, \rho_i | a_i) (v_i(a_{-i}, b_i) - v_i(a)) \\ & \leq \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \sum_{i \in I} z_i \end{aligned}$$

for every (α, π) such that α is a strategy profile and $\pi = (\pi_1, \dots, \pi_n) \geq 0$, where $\widehat{\eta} := \sum_{(a,s)} \min_i \{\Pr(s|a, \alpha_i) - (1 + \pi_i) \Pr(s|a)\}$ and $v_i(\sigma) = \sum_a \sigma(a) v_i(a)$.

Theorem 9 generalizes Theorems 7 and 8, as the next result shows.

COROLLARY 1: *Suppose that σ is enforceable with individual rationality and liability limited by z . (i) If $v_i(\sigma) \geq z_i$, then agent i 's participation is not a binding constraint. (ii) The distribution of liability does not matter within the subset t of agents whose participation constraint is not binding, that is, σ is also enforceable with individual rationality and liability limited by any z' with $z_j = z'_j$ for $j \in I \setminus t$ and $\sum_{i \in t} z_i = \sum_{i \in t} z'_i$.*

6. LITERATURE

To help identify this paper's contribution, let us now compare its results with the literature. Broadly, the paper contributes (i) a systematic analysis of partnerships that fully exploit internal communication, and (ii) results that show that attribution and IOA yield the weakest requirements on a monitoring technology for enforcement and virtual enforcement. IOA enjoys the key property that different action profiles can be used to attribute different disobedient strategy profiles, in contrast with the literature, which we discuss below.

In contract theory, [Legros and Matsushima \(1991\)](#) characterized exact enforcement with standard contracts and publicly verifiable signals, but they did not interpret their results in terms of attribution, nor did they consider virtual enforcement. Another related paper is [d'Aspremont and Gérard-Varet \(1998\)](#). In the same context as [Legros and Matsushima \(1991\)](#), they derived intuitive sufficient conditions for enforcement. A closer paper to ours is by [Legros and Matthews \(1993\)](#), who studied virtual enforcement with standard contracts and deterministic output. They proposed a contract that uses mixed strategies to identify nonshirkers whenever possible,²² but the same correlated strategy must identify nonshirkers after every deviation, unlike mediated contracts. Their contract fails to provide the right incentives if output given efforts is stochastic and its distribution does not have a “moving support,” that is, the support does not depend on efforts. The key difference between their contract and ours is that mediated partnerships correlate agents' payoffs not just to output, but also to others' mixed strategies. As a result, mediated partnerships can virtually enforce efficient behavior even without a moving support, as Example 1 and Theorem 1 show.²³

In the context of repeated games, the closest papers to ours may be [Kandori \(2003\)](#), [Aoyagi \(2005\)](#), and [Tomala \(2009\)](#). They establish versions of the Folk theorem by interpreting players' continuation values as linear transfers. [Kandori](#) allowed agents to play mixed strategies and reported on the realization of such mixtures after observing a public signal. He considered contracts contingent on the signals and these reports.²⁴ Although his contracts are non-standard, they fail to fully employ communication. For instance, they fail to provide incentives in Example 1. [Aoyagi](#) used dynamic mediated strategies that rely on “ ε -perfect” monitoring and fail if monitoring is costly or one-sided. Our results accommodate these issues. Finally, [Tomala](#) studied a class of recursive communication equilibria.

²²A (stronger) form of identifying nonshirkers was suggested in mechanism design by [Kosenok and Severinov \(2008\)](#). However, they characterized full surplus extraction rather than enforcement.

²³[Fudenberg, Levine, and Maskin \(1994\)](#) considered a form of virtual enforcement without a moving support. However, they required much stronger assumptions than ours, discussed momentarily.

²⁴[Obara \(2008\)](#) extended [Kandori](#)'s contracts to study full surplus extraction with moral hazard and adverse selection in the spirit of [Cremer and McLean \(1988\)](#), ignoring budget balance.

There are several differences between these papers and ours. One especially noteworthy difference is that to prove the Folk theorem they make much more restrictive assumptions than IOA, structurally similar to the *pairwise full rank* (PFR) of Fudenberg, Levine, and Maskin (1994). Intuitively, PFR-like conditions ask to identify deviators instead of just nondeviators. To see this, let us focus for simplicity on public monitoring and recall the decomposition of IOA into DUD and DIA (Theorem 5).

For every i , let C_i (called the *cone* of agent i) be the set of all $\eta \in \mathbb{R}^{A \times S}$ with,

$$\forall(a, s), \quad \eta(a, s) = \sum_{b_i \in A_i} \alpha_i(b_i|a_i) (\Pr(s|a_{-i}, b_i) - \Pr(s|a))$$

for some $\alpha_i : A_i \rightarrow \Delta(A_i)$. DIA imposes on agents' cones the restriction

$$\bigcap_{i \in I} C_i = \{\mathbf{0}\},$$

where $\mathbf{0}$ stands for the origin of $\mathbb{R}^{A \times S}$.

In other words, agents' cones do not overlap. PFR implies that for *every pair* of agents, their cones do not overlap. Intuitively, this means that upon any deviation, it is possible to identify the deviator's identity. On the other hand, DIA only requires that *all* agents' cones fail to overlap simultaneously. Thus, it is possible to provide budget-balanced incentives even if there are two agents whose cones overlap (i.e., their intersection is larger than just the origin), so PFR fails. In general, DIA does not even require that there exist two agents whose cones fail to overlap, in contrast with *local compatibility* of d'Aspremont and Gérard-Varet (1998). Figure 1 illustrates this point.²⁵

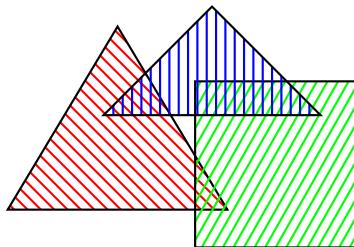


FIGURE 1.—A cross section of three nonoverlapping cones in \mathbb{R}^3 (pointed at the origin behind the page) such that every pair of cones overlaps.

²⁵Figure 1 is not pathological. Indeed, Example 1 may be viewed as a version of Figure 1.

7. CONCLUSION

Mediated partnerships embrace the idea that—as part of an economic organization—it may be beneficial for private information to be allocated differently across agents to provide the right incentives. As Example 1 illustrates, mediated partnerships can enforce outcomes that standard ones simply cannot. Indeed, mediated contracts can provide the right incentives in partnerships with stochastic output whose distribution fails to exhibit a “moving support” (i.e., the support is independent of effort), even without complementarities in production. Standard contracts cannot.

In general, mediated partnerships are enforceable if and only if it is possible to identify obedient agents. This means that after any unilateral deviation, innocence is statistically attributable to someone, although different actions may be used to attribute innocence after different deviations.²⁶ Informationally, this is clearly less costly than attempting to attribute guilt, as well as using the same actions to attribute innocence after every deviation. This latter difference exactly captures the value of mediated partnerships.

APPENDIX: PROOFS

PROOF OF LEMMA 1: With only finitely many actions and finitely many agents, the second half of the lemma holds if and only if there exists ξ such that $\sum_i \xi_i(a, s) = 0$ for all (a, s) and

$$\forall(i, a_i, b_i, \rho_i), \quad \Delta_i(a_i, b_i) \leq \sum_{(a_{-i}, s)} \xi_i(a, s) (\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)),$$

where $\Delta_i(a_i, b_i) = 1$ if $a_i \neq b_i$ and 0 otherwise. Consider the linear program that consists of choosing ξ to minimize 0 subject to the above constraints. The dual problem is to choose a vector (λ, η) such that $\lambda \geq 0$ to maximize $\sum_{(i, a_i, b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i) \Delta_i(a_i, b_i)$ subject to

$$\forall(i, a, s), \quad \sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i) (\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)) = \eta(a, s).$$

Here, the vector $\lambda \geq 0$ collects the multipliers on incentive constraints and η collects those of the budget balance constraints. Since the dual is feasible (with $(\lambda, \eta) = 0$), by the strong duality theorem (see, e.g., Schrijver (1986, p. 92)), the condition on ξ above fails if and only if there exists a dual feasible solution

²⁶Although identifying obedient agents is impossible with only two agents and public monitoring, it holds generically in richer environments, even with just three agents or a minimal amount of public information. See the working paper version of this article for these results.

(λ, η) such that $\lambda_i(a_i, b_i, \rho_i) > 0$ for some (i, a_i, b_i, ρ_i) with $a_i \neq b_i$. Let $\Lambda = \max_{(i, a_i)} \sum_{(b_i, \rho_i)} \lambda_i(a_i, b_i, \rho_i) > 0$ and define

$$\alpha_i(b_i, \rho_i | a_i) := \begin{cases} \lambda_i(a_i, b_i, \rho_i) / \Lambda, & \text{if } (b_i, \rho_i) \neq (a_i, \tau_i), \\ 1 - \sum_{(b_i, \rho_i) \neq (a_i, \tau_i)} \lambda_i(a_i, b_i, \rho_i) / \Lambda, & \text{otherwise.} \end{cases}$$

By construction, α_i is disobedient and unattributable (using α_{-i}): IOA fails. $Q.E.D.$

PROOF OF THEOREM 1: Sufficiency follows from the paragraph preceding the statement of the theorem. For necessity, suppose that IOA fails, that is, there is a disobedient profile $\alpha = (\alpha_1, \dots, \alpha_n)$ that is also unattributable. Let $a^* \in A$ be an action profile where α is disobedient, that is, there exists an agent i^* such that $\alpha_{i^*}(b_{i^*}, \rho_{i^*} | a_{i^*}^*) > 0$ for some $b_{i^*} \neq a_{i^*}^*$. Let $v_i(a) = 0$ for all (i, a) except for $v_{i^*}(b_{i^*}, a_{-i^*}^*) = 1$. Consider any correlated strategy σ that places positive probability on a^* . For a contradiction, assume that there is a payment scheme ζ that enforces σ . Summing the incentive constraints at a^* across agents and using budget balance together with the definition of v , we obtain

$$0 < \sigma(a^*) = \sum_{(i, b_i, a_{-i})} \sigma(a_i^*, a_{-i})(v_i(b_i, a_{-i}) - v_i(a_i^*, a_{-i})) \leq 0.$$

Therefore, σ is not enforceable. Finally, this implies that a^* is not virtually enforceable. $Q.E.D.$

PROOF OF THEOREM 2: The proof follows that of Lemma 1. By the strong duality theorem, \Pr satisfies $\text{IOA-}\sigma$ if and only if there exists a payment scheme $\zeta: I \times S \rightarrow \mathbb{R}$ that only depends on reported signals for each agent such that $\sum_i \zeta_i(s) = 0$ for all s and

$$\begin{aligned} \forall i \in I, a_i \in B_i, (b_i, \rho_i) \in A_i \times R_i, \\ 0 \leq \sum_{(a_{-i}, s)} \sigma(a) \zeta_i(s) (\Pr(s | a_{-i}, b_i, \rho_i) - \Pr(s | a)), \end{aligned}$$

with a strict inequality if $a_i \neq b_i$, where $B_i = \{a_i \in A_i : \exists a_{-i} \text{ s.t. } \sigma(a) > 0\}$. Call this dual condition $\text{IOA}^*\text{-}\sigma$. By scaling ζ as necessary, $\text{IOA}^*\text{-}\sigma$ clearly implies that any deviation gains can be outweighed by monetary losses. Conversely, if $\text{IOA-}\sigma$ fails, then there is a profile of deviation plans α such that $\Pr(\sigma, \alpha_i) = \Pr(\sigma, \alpha_j)$ for all (i, j) and there is an agent i^* such that α_{i^*} satisfies $\alpha_{i^*}(b_{i^*}, \rho_{i^*} | a_{i^*}^*) > 0$ for some $a_{i^*} \in B_{i^*}$ and $b_{i^*} \neq a_{i^*}^*$. For all a_{-i^*} , let $0 = v_{i^*}(a) < v_{i^*}(a_{-i^*}, b_{i^*}) = 1$ and $v_j(a) = v_j(a_{-i^*}, b_{i^*}) = 0$ for all $j \neq i^*$. Now σ

cannot be enforced by any $\zeta: I \times S \rightarrow \mathbb{R}$ such that $\sum_i \zeta_i(s) = 0$ for all s , since $\sum_{(i, b_i, \rho_i)} \alpha_i(b_i, \rho_i | a_i) \sum_{a_{-i}} \sigma(a)(v_i(a_{-i}, b_i) - v_i(a)) > \sum_{(i, s)} \zeta_i(s)(\Pr(s|\sigma, \alpha_i) - \Pr(s|\sigma)) = 0$, being a nonnegative linear combination of incentive constraints, will violate at least one. *Q.E.D.*

PROOF OF THEOREM 3: (i) \Leftrightarrow (ii) follows by applying a version of the proof of Lemma 1 and Theorem 1 after replacing B with A . (i) \Leftrightarrow (iii) follows similarly, after fixing any correlated strategy σ with support equal to B . *Q.E.D.*

PROOF OF THEOREM 4: (i) follows by applying the proof of Lemma 1 with both σ and v fixed to the incentive compatibility constraints (*). (ii) follows by a similar version of the proof of Theorem 2, again with both σ and v fixed. *Q.E.D.*

PROOF OF THEOREM 5: IOA clearly implies DUD (just replace α_{-i} with honesty and obedience for every α_i in the definition of attribution). By IOA, if a profile α is unattributable, then it is obedient, hence every deviation plan in the profile is undetectable (since the monitoring technology is publicly verifiable) and DIA follows. Conversely, DIA implies that every unattributable α_i is undetectable, and by DUD, every undetectable α_i is obedient. *Q.E.D.*

PROOF OF THEOREM 6: Consider the following primal problem: Find a feasible ξ to solve,

$$\forall(i, a_i, b_i), \quad 0 \leq \sum_{(a_{-i}, s)} \xi_i(a, s)(\Pr(s|a_{-i}, b_i) - \Pr(s|a)),$$

and

$$\forall(a, s), \quad \sum_{i \in I} \xi_i(a, s) = K(a, s).$$

The dual of this problem is given by

$$\begin{aligned} \inf_{\lambda \geq 0, \eta} & \sum_{(a, s)} \eta(a, s)K(a, s) \quad \text{s.t.} \\ \forall(i, a, s), \quad & \sum_{b_i \in A_i} \lambda_i(a_i, b_i)(\Pr(s|a_{-i}, b_i) - \Pr(s|a)) = \eta(a, s). \end{aligned}$$

If ICE is satisfied, then the value of the primal equals 0 for any $K: A \times S \rightarrow \mathbb{R}$. By the strong duality theorem, the value of the dual is also 0 for any $K: A \times S \rightarrow \mathbb{R}$. Therefore, any η satisfying the constraint for some λ must be 0 for all (a, s) , so DIA is satisfied.

For sufficiency, if DIA holds, then the value of the dual is always 0 for any $K: A \times S \rightarrow \mathbb{R}$. By strong duality, the value of the primal is also 0 for any K .

Therefore, given K , there is a feasible primal solution $\xi_i(a, s)$ that satisfies all primal constraints, and ICE holds. $Q.E.D.$

PROOF OF THEOREM 7: We use the following notation. Given a correlated strategy σ and a deviation plan α_i , let $\Delta v_i(\sigma, \alpha_i) = \sum_{(a, b_i, \rho_i)} \sigma(a) \alpha_i(b_i, \rho_i | a_i) \times (v_i(a_{-i}, b_i) - v_i(a))$ be the utility gain from α_i at σ and let $\Delta \Pr(s|a, \alpha_i) = \sum_{(a, b_i, \rho_i)} \alpha_i(b_i, \rho_i | a_i) (\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a))$ be the change in the probability that s is reported from α_i at a . Enforcing an arbitrary correlated strategy σ subject to participation constraints reduces to finding transfers ζ to solve the family of linear inequalities

$$\begin{aligned} \forall(i, a_i, b_i, \rho_i), \quad & \sum_{a_{-i}} \sigma(a) (v_i(a_{-i}, b_i) - v_i(a)) \\ & \leq \sum_{(a_{-i}, s)} \sigma(a) \zeta_i(a, s) (\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)), \\ \forall(a, s), \quad & \sum_{i=1}^n \zeta_i(a, s) = 0, \\ \forall i \in I, \quad & \sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a, s)} \sigma(a) \zeta_i(a, s) \Pr(s|a) \geq 0. \end{aligned}$$

The dual of this problem subject to participation is

$$\begin{aligned} \max_{\lambda, \pi \geq 0, \eta} \quad & \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - \pi_i v_i(\sigma) \quad \text{s.t.} \\ \forall(i, a, s), \quad & \sigma(a) \Delta \Pr(s|a, \lambda_i) = \eta(a, s) + \pi_i \sigma \Pr(s|a), \end{aligned}$$

where π_i is a multiplier for agent i 's participation constraint and $v_i(\sigma) = \sum_a \sigma(a) v_i(a)$. Adding the dual constraints with respect to $s \in S$, it follows that $\pi_i = \pi$ does not depend on i . Redefining $\eta(a, s)$ as $\eta(a, s) + \pi \Pr(s|a)$, the set of feasible $\lambda \geq 0$ is the same as without participation constraints. Since $\sum_i v_i(a) \geq 0$ for all a , the dual is maximized by $\pi = 0$. $Q.E.D.$

PROOF OF THEOREM 8: We use the same notation as in the proof of Theorem 7. Let $z = (z_1, \dots, z_n)$ be a vector of liability limits for each agent. Enforcing σ subject to limited liability reduces to finding ζ such that

$$\begin{aligned} \forall(i, a_i, b_i, \rho_i), \quad & \sum_{a_{-i}} \sigma(a) (v_i(a_{-i}, b_i) - v_i(a)) \\ & \leq \sum_{(a_{-i}, s)} \sigma(a) \zeta_i(a, s) (\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)), \end{aligned}$$

$$\forall(a, s), \quad \sum_{i=1}^n \zeta_i(a, s) = 0,$$

$$\forall(i, a, s), \quad \zeta_i(a, s) \leq z_i.$$

The dual of this metering problem subject to one-sided limited liability is given by

$$\max_{\lambda, \beta \geq 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) - \sum_{(i, a, s)} \beta_i(a, s) z_i \quad \text{s.t.}$$

$$\forall(i, a, s), \quad \sigma(a) \Delta \Pr(s|a, \lambda_i) = \eta(a, s) + \beta_i(a, s),$$

where $\beta_i(a, s)$ is a multiplier on the liability constraint for agent i at (a, s) . Adding the dual equations with respect to s implies $-\sum_s \beta_i(a, s) = \sum_s \eta(a, s)$ for all (i, a) . Therefore,

$$-\sum_{(i, s)} \beta_i(a, s) z_i = \sum_{(i, s)} \eta(a, s) z_i = \hat{z} \sum_{s \in S} \eta(a, s),$$

where $\hat{z} = \sum_i z_i$, so we may eliminate $\beta_i(a, s)$ from the dual and get the equivalent problem:

$$\max_{\lambda \geq 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) + \hat{z} \sum_{(a, s)} \eta(a, s) \quad \text{s.t.}$$

$$\forall(i, a, s), \quad \sigma(a) \Delta \Pr(s|a, \lambda_i) \geq \eta(a, s).$$

Any two liability profiles z and z' with $\hat{z} = \hat{z}'$ lead to this dual with the same value. $Q.E.D.$

PROOF OF THEOREM 9: We use the same notation as in the proof of Theorem 7. Enforcing σ subject to participation and liability is equivalent to the value of the following problem being zero:

$$\min_{\zeta} \sum_{(i, a_i)} \varepsilon_i(a_i) \quad \text{s.t.}$$

$$\forall(i, a, s), \quad \zeta_i(a, s) \leq z_i, \quad \forall(i, a_i, b_i, \rho_i),$$

$$\sum_{a_{-i}} \sigma(a) (v_i(a_{-i}, b_i) - v_i(a))$$

$$\leq \sum_{(a_{-i}, s)} \sigma(a) \zeta_i(a, s) (\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)) + \varepsilon_i(a_i),$$

$$\forall(a, s), \quad \sum_{i \in I} \zeta_i(a, s) = 0,$$

$$\forall i \in I, \quad \sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a) \zeta_i(a, s) \Pr(s|a) \geq 0.$$

The first family of constraints imposes incentive compatibility, the second family imposes budget balance, the third family imposes individual rationality, and the last family corresponds to one-sided limited liability. The dual of this metering problem is given by the following program, where λ , η , π , and β represent the respective multipliers on each of the primal constraints:

$$\begin{aligned} & \max_{\alpha, \pi, \beta \geq 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \alpha_i) - \sum_{i \in I} \pi_i v_i(\sigma) - \sum_{(i,a,s)} \beta_i(a, s) z_i \quad \text{s.t.} \\ & \forall (i, a_i), \quad \sum_{(b_i, \rho_i)} \alpha_i(b_i, \rho_i | a_i) = 1, \\ & \forall (i, a, s), \quad \sigma(a) \Delta \Pr(s|a, \alpha_i) = \eta(a, s) + \pi_i \sigma(a) \Pr(s|a) + \beta_i(a, s). \end{aligned}$$

Adding the dual constraints with respect to $s \in S$, it follows that

$$-\sum_{(a,s)} \beta_i(a, s) = \sum_{(a,s)} \eta(a, s) + \pi_i = \widehat{\eta} + \pi_i,$$

where $\widehat{\eta} := \sum_{(a,s)} \eta(a, s)$. After substituting and eliminating β , the dual is equivalent to

$$\begin{aligned} V := \max_{\alpha, \pi \geq 0, \eta} & \sum_{i \in I} \Delta v_i(\sigma, \alpha_i) - \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \widehat{z} \quad \text{s.t.} \\ & \forall (i, a, s), \quad \sigma(a) \Delta \Pr(s|a, \alpha_i) \geq \eta(a, s) + \pi_i \sigma(a) \Pr(s|a). \end{aligned}$$

Now, σ is enforceable if and only if $V = 0$, that is, if and only if for any dual-feasible (α, π, η) such that $\sum_i \Delta v_i(\sigma, \alpha_i) > 0$, we have that

$$\sum_{i \in I} \Delta v_i(\sigma, \alpha_i) \leq \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \widehat{z}.$$

Finally, since the dual objective is increasing in η , an optimal solution for η must solve

$$\eta(a, s) = \min_{i \in I} \{\Delta \Pr(s|a, \alpha_i) - \pi_i \Pr(s|a)\}.$$

This completes the proof. *Q.E.D.*

PROOF OF COROLLARY 1: Given the dual problem from the proof of Theorem 9, the first statement follows because if $v_i(\sigma) \geq z_i$, then the objective

function is decreasing in π_i and reducing π_i relaxes the dual constraints. The second statement follows by rewriting the objective as

$$\sum_{i \in I} \Delta v_i(\sigma, \alpha_i) - \sum_{i \in I \setminus t} \pi_i(v_i(\sigma) - z_i) + \hat{\eta} \sum_{i \in I} z_i,$$

where t is the set of agents whose participation constraint will not bind ($\pi_i^* = 0$ for $i \in t$). *Q.E.D.*

REFERENCES

- ALCHIAN, A., AND H. DEMSETZ (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777–795. [285]
- AOYAGI, M. (2005): "Collusion Through Mediated Communication in Repeated Games With Imperfect Private Monitoring," *Economic Theory*, 25, 455–475. [299]
- AUMANN, R. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96. [285]
- (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18. [289]
- CREMER, J., AND R. MCLEAN (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257. [299]
- D'ASPREMONT, C., AND L.-A. GÉRARD-VARET (1998): "Linear Inequality Methods to Enforce Partnerships Under Uncertainty: An Overview," *Games and Economic Behavior*, 25, 311–336. [285,298–300]
- D'ASPREMONT, C., J. CREMER, AND L.-A. GÉRARD-VARET (2004): "Balanced Bayesian Mechanisms," *Journal of Economic Theory*, 115, 385–396. [297]
- FORGES, F. (1986): "An Approach to Communication Equilibria," *Econometrica*, 54, 1375–1385. [285,290]
- FUDENBERG, D., D. LEVINE, AND E. MASKIN (1994): "The Folk Theorem With Imperfect Public Information," *Econometrica*, 62, 997–1039. [286,299,300]
- HOLMSTRÖM, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340. [285, 287,288,296]
- KANDORI, M. (2003): "Randomization, Communication, and Efficiency in Repeated Games With Imperfect Public Monitoring," *Econometrica*, 71, 345–353. [286,299]
- KOSENOK, G., AND S. SEVERINOV (2008): "Individually Rational, Balanced-Budget Bayesian Mechanisms and the Allocation of Surplus," *Journal of Economic Theory*, 140, 126–261. [299]
- LEGROS, P., AND H. MATSUSHIMA (1991): "Efficiency in Partnerships," *Journal of Economic Theory*, 55, 296–322. [285,298,299]
- LEGROS, P., AND S. MATTHEWS (1993): "Efficient and Nearly Efficient Partnerships," *Review of Economic Studies*, 60, 599–611. [285,286,298,299]
- MYERSON, R. (1986): "Multistage Games With Communication," *Econometrica*, 54, 323–358. [285,290]
- OBARA, I. (2008): "The Full Surplus Extraction Theorem With Hidden Actions," *The B.E. Journal of Theoretical Economics*, 8, 1–26. [299]
- RADNER, R., R. MYERSON, AND E. MASKIN (1986): "An Example of a Repeated Partnership Game With Discounting and With Uniformly Inefficient Equilibria," *Review of Economic Studies*, 53, 59–69. [285,287,288]
- RAHMAN, D. (2008): "But Who Will Monitor the Monitor?" Working Paper, University of Minnesota. [296]
- SCHRIJVER, A. (1986): *Theory of Linear and Integer Programming*. New York: Wiley-Interscience. [301]

TOMALA, T. (2009): “Perfect Communication Equilibria in Repeated Games With Imperfect Monitoring,” *Games and Economic Behavior*, 67, 682–694. [299]

*Dept. of Economics, University of Minnesota, 4-101 Hanson Hall, 1925 Fourth Street South, Minneapolis, MN 55455, U.S.A.; dmr@umn.edu
and*

Dept. of Economics, University of California, Los Angeles, 8283 Bunche Hall, Los Angeles, CA 90096, U.S.A.; iobara@econ.ucla.edu.

Manuscript received October, 2005; final revision received October, 2009.

RECURSIVE EQUILIBRIUM IN STOCHASTIC OVERLAPPING-GENERATIONS ECONOMIES

BY ALESSANDRO CITANNA AND PAOLO SICONOLFI¹

We prove the generic existence of a recursive equilibrium for overlapping-generations economies with uncertainty. “Generic” here means in a residual set of utilities and endowments. The result holds provided there is a sufficient number of potentially different individuals within each cohort.

KEYWORDS: Overlapping generations, Markov equilibrium, recursive equilibrium, transversality theorem.

1. INTRODUCTION

THE OVERLAPPING-GENERATIONS (OLG) model, introduced first by Allais (1947) and Samuelson (1958), is one of the two major workhorses for macroeconomic and financial modeling of open-ended dynamic economies. Following developments in the study of two-period economies, the OLG model has been extended to cover stochastic economies with production and possibly incomplete financial markets. As is the consensus, in dynamic economies the general notion of competitive equilibrium à la Arrow and Debreu, which allows for prices and allocations to depend on histories of arbitrary length, is not always fully satisfactory for a variety of reasons. Among them, it is worth recalling at least the following two. First, from a theoretical viewpoint, when prices have unbounded memory, the notion of rational expectations equilibrium is strained because of the complexity of the forecasts and the expectations coordination involved. Second, the ensuing large dimensionality of the allocation and price sequences strains the ability of approximating solutions with present-day computers through truncations, rendering this general notion of equilibrium not very useful for applied, quantitative work, even in stationary Markovian environments.

Duffie, Geanakoplos, Mas-Colell, and McLennan (1994) gave a general theorem for the existence of stationary Markov equilibria for OLG economies with associated ergodic measure.² While these equilibria help bypass the two above-mentioned issues, they are still quite complicated as the state space contains a number of past and current endogenous variables. Therefore, the conceptual issue of whether it is possible to find simpler equilibria—stationary Markov equilibria based on a minimal state space—remains open.

¹We thank John Geanakoplos, Felix Kubler, David Levine, Michael Magill, Herakles Polemarchakis, Martine Quinzii, and three anonymous referees for useful comments. The paper was written while the first author was visiting the Columbia Business School, whose financial support is gratefully acknowledged.

²See, for example, also the earlier work by Spear (1985), Cass, Green, and Spear (1992), and Gottardi (1996).

In parallel developments, and to allow for computational work, the literature in macroeconomics and finance has focused on a simpler notion of time-homogeneous Markov equilibrium, also known as recursive equilibrium.³ In a recursive equilibrium, the state space is reduced to the exogenous shocks and the initial distribution of wealth for the agents—asset portfolios from the previous period, and capital and storage levels if production is considered. The notion of recursive equilibrium also originates in the long-standing tradition of using recursive methods in economics (see, e.g., Stokey and Lucas (1989) and Ljungqvist and Sargent (2000)), and is the “natural” extension of those methods to stochastic OLG models with heterogeneity. A recursive equilibrium can be thought of as a time-homogeneous Markov equilibrium that is based on a minimal state space.

However, no existence theorem is available for such recursive equilibria. In fact, Kubler and Polemarchakis (2004) provided two examples of nonexistence of recursive equilibrium in OLG exchange economies. The idea that recursive equilibria may not exist is based on the observation that when there are multiple equilibria, the continuation of an equilibrium may depend on past economic variables other than the wealth distribution. That is, the current wealth distribution may not be enough to summarize the information contained in past equilibrium prices and marginal utilities.

While this phenomenon may occur, we prove that it is nongeneric under some qualifying condition. The argument follows from two fundamental observations.

First, we show that competitive equilibria, which are time-homogeneous Markov over a *simple* state space, exist. This is the state space made of the current exogenous state, the current wealth distribution, current commodity prices, and marginal utilities of income for all generations except for the first and the last (“newly born” and “eldest”). This result is similar to that obtained by Duffie et al. (1994, Section 2.5). In fact, we even construct such equilibria that guarantee that the Markov state space contains all the wealth distributions that can be taken to be initial conditions for competitive equilibria of each economy considered.

Second, we show that simple Markov equilibria typically satisfy the condition that prices and multipliers are a function of the current state and of the current wealth distribution, provided that there is a large number of individuals within each generation. We call these equilibria nonconfounding, while we call equilibria that do not satisfy that condition confounding. Clearly, nonconfounding simple time-homogeneous Markov equilibria are recursive.

The trick we use is to find a finite system of equations that must necessarily have a solution so that simple Markov equilibria fail to be nonconfounding.

³Examples by now abound; see, for example, Rios Rull (1996), Constantinides, Donaldson, and Mehra (2002), Geanakoplos, Magill, and Quinzii (2004), and Storesletten, Telmer, and Yaron (2004).

This trick allows one to bypass the infinite-dimensional nature of the equilibrium set and the fact that, with overlapping generations, there is an infinite number of individuals and an infinite number of market clearing equations, rendering direct genericity analysis quite problematic.

Specifically, if there exists a confounding equilibrium, there is a pair of Markov states with identical current shock and wealth distribution, but different multipliers or different current prices. However, the joint requirement of equal wealth but different multipliers restricts the future prices following the realization of the two critical Markov states to be such that individuals with equal wealth make different choices. Since individuals have finite lives, the latter is a restriction over finitely many future prices. Equivalently, both the dimension of the relevant price processes and the number of wealth equalities are finite. We show that with a large enough number of potentially different individuals of the same generation (Assumption A1), typically the wealth equalities are not satisfied, thereby establishing the generic existence of recursive equilibria. Since we need to check this for all admissible—and not just equilibrium—pairs of price processes, the number of individuals must be large. It should be noted in passing that this is not at odds with the notion of price-taking behavior, which is assumed in competitive models such as ours.

The notion of genericity we will use will rely also on utility perturbations and, therefore, will only be topological. In fact, due to the infinite dimension of the equilibrium set, we will not be able to establish local uniqueness of competitive equilibria, whether or not it is time-homogeneous. Without this prerequisite, the argument that essentially shows some one-to-one property of prices will have to be made without knowing whether or not such prices are “critical”; in fact, without knowing if they are equilibrium prices. Therefore, we will resort to an argument reminiscent of Mas-Colell and Nachbar (1991), and we will show the existence of recursive equilibria for a *residual* or *nonmeager* subset of parameters, that is, a set of stationary utilities and endowments that is dense and is the countable intersection of open and dense sets. This is a well established notion of genericity for dynamic systems.

Our class of OLG economies has multiple goods, generations, and types within each generation. We present first the simplest model where there are only short-lived real—namely, numéraire—assets in zero net supply. Long-lived assets with nonzero real payoffs and in positive net supply, production (as in Rios Rull (1996)), and economies with individual risk are briefly discussed at the end of the paper. Our result encompasses all these extensions and the argument of the paper goes through unchanged. In fact, in the extensions the Markov state space becomes richer, thereby possibly relaxing Assumption A1.

To simplify the exposition, in this paper we carry out the density part of the argument for complete financial markets. When markets are incomplete, density proofs are much more elaborate. Hence, we leave the density analysis with incomplete markets to a companion paper (Citanna and Siconolfi (2007)).

Our result suggests that the notion of recursive equilibrium is coherent as an exact concept, adding robustness to its interpretation and quantitative use.⁴ While the number of agents in each cohort must be large, their characteristics can always be taken to be very similar for the underlying economy to have exact recursive equilibria. We must stress that although we cannot claim that economies with a large number of identical individuals are part of the generic set where recursive equilibrium exists, nothing in the argument indicates otherwise.⁵

The paper is organized as follows. Section 2 introduces the stochastic OLG economy model. Section 3 gives the definition of the relevant notions of equilibrium: competitive, simple Markov, and recursive. Section 4 states the main theorem and outlines the fundamental aspects of the argument. In particular, Section 4.1 shows existence of simple Markov equilibria, while Section 4.2 contains the properties of demand that arise from a distribution of agents in two cohorts. Section 5 discusses extensions to the basic model and related results.

2. THE MODEL

We consider standard stationary OLG economies. Time is discrete, indexed by $t = 0, 1, 2, \dots$. There are $S > 1$ states of the world that may be realized in each period. At each t , H individuals enter the economy. Each individual $h \in H$ lives $G + 1 \geq 2$ periods, indexed by $a = 0, \dots, G$, from the youngest ($a = 0$) to the oldest ($a = G$) age. If and when we need to make explicit that a variable or function is of a specific individual of type h (and of age a), we use the superscript h (and ha). At each t , $C \geq 1$ physical commodities are available for consumption. The consumption bundle of an individual of age a is $x_t^a \in \mathbb{R}_{++}^C$. Each individual h has a discounted, time-separable, von Neumann-Morgenstern utility function with time-, state-, and age-invariant Bernoulli utility index u^h at age a . Each $u^h : \mathbb{R}_{++}^C \rightarrow \mathbb{R}$ is twice continuously differentiable, differentially strictly increasing, and differentially strictly concave (the Hessian is negative definite), and satisfies the following boundary condition: if $x_{c,t}^{ha} \rightarrow 0$, then $\|Du^h(x_t^a)\| \rightarrow +\infty$, where Du^h denotes the vector of partial derivatives. The common discount factor is $\delta \in (0, 1]$. At each t , endowments are $e_t^{ha} \in \mathbb{R}_{++}^C$ for $h \in H$ and $0 \leq a \leq G$.

⁴For their interpretation as mere computing devices, see Kubler and Schmedders (2005). When feedback policies are examined in a recursive formulation, their interpretation is of course strengthened by the exact, as opposed to only approximate, nature of recursivity. As we stressed earlier, the focus on recursive equilibria can also be justified by their “simplicity” when they are exact.

⁵On the other hand, if an economy with possibly many identical individuals does not have a recursive equilibrium, independent and arbitrarily small perturbations of their characteristics restore its existence. While Assumption A1 does not necessarily encompass the economies of the examples constructed by Kubler and Polemarchakis (2004), in a previous paper (Citanna and Siconolfi (2008)) we showed that such examples are nonrobust.

At each t , there are competitive spot markets for the exchange of physical commodities. The price vector of the C commodities is $p_t = (\dots, p_{c,t}, \dots) \in \mathbb{R}_{++}^C$. Commodity $c = 1$ is dubbed the numéraire commodity and hereafter, unless we say otherwise, we adopt the normalization $p_{1,t} = 1$ for all t . There are also competitive markets for trading S one-period securities in zero net supply, with prices $q_t \in \mathbb{R}^S$. We set $\psi_t \equiv (p_t, q_t)$.

The security payoffs $d_t \in \mathbb{R}^S$ are in units of the numéraire commodity. The portfolio of an individual of age a is $\theta_t^a \in \mathbb{R}^S$. We assume that $\theta_t^{-1} \equiv \theta_t^G \equiv 0$ and, denoting with $w_t^a \equiv d_t \theta_{t-1}^{a-1}$ the financial wealth of an individual of age a at t , that $w_t^0 = 0$ for all t . Also, $w_0^a = d_0 \theta_{-1}^{a-1} \in \mathbb{R}$ is the individual initial wealth at $t = 0$ for $a > 0$, while $w_0 \equiv (w_0^h)_{h \in H}$ denotes the initial wealth distribution of the economy.

Finally, we let λ_t^{ha} be the marginal utility of income for all (h, a) ; that is, $\lambda_t^{ha} = D_1 u^h(x_t^a)$, where D_1 denotes the derivative with respect to the first entry of the vector x_t^a .

To define the various notions of equilibrium, all exogenous and endogenous variables are seen as stochastic processes on a probability space $(\mathcal{X}, \mathcal{F}, \Pr)$. First, the sequence of exogenous shocks is seen as a realization of the process $\tilde{s} = (s_t, t = 0, 1, \dots)$ that is constructed in a standard way through a given $S \times S$ stochastic matrix π , where each $s_t : \mathcal{X} \rightarrow S$ is \mathcal{F} -measurable, and with initial shock $s_0(\chi) = s_0$ for some $s_0 \in S$ for \Pr -a.a. χ . The matrix π is the time-invariant transition of this process and $\pi(s_{t+1}|s_t) = \Pr\{\chi : s_{t+1} = s_{t+1}(\chi)|s_t = s_t(\chi)\}$ defines the probability of shock realization s_{t+1} given s_t , that is, \tilde{s} is a time-homogeneous, first-order Markov chain.

Endowments e_t^{ha} are assumed to be affected only by current realizations of $s \in S$ and are denoted by $e_s^{ha} \in \mathbb{R}_{++}^C$ for $h \in H$, $0 \leq a \leq G$, and $s \in S$. The security payoff d_t is also only affected by the state $s_t \in S$ and is d_{s_t} , and is assumed to give rise to a full-rank, $S \times S$ -dimensional matrix \mathbf{D} which is then time and history invariant. Hence, state realizations affect endowments and preferences of all individuals of all ages and security payoffs, while time does not. Thus, the fundamentals of our economies follow a time-homogeneous, first-order Markov process, so that the economy is *stationary*.

In what follows, we keep (π, \mathbf{D}) fixed and we identify an economy with an endowment and utility profile e_s^{ha}, u^h , for all $h \in H$, $a = 0, \dots, G$, $s \in S$, and discount factor $\delta \in (0, 1]$. The endowment space E is an open subset of $\mathbb{R}_{++}^{H(G+1)SC}$. The set \mathcal{U} of utilities u^h is the G_δ subset of $\bigtimes_h \mathcal{C}^2(\mathbb{R}_{++}^C, \mathbb{R})$, which satisfies our maintained assumptions and is endowed with the topology of \mathcal{C}^2 -uniform convergence on compacta. The set \mathcal{U} is a complete metric space, and, by the Baire theorem, countable intersections of open and dense sets are dense in \mathcal{U} . Let $\Omega = E \times \mathcal{U} \times (0, 1]$ be the space of economies with the product topology.

3. EQUILIBRIUM

We give a unified view of the different notions of equilibria that we study as Markov processes defined over a common state space. Competitive equilibria

of an OLG economy are stochastic processes attached to an initial condition, that is, an initial shock and a wealth distribution. In principle, the entire equilibrium trajectory may depend on the initial condition, and its continuation path may be history dependent. Since we are aiming at a unified treatment, histories and initial conditions must be part of our state space.

For given χ and \tilde{s} , a history of length t is an array of shock realizations with identical initial shock, $s^t(\chi) = (s_0, s_1, \dots, s_t)(\chi)$. For a given process \tilde{s} , let $S_{s_0}^t$ be the set of all possible histories of length t with initial shock s_0 . History s^t precedes history \tilde{s}^t and we write $\tilde{s}^t \succsim s^t$ if there is an array of $t' - t$ realizations of shocks $s^{t'-t}$ such that $\tilde{s}^t = (s^t, s^{t'-t})$. We denote with $\tilde{S}_{s_0} = \bigcup_t S_{s_0}^t$ the set of all possible histories of any length, and denote with initial shock s_0 , a tree with nodes s^t and root s_0 . When all possible initial conditions need to be considered at once, the union $\tilde{S} = \bigcup_{s_0 \in S} \tilde{S}_{s_0}$ of S trees with distinct roots comes into play.

Let the hyperplane $W \equiv \{w \in \mathbb{R}^{(G+1)H} : w^{h0} = 0, \text{ for all } h, \text{ and } \sum_{ha} w^{ha} = 0\}$ be the space of wealth distributions. The space of endogenous variables $\Xi = (\mathbb{R}_{++}^C \times \mathbb{R}^S \times \mathbb{R}_{++})_{h,a} \times \mathbb{R}_{++}^C \times \mathbb{R}^S$ has element $\xi = ((x^{ha}, \theta^{ha}, \lambda^{ha})_{h,a}, \psi)$, which comprises the vectors of consumption bundles, of portfolios, of marginal utilities of income for all individuals of all ages, and of prices.

To define trajectories which are history and initial-condition dependent, an array $(s^t, w, \xi) \in \tilde{S} \times W \times \Xi$, specifying a history, current values of the wealth distribution, and endogenous variables, has to be augmented with an initial state $(s_0, w_0, \xi_0) \in Z_0 = S \times W \times \Xi$. The state space is then $Z \subset \tilde{S} \times W \times \Xi \times Z_0$, and we write z to denote a generic element of Z , write ζ for an element of $W \times \Xi$, and write $z_0 = (s_0, w_0, \xi_0)$ for an initial state. We assume that the projection of Z on \tilde{S} is onto, that is, no histories s^t are excluded.

Stochastic processes over Z are defined by specifying a transition. We limit attention to spotless transitions, that is, at each t the process can take as many values as realizations of the exogenous shocks. We can therefore identify a transition with a deterministic map $T:Z \rightarrow Z^S$ such that $T(s^t, \zeta, z_0) = ((s^t, s), \zeta_s, z_0)_{s \in S}$. Therefore, $\Pr(z'|z) = \pi(s_{t+1}|s_t)$ for $z = (s^t, \cdot)$ and $z' = ((s^t, s_{t+1}), \cdot)$. We also write $T = (T_s)_{s \in S}$ with $T_s:Z \rightarrow Z$ and $z' = ((s^t, s), \zeta, z_0) = T_s(z)$. Notice two aspects of this construction. First, the transition T leaves the initial condition unaltered. Second, the spotless nature of the transition allows us to dispense with any additional restriction on the transition map T such as, for instance, (Borel) measurability with respect to $W \times \Xi$.

The pair (Z, T) generates a family of stochastic processes $\tilde{z} = (\tilde{z}_{z_0}, z_0 \in Z_0)$ or simply a process, each with $z_{z_0,0}(\chi) = (z_0, z_0)$, $z_{z_0,1}(\chi) = T_{s_1(\chi)}(z_{z_0,0}(\chi))$, and, recursively, $z_{z_0,t}(\chi) = T_{s_t(\chi)}[\cdots T_{s_1(\chi)}(z_{z_0,0}(\chi))]$ for Pr-a.a. χ . Naturally a process \tilde{z} generated by (Z, T) induces an endogenous variables process $\tilde{\xi}$ via the natural projection, that is, $\xi_t = \text{proj}_{\Xi} z_t$, and we say that $\tilde{\xi}$ is generated by (Z, T) . Since the stochastic engine of \tilde{z}_{z_0} is the shock process \tilde{s} , $\tilde{\xi}_{z_0}$ is adapted to \tilde{S} for all initial conditions z_0 .

A process $\tilde{\xi}$ generated by (Z, T) is a family of *competitive equilibrium processes* if, for Pr-a.a. χ and all $t \geq 0$, the following conditions hold:

CONDITION H: For all h ,

$$(1a) \quad Du^h(x_t^{ha}) - \lambda_t^{ha} p_t = 0 \quad \text{for all } a,$$

$$(1b) \quad -\lambda_t^{ha} q_t + \delta \mathbb{E}_t(\lambda_{t+1}^{h(a+1)} d_{t+1}) = 0 \quad \text{for all } a < G,$$

$$(1c) \quad \psi_t[(x_t^{ha} - e_t^{ha}), \theta_t^{ha}] = d_t \theta_{t-1}^{h(a-1)} \quad \text{for all } z' = T(z), \text{ all } a.$$

CONDITION M: $\sum(x_t^{ha} - e_t^{ha}, \theta_t^{ha}) = 0$.

Condition **M** is the familiar market clearing equation. Condition **H** is optimality, that is, it is the first-order conditions for the utility maximization problem of any individual h of age a at any t , choosing consumption and asset portfolios facing the competitive sequential budget constraint (1c) for $\tau = t, \dots, G - a$. By the assumptions on u^h and the Markovian structure of the economy, time consistency is satisfied and these conditions apply to the problem faced at any $\tau = t - a, \dots, t$.

A competitive equilibrium is an element $\tilde{\xi}_{\bar{z}}$ from a family of competitive equilibrium processes where $z_0 = \bar{z} = (\bar{s}, \bar{w}, \tilde{\xi})$ for Pr-a.a. χ is a given initial condition for the economy. Since the initial condition specifies also the values of the endogenous variables at \bar{s} , the same initial pair (\bar{s}, \bar{w}) may be associated to multiple continuation paths, one for each distinct vector $\tilde{\xi}$. Equivalently, our definition allows for multiple competitive equilibria of an economy ω starting off at (\bar{s}, \bar{w}) .

As is well known, the existence of a competitive equilibrium $\tilde{\xi}_{\bar{z}}$ for an economy ω restricts the choice of the initial wealth distributions \bar{w} to a bounded subset $W_{\bar{s}, \omega}$ of the hyperplane W , with nonempty interior $\overset{\circ}{W}_{\bar{s}, \omega}$. Given the boundary condition on utilities, for all economies ω and for all initial conditions (\bar{s}, \bar{w}) , $\bar{w} \in W_{\bar{s}, \omega}$, the existence of a competitive equilibrium is established using a standard truncation argument (see Balasko and Shell (1980)).

The equilibrium process $\tilde{\xi}$ is a Markov process, but it is not time-homogeneous. The focus of this paper is on a special kind of stationary, time-homogeneous competitive equilibrium, also known as recursive equilibrium. To study recursive equilibria, our argument will go through and use an intermediate notion of equilibrium that is stationary, time-homogeneous, and Markov, and that we call simple Markov.

A *simple Markov equilibrium* is a family of competitive equilibrium processes $\tilde{\xi}$ generated by a pair (Z, T) that satisfies:

$$T(s^t, \zeta, z_0) = T(\hat{s}^t, \hat{\zeta}, \hat{z}_0) \quad \text{if} \quad (s_t, w^{ha}, \lambda^{ha}, p) = (\hat{s}_t, \hat{w}^{ha}, \hat{\lambda}^{ha}, \hat{p}) \\ \text{for all } h, a.$$

A *recursive equilibrium* is a simple Markov equilibrium $\tilde{\xi}$ that satisfies:

$$T(s^t, \zeta, z_0) = T(\hat{s}^t, \hat{\zeta}, \hat{z}_0) \quad \text{if } (s_t, w^{ha}) = (\hat{s}_t, \hat{w}^{ha}) \quad \text{for all } h, a.$$

Two aspects are worth noticing. Different restrictions on the transition T deliver different equilibrium notions. The restrictions on the map T translate into restrictions on Z : all elements of Z satisfy the same constraints imposed on T , thereby effectively generating a reduced state space, the subspace of the coordinates making the transition T (potentially) injective. Thus, the reduced state space of a simple Markov equilibrium includes only the current exogenous state s and, as endogenous states, wealth distribution w , commodity prices p , and marginal utilities of income for all generations except for the first and the last, $\lambda \equiv ((\lambda^{ha})_{0 < a < G})_{h \in H}$: with some abuse of notation, for a simple Markov, Z is the set of vectors $(s, \zeta) = (s, w, p, \lambda)$. For a recursive equilibrium, the endogenous state is only the wealth distribution w , and Z is the set of vectors $(s, \zeta) = (s, w)$. The transition $T: Z \rightarrow Z^S$ maps the current state (s, ζ) into all its S immediate successors (s', ζ') . An endogenous variables function $\xi: Z \rightarrow \Xi$ completes the construction. Hence, our definition of recursive equilibrium coincides with the usual formulation found in the literature (see, e.g., Kubler and Polemarchakis (2004) or Rios Rull (1996)). We write Z_ω , T_ω , and ξ_ω when we want to stress their dependence on the economy ω , and write W_s^Z for the s section of the projection of Z on W .⁶

The following notions of initial Markov state and initial state–wealth pair will be important for the subsequent analysis. Consider a simple Markov equilibrium of an economy $\omega \in \Omega$. A state $\bar{z} = (\bar{s}, \bar{\zeta}) \in Z$ is an *initial Markov state* if there does not exist $s \in S$ such that $\bar{z} = T_{\bar{s}}(z)$ for some $z \in Z$. Thus, the economy can just start from an initial Markov state, but it can never reach that state starting from somewhere else—an initial Markov state is an extreme notion of transient state. A pair (\bar{s}, \bar{w}) is an *initial state–wealth pair* if $(\bar{s}, \bar{w}, p, \lambda)$ is an initial Markov state for some (p, λ) such that $(\bar{s}, \bar{w}, p, \lambda) \in Z$.

4. GENERIC EXISTENCE OF RECURSIVE EQUILIBRIA

We are going to show that under a qualifying condition on H , recursive equilibria typically exist and are compatible with a large set of initial conditions, that is, of initial wealth distributions.

The qualifying condition on H , which is used below, is the following inequality.

$$\text{ASSUMPTION A1: } H > 2[(C - 1) \sum_{a=0}^G S^a + S \sum_{a=0}^{G-1} S^a].$$

⁶That is, $W_s^Z = \{w \in W : (s, w, p, \lambda) \in Z \text{ for some } (p, \lambda)\}$ for a simple Markov and $W_s^Z = \{w \in W : (s, w) \in Z\}$ for a recursive equilibrium.

Our main result is then stated as follows.

THEOREM 1: *Under A1, there exists a residual subset Ω^* of Ω such that every economy ω in Ω^* has a recursive equilibrium. Furthermore, $W_{s,\omega}^Z$ contains an open and full Lebesgue measure subset $W_{s,\omega}^*$ of $\overset{\circ}{W}_{s,\omega}$ for all $s \in S$.*

We summarize the logic and the various steps involved in proving Theorem 1. We will first prove that simple Markov equilibria with a large wealth space exist (Proposition 2 in Section 4.1). Then, using a transversality argument, we will show that the wealth distribution is typically a sufficient statistic of the Markov states, that is, that typically a Markov equilibrium is recursive.

More precisely, we want to show that simple Markov equilibria, typically in ω and under A1, have the following injection property: if two Markov states z , \hat{z} are given, with $(s, w) = (\hat{s}, \hat{w})$, then $(p, \lambda) = (\hat{p}, \hat{\lambda})$. We call Markov equilibria that satisfy this property *nonconfounding*. It is immediate that a simple Markov equilibrium is nonconfounding if and only if it is a recursive equilibrium. Instead, a pair of Markov states z and \hat{z} violating this property is called *critical* and the corresponding equilibrium is called *confounding*. We also call an exogenous state s critical if $(s, \zeta_1) \in Z$ and $(s, \zeta_2) \in Z$ are critical Markov states for a pair ζ_k , $k = 1, 2$.

If (\bar{s}, ζ_1) and (\bar{s}, ζ_2) are a pair of critical Markov states, then

$$(2) \quad w_1 = w_2,$$

and either

$$(3) \quad p_1 \neq p_2$$

or

$$(4) \quad \lambda_1 \neq \lambda_2.$$

Clearly, (2) and (3) or (4) cannot have a solution if (2) does not have a solution for the H individuals of age $a = 1$ when (3) or (4) holds. The next step is then to establish via perturbation methods that indeed (2) for $a = 1$, and (3) and (4) cannot have a solution.

We can prove that system (2) and conditions (3) and (4) are incompatible at a simple Markov equilibrium if they also are incompatible at prices which are not necessarily equilibrium prices. In fact, we will only use restrictions on prices that arise from equilibrium to put them in a compact set $P(\omega)$. Since (w_1, λ_1) and (w_2, λ_2) are wealth and multiplier values that arise in the finite-dimensional optimization problems of H individuals of various ages, using stationarity and finite lives, the relevant price set $P(\omega)$ will also have finite dimension bounded by the right-hand side of Assumption A1. Hence we reduce

the problem of the existence of a recursive equilibrium to the problem of establishing that generically the joint wealth differentials $w_1^{h1} - w_2^{h1}$ that arise in the finite-dimensional optimization problems of H individuals in two cohorts cannot be zero in the price domain $P(\omega)$ that satisfies (3) or (4). This is done in Section 4.2 via Propositions 3 and 4.

4.1. Existence of Simple Markov Equilibrium

Markov equilibria have a “large” wealth space when the latter is compatible with a large set of initial wealth distributions of the competitive economy. The next proposition states that simple Markov equilibria exist and their wealth space is large, and also gives two properties of the initial state–wealth pairs that will be important for the generic existence of recursive equilibria.

PROPOSITION 2: *For all $\omega \in \Omega$ and any subsets $O_s \subseteq W_{s,\omega}$, $s \in S$, (i) there exists a simple Markov equilibrium with $O_s \subset W_{s,\omega}^Z$ for all $s \in S$, (ii) if (\bar{s}, \bar{w}) is an initial state–wealth pair, then $\bar{w} \in O_{\bar{s}}$, and (iii) for every $\bar{s} \in S$, there is a unique $(\bar{p}, \bar{\lambda})$ such that $(\bar{s}, \bar{w}, \bar{p}, \bar{\lambda})$ is an initial Markov state.*

See the Appendix for the proof.

Since O_s is any subset of $W_{s,\omega}$, simple Markov equilibria can be constructed so that $W_{s,\omega}^Z$ contains all initial conditions of the competitive economy. However, if we take $O_s = \overset{\circ}{W}_{s,\omega}$, by Proposition 2(ii) we construct a simple Markov equilibrium with initial wealth distribution contained in an open set, a precondition for later perturbations. Proposition 2(iii) will later allow us to exclude certain configurations of bad states. To get the idea of the proof of Proposition 2(i), pick an economy $\omega \in \Omega$ and consider the family of competitive equilibria $\tilde{\xi}_z$, $z = (s, w, \xi) \in Z_0$, with $w \in O_s$. Competitive equilibria may fail to be simple Markov because at some histories they may generate identical simple Markov states, but different values of some current endogenous variables or different continuation paths. Consider two competitive equilibria $\tilde{\xi}_{\bar{z}}$ and $\tilde{\xi}_z$ (with possibly $\bar{z} = z$) that generate at some histories s^t and $\hat{s}^{t'}$, respectively, the same Markov state. Construct a new equilibrium by grafting $\tilde{\xi}_{\bar{z}}$ onto $\tilde{\xi}_z$: follow the equilibrium process $\tilde{\xi}_z$, but modify it from $\hat{s}^{t'}$ on by using the process $\tilde{\xi}_{\bar{z}}$ as if the history were s^t . Our choice for Z implies that, checking the first-order conditions (1), the grafting technique still defines a competitive equilibrium. Then if, at some histories, multiple competitive equilibria generate identical Markov states, we can select one of them, thereby obtaining unique realizations of the endogenous variables and of the continuation paths.

Duffie et al. (1994) proved the existence of simplified equilibria where, when $G = 1$, the state space can be reduced to the current exogenous shock, the consumption of the young, and their portfolio choices. For comparison, in this

same case in our simple Markov equilibria, the state space reduces to the current shock, the current commodity prices, and the wealth distribution of the *current old*. It is this last state component that is the fundamental ingredient of our analysis; instead, it is missing in the construction of their simplified equilibria. We construct a selection directly by applying the grafting technique to the current state, making sure the previous and current equilibrium conditions are satisfied. Duffie et al.'s construction is forward looking and therefore does not make use of the wealth distribution of the old. Furthermore, their use of a measurable selection argument does not allow them to control the position of the initial state–wealth pairs—an important step of our argument (Proposition 2(ii)). Finally, their construction requires further measurability and topological assumptions on the endogenous variable functions and spaces what we avoid.

4.2. Some Properties of the Wealth Differentials

As argued above, the existence of recursive equilibria depends on generic properties of the joint wealth differentials of H individuals. Two such properties are of interest, depending on whether the wealth levels w_k^{h1} for some state k are exogenously given as initial wealth levels \bar{w}_k^{h1} or the wealth levels at both states $k = 1, 2$ result from optimization.

Observe that an individual of age a at t faces $N_a = \sum_{a'=0}^{G-a} S^{a'}$ histories $s^{t+a'} \geq s^t$ before death, $a = 0, \dots, G - a$. Thus, to study the joint wealth differentials, in the first case consider a tree of length G with initial node $s_{01} \in S$ and define over it an arbitrary finite price process $\underline{\psi} \in \mathbb{R}_{++}^{(C-1)N_0} \times \mathbb{R}^{SN_0}$, that is, a vector of prices (p_{sa}, q_{sa}) at each node s^a of the tree. As is well known, under our maintained assumptions—namely, stationarity and the boundary condition on u^h —competitive (and, therefore, Markov) equilibrium prices are bounded uniformly in $s^t \in \tilde{S}$ (or $s \in S$) and $(\bar{w}^{ha})_{h \in H, a > 0}$, the exogenously given financial wealth of the economy. That is, commodity prices are uniformly bounded above and bounded away from zero, while asset prices are uniformly bounded and bounded away from the boundary of the no-arbitrage region. Hence, if we think of $\underline{\psi}$ as the restriction of a simple Markov equilibrium price process to the finite histories represented by the tree, the equilibrium nature of this price process implies that we can take it in $P(\omega)$, a compact subset of $\mathbb{R}_{++}^{(C-1)N_0} \times \mathbb{R}^{SN_0}$ independent of \bar{w} .

We just look at the H individuals born at s_{01} , hence $P(\omega)$ will also satisfy the innocuous additional restriction $q_{s^G} = 0$, for all terminal nodes s^G : individuals born at s_{01} will be old and, in the absence of arbitrage (a necessary condition for the existence of equilibria, embedded in $\underline{\psi}$), they will not trade on the asset market. Individual optimization (regularity of demand) pins down all the endogenous variables of these individuals as (smooth) functions of p , q , and ω . In particular, consider all individuals of age 0 who solve at s_{01} their program-

ming problem when facing the finite process $\underline{\psi} \in P(\omega)$. Let $(\underline{x}^h, \underline{\theta}^h)(\underline{\psi}, \omega)$ be their optimal solution that takes values $(x_{s^a}^{ha}, \theta_{s^a}^{ha})(\underline{\psi}, \omega)$ at s^a and let

$$w_{\bar{s}^1}^{h1}(\underline{\psi}, \omega) = d_{\bar{s}} \theta_{s_01}^{h0}(\underline{\psi}, \omega)$$

be the wealth of an individual born at s_{01} and of age $a = 1$ at node $\bar{s}^1 = (s_{01}, \bar{s})$. For given $\bar{w} \in \overset{\circ}{W}_{\bar{s}, \omega}$, let

$$f_{\bar{s}^1}(\underline{\psi}, \omega, \bar{w}^1) \equiv (w_{\bar{s}^1}^{h1}(\underline{\psi}, \omega) - \bar{w}^{h1})_{h \in H}.$$

Under Assumption A1, for given $(\underline{\psi}, \bar{w}^1)$, the H equations $f_{\bar{s}^1}(\underline{\psi}, \omega, \bar{w}^1) = 0$ outnumber the unknowns. We use this to show that, for each $\omega \in \Omega$ and $\bar{s} \in S$, and for generic choices of $\bar{w} \in \overset{\circ}{W}_{\bar{s}, \omega}$, there is no such finite tree, that is, no \bar{s}^1 and no $\underline{\psi} \in P(\omega)$, where $f_{\bar{s}^1}(\underline{\psi}, \omega, \bar{w}^1) = 0$.

PROPOSITION 3: *Let $\omega \in \Omega$ be given. For each $\bar{s} \in S$ there is an open and full Lebesgue measure subset $W_{\bar{s}, \omega}^*$ of $\overset{\circ}{W}_{\bar{s}, \omega}$ such that $f_{\bar{s}^1}(\underline{\psi}, \omega, \bar{w}^1) = 0$ does not have a solution in $P(\omega)$ for all $\bar{w} \in W_{\bar{s}, \omega}^*$ and all (s_{01}, \bar{s}) .*

For the proof see the Appendix.

The proof of Proposition 3 uses a standard transversality argument through the computation of the derivative of $f_{\bar{s}^1}$ with respect to \bar{w}^1 .

Continuing the analysis of wealth differentials, we now study the second case. To this end, consider two trees of finite length G each with initial node s_{0k} , $k = 1, 2$. Otherwise identical histories s^a on the two trees may only differ in their initial node. When we want to stress this, we denote with (k, s^a) the history s^a on the k tree. Consider a pair of finite price processes $\underline{\psi}_k \in P(\omega)$ defined over the two trees. With some abuse of notation, the process $\underline{\psi}$ denotes now the pair $(\underline{\psi}_1, \underline{\psi}_2) \in P(\omega) \times P(\omega) = P(\omega)^2$.

Once again, we focus on all the individuals h born at s_{0k} , hence we can further restrict prices in $P(\omega)^2$ to satisfy $q_{k, s^G} = 0$ for all terminal nodes s^G and $k = 1, 2$. The optimizing behavior of the two cohorts (individuals born at s_{0k} ; one for each k) is entirely determined by $\underline{\psi} \in P(\omega)^2$ and ω . In particular, for each k , consider all individuals of age 0 solving at s_{0k} their programming problem when facing the finite process $\underline{\psi}_k \in P(\omega)$. Let $(\underline{x}_k^h, \underline{\theta}_k^h)(\underline{\psi}_k, \omega)$ be their optimal solution, taking values $(x_{k, s^a}^{ha}, \theta_{k, s^a}^{ha})(\underline{\psi}_k, \omega)$ at (k, s^a) , and let

$$w_{k, \bar{s}^1}^{h1}(\underline{\psi}_k, \omega) = d_{\bar{s}} \theta_{s_{0k}}^{h0}(\underline{\psi}_k, \omega)$$

be the wealth of an individual born at s_{0k} and of age $a = 1$ at node (k, \bar{s}^1) . To simplify the notation, hereafter we set $\sigma \equiv (s_{01}, s_{02}, \bar{s})$ and let

$$\hat{f}_\sigma(\underline{\psi}, \omega) \equiv (w_{1, \bar{s}^1}^{h1}(\underline{\psi}_1, \omega) - w_{2, \bar{s}^1}^{h1}(\underline{\psi}_2, \omega))_{h \in H}$$

be the wealth difference between the two age $a = 1$ cohorts at $\bar{s}^1 = (s_{0k}, \bar{s})$, $k = 1, 2$. We further restrict attention to $P(\sigma; \omega) \subset P(\omega)^2$, the set of price processes $\underline{\psi}$ that satisfy inequality (3) or induce (4) at node \bar{s}^1 among older cohorts, given that all these cohorts' wealth (w_{k, \bar{s}^1}^{ha} for $a \geq 1$) is k -invariant. We show below that in a residual set of parameters, the system of equations $\hat{f}_\sigma(\underline{\psi}, \omega) = 0$ does not have a solution for any σ and any price process in $P(\sigma; \omega)$.

PROPOSITION 4: *There exists a residual subset Ω^* of Ω such that $\hat{f}_\sigma(\underline{\psi}, \omega) = 0$ does not have a solution in $P(\sigma; \omega)$ for all $\omega \in \Omega^*$ and all σ .*

The proof of Proposition 4 is quite elaborate, but interesting in its own. Since it is central to our technique, reducing the whole issue to a finite-dimensional problem, we devote Section 4.4 to explaining its logic (a perturbation argument), while details and computations are in the Appendix. Although density, in general, is stated in the space of all parameters, the argument will show that when $G = 1$, genericity is only in endowments.

Taking for granted Propositions 2–4, we are now ready to prove Theorem 1.

4.3. Proof of Theorem 1

Pick an economy $\omega \in \Omega^*$, the set constructed in Proposition 4. Use Proposition 2(i) and (ii) to construct the state space Z_ω of the simple Markov equilibrium so that if $\bar{w} \in W_{s, \omega}^Z$ and (\bar{w}, s) is an initial state–wealth pair, then $\bar{w} \in W_{s, \omega}^*$ for all $s \in S$, the set defined in Proposition 3. We need to show that such equilibrium is void of critical Markov states. At this junction, two possibilities arise: Case 1, neither (\bar{s}, ζ_1) nor (\bar{s}, ζ_2) is initial Markov states; Case 2, there exists k such that (\bar{s}, ζ_k) is an initial Markov state of the economy. While in Case 1 we just refer to the Markov states as critical, in Case 2 we add the qualification “initial.”

CASE 1: Any simple Markov equilibrium of an economy $\omega \in \Omega^*$ is void of critical pairs of Markov states. Suppose not. Pick a pair of critical Markov states (\bar{s}, ζ_k) , $k = 1, 2$. Consider the H individuals born at the node on \tilde{S} predecessor to the one where (\bar{s}, ζ_k) has realized (s_{0k}) and of age $a = 1$ at the critical state.⁷ The (simple Markov) equilibrium price process matters to them only as restricted to the finite histories represented by the pair of trees since they were born, that is, as $\underline{\psi} \in P(\omega)^2$. Thanks to the stationarity of the economy, each pair of such trees is identified only by their initial nodes (s_{01}, s_{02}) . The individuals' wealth difference at age $a = 1$ is $\hat{f}_\sigma(\underline{\psi}, \omega)$, defined in Section 4.2. If

⁷The choice of $a = 1$ is dictated to avoid considering combinations of critical states other than the two studied here.

the pair (\bar{s}, ζ_k) , $k = 1, 2$, is critical, then (2), and conditions (3) and (4) correspond to $\hat{f}_\sigma(\underline{\psi}, \omega) = 0$ for some $\underline{\psi} \in P(\sigma; \omega)$ and some (s_{01}, s_{02}) , contradicting Proposition 4.

CASE 2: Consider now initial critical pairs of Markov states. By Proposition 2(iii), (\bar{s}, ζ_1) and (\bar{s}, ζ_2) cannot both be initial Markov states. Suppose that (\bar{s}, ζ_2) , say, is an initial Markov state of the economy where individuals of age $0 < a \leq G$ are endowed with exogenously given financial wealth $(\bar{w}^{ha})_{h \in H}$. Proposition 2(ii) guarantees that for (\bar{s}, \bar{w}) we have $\bar{w} \in W_{\bar{s}, \omega}^*$. As in Case 1, we look at the H individuals born at the node on \tilde{S} predecessor to the one where (\bar{s}, ζ_1) has realized, s_{01} . Their wealth when their age is $a = 1$ is their wealth at the critical state, and it is a function of the equilibrium price process over the tree spanning their finite life, that is, of $\underline{\psi} \in P(\omega)$. It is compared with the exogenous wealth \bar{w}^1 of individuals of the same age at state (\bar{s}, ζ_2) , and the difference is $f_{(s_{01}, \bar{s})}(\underline{\psi}, \omega, \bar{w}^1)$, which is defined in Section 4.2. If (\bar{s}, ζ_k) , $k = 1, 2$, is an initial critical pair, then $f_{(s_{01}, \bar{s})}(\underline{\psi}, \omega, \bar{w}^1) = 0$ for some $\underline{\psi} \in P(\omega)$ and some s_{01} . However, this contradicts Proposition 3.

Hence, the Markov equilibrium we constructed is void of both initial critical pairs and critical pairs of Markov states. Thus, the Markov states $(s, \zeta) \in Z_\omega$ are one-to-one in (s, w) ; equivalently, there exists a function $(p, \lambda)(s, w)$ such that each Markov state z can be written as $(s, w, (p, \lambda)(s, w))$. Equivalently, the transition function of the simple Markov equilibrium can be decomposed as $T_s(z) = [T_s^r(s, w), (p, \lambda)(T_s^r(s, w))]$. Thus, the recursive equilibrium of the economy $\omega \in \Omega^*$ is the recursive state space Z_ω^r and transition and endogenous variables functions T^r, ξ^r defined as

$$\begin{aligned} Z_\omega^r &= \text{proj}_{S \times W} Z_\omega, \quad \text{with } W_{s, \omega}^{Z_\omega^r} \supset W_{s, \omega}^*, \quad \text{for all } s, \\ T^r : Z_\omega^r &\rightarrow (Z_\omega^r)^S, \\ \xi^r : Z_\omega^r &\rightarrow \Xi \quad \text{is } \xi^r(z) = \xi(s, w, (p, \lambda)(s, w)), \end{aligned}$$

and it has the desired properties, ending the proof. *Q.E.D.*

Of course, at this stage nothing is said about any regularity property of the transition or the endogenous variables functions, an important topic for future research.

4.4. Perturbation Analysis

In this technical subsection we prove Proposition 4. Proposition 4 asserts that while facing different prices, at least one among H individuals will typically have different expenses on a subset of goods—those purchased after a certain date–event. If we look at just one individual, and put no qualification on what

“different prices” means, this may not be true. Consider, for example, a Walrasian individual with log-linear utility over three commodities. The individual will spend the same on commodity one, even if he faces different prices for commodities two—and three—provided that prices of commodities two and three are such that the value of the endowment is the same. We could perturb the utility and the endowments of this individual, but still obtain a region of different prices, yielding the same expense on commodity one. It is clear that we need to consider simultaneously many potentially different individuals and that the dimension of potential difference across individuals should be larger than the price dimension—Assumption A1.

We proceed as follows. Since openness and density are local properties, for each ω we need to define a superset of $P(\omega)$ which is locally independent of ω . This will allow perturbations of ω independent of prices. Then, to make the analysis as simple as possible, we transform the price space, the individual programming problems, and the equations $\hat{f}_\sigma(\cdot) = 0$ into a more convenient, but equivalent form. This first transformation suffices to prove Proposition 4 for $G = 1$. However, it will not be powerful enough to carry the result for $G > 1$.

Local Independence of the Price Set From ω : Pick $\omega \in \Omega$, and let $B_\varepsilon(\omega) \subset \Omega$ be an open ball centered at ω for some fixed $\varepsilon > 0$. Under the maintained assumptions, the optimal consumption-portfolio plans $(\underline{x}_k^h, \underline{\theta}_k^h)(\underline{\psi}_k, \omega)$ are continuous and then, by the strict monotonicity of preferences, there is a compact set of prices P such that $P(\omega') \subset P \subset \mathbb{R}_{++}^{(C-1)N_0} \times \mathbb{R}^{SN_0}$ for all $\omega' \in B_\varepsilon(\omega)$. Hence, by the compactness of P , for all h, k, s^a , and $(\underline{\psi}, \omega') \in P \times B_\varepsilon(\omega)$, $x_{k,s^a}^{ha}(\underline{\psi}_k, \omega') \in \bar{X} \subset \mathbb{R}_+^C$ —a compact set—and since preferences satisfy the boundary condition, \bar{X} is contained in the interior of the positive cone, that is, $\bar{X} \subset \mathbb{R}_{++}^C$.

What is open and dense in $B_\varepsilon(\omega)$ for any arbitrary such $B_\varepsilon(\omega)$ is open and dense in Ω . Therefore, to keep notation simple, hereafter we identify Ω with the arbitrary $B_\varepsilon(\omega)$.

Transforming the Programming Problems and the Price Space: In our economy, assets pay off in the numéraire commodity and their payoff matrix \mathbf{D} is invertible. Thus, individuals face sequentially complete asset markets. Therefore, the sequence of budget constraints can be compressed into a single one, getting rid of asset prices and portfolios. Hence, we change the price space to

$$P' = \{\underline{p}' \in \mathbb{R}_{++}^{2CN_0} : p'_{1,k,s_0} = 1, k = 1, 2\}.$$

We introduce the operator $\mathbb{E}_{s_t}^\delta$, which applies to any finite process $\underline{L}^a = (L_{t+a'})_{a'=0}^{G-a}$ of N_a histories and is defined as $\mathbb{E}_{s_t}^\delta(\underline{L}^a) = \mathbb{E}_{s_t}\{\sum_{a'=0}^{G-a} \delta^{a'} L_{t+a'}\}$; for simplicity, for $a = 0$ we omit the superscript from \underline{L}^a . The programming problems of the individuals born at s_{0k} are

$$(5) \quad \max \mathbb{E}_{s_{0k}}^\delta \{u^h(\underline{x})\} \quad \text{s.t.} \quad \mathbb{E}_{s_{0k}}^\delta \{\underline{p}'_k(\underline{x} - \underline{e}_k^h)\} = 0.$$

We need to reformulate the programming problems of individuals of age $a^* = 1, \dots, G$ who have wealth $w_{k,\bar{s}^1}^{ha^*} = d_{\bar{s}} \theta_{s_{0k}}^{h(a^*-1)}$ at $\bar{s}^1 = (s_{0k}, \bar{s})$. Since p_{k,\bar{s}^1} is normalized, while p'_{k,\bar{s}^1} is not, the prices in the budget constraint at \bar{s}^1 for these individuals must be divided by $\delta \pi(\bar{s}^1 | s_{0k}) p'_{1,k,\bar{s}^1}$, and their programming problem is

$$(6) \quad \max \mathbb{E}_{\bar{s}^1}^\delta \left\{ u^h(\underline{x}^{a^*}) \right\} \quad \text{s.t.} \quad \mathbb{E}_{\bar{s}^1}^\delta \left\{ \frac{\underline{p}'_k}{p'_{1,k,\bar{s}^1}} (\underline{x}^{a^*} - \underline{e}_k^{ha^*}) \right\} = w_{k,\bar{s}^1}^{ha^*}.$$

Notice that by the definition of the operator $\mathbb{E}_{s_{0k}}^\delta$, the coefficient that multiplies $u^h(x_{s^a})$ in the objective function of problem (5) coincides with the coefficient that multiplies $p'_{k,s^a}(x_{s^a} - e_{k,s^a}^h)$ in the budget constraint, and the same applies (modulo changing p'_{k,s^a} with $p'_{k,s^a}/p'_{1,k,\bar{s}^1}$) for problem (6).

By Arrow's equivalence theorem, to each $\underline{\psi} \in P^2$ corresponds a unique pair $\underline{p}' \in P'$ such that $\underline{\psi}$ and \underline{p}' are equivalent: the consumption bundles that satisfy the sequential budget constraints (1c) at $\underline{\psi}$ coincide with the consumption bundles that satisfy at \underline{p}' the single budget constraint of (5) (and hence of (6)). We take P' to be the set of processes \underline{p}' equivalent to price pairs $\underline{\psi} \in P^2$, which we refer to as Arrow prices; the two sets have obviously identical dimension $2CN_0 - 2$, the right-hand side of the inequality in Assumption A1. As we took P^2 to be a compact set, P' is also a compact subset of $\mathbb{R}_{++}^{2CN_0-2}$. Also, we take $P'(\sigma; \omega) \subset P'$ to be the set of Arrow prices \underline{p}' equivalent to the prices $\underline{\psi} \in P(\sigma; \omega)$.

Transforming the Wealth Equations: We reformulate the wealth equation $\hat{f}_\sigma(\cdot) = 0$ without making reference to portfolios. For $\underline{p}' \in P'(\sigma; \omega)$, let $x_{k,s^a}^{ha}(\underline{p}_k^1, \omega)$ be the optimal solutions at (k, s^a) to problems (5). Then, in analogy with the form of the budget constraints (6), we get that the wealth functions are

$$w_{k,\bar{s}^1}^{h1}(\underline{p}'_k, \omega) = \mathbb{E}_{\bar{s}^1}^\delta \left(\frac{\underline{p}'_k}{p'_{1,k,\bar{s}^1}} (\underline{x}_k^{h1}(\underline{p}_k^1, \omega) - \underline{e}_k^{h1}) \right)$$

for $k = 1, 2$, and we define the functions

$$f_\sigma^h(\tilde{p}^{h0}, \omega) = w_{1,\bar{s}^1}^{h1}(\underline{p}'_1, \omega) - w_{2,\bar{s}^1}^{h1}(\underline{p}'_2, \omega)$$

and let $f_\sigma = (f_\sigma^h)_{h \in H}$. We are ready to prove Proposition 4 for $G = 1$.

4.4.1. $G = 1$

Remember that for $G = 1$, the endogenous Markov state is reduced to (w, p) . If $C = 1$, simple Markov equilibria are already recursive. Therefore, let $C > 1$ and let

$$P'(\sigma; \omega) = P'(\bar{s}) = \left\{ \underline{p}' \in P' : \left\| \frac{p'_{1,\bar{s}^1}}{p'_{1,1,\bar{s}^1}} - \frac{p'_{2,\bar{s}^1}}{p'_{1,2,\bar{s}^2}} \right\| \neq 0 \right\},$$

and for any integer $n > 0$, let

$$P^n(\bar{s}) = \left\{ \underline{p}' \in P' : \left\| \frac{p'_{1,\bar{s}^1}}{p'_{1,1,\bar{s}^1}} - \frac{p'_{2,\bar{s}^1}}{p'_{1,2,\bar{s}^1}} \right\| \geq \frac{1}{n} \right\}.$$

Obviously, $P^n(\bar{s}) \subset P'(\bar{s})$, and both $P^n(\bar{s})$ and $P'(\bar{s})$ are sets that are (locally) independent of ω . Let Ω_σ^n denote the subset of Ω where system $f_\sigma(\underline{p}', \omega) = 0$ does not have a solution in $P^n(\bar{s})$. If Ω_σ^n is open and dense in Ω , then

$$\Omega^* = \bigcap_{n>0} \bigcap_{\sigma} \Omega_\sigma^n$$

is the intersection of a countable family of open and dense sets; therefore, it is a residual set of Ω (hence, it contains a dense subset), where system $\hat{f}_\sigma(\underline{p}', \omega) = 0$ does not have a solution in $\bar{P}'(\bar{s})$ for all σ . Suppose not. Then there is $\omega^* \in \Omega^*$, σ , and $\underline{p}' \in \bar{P}'(\bar{s})$ such that $f_\sigma(\underline{p}', \omega^*) = 0$. By definition of $P'(\bar{s})$, there must be $\hat{n} > 0$ such that $\underline{p}' \in P^{\hat{n}}(\bar{s})$. However, the latter implies that $\omega^* \notin \Omega_\sigma^{\hat{n}}$, a contradiction.

To show that Ω_σ^n is open, pick $\omega \in \Omega_\sigma^n$. The compactness of $P^n(\bar{s})$ implies that $|f_\sigma(\underline{p}', \omega)| \geq \eta$ for some $\eta > 0$ and all $\underline{p}' \in P^n(\bar{s})$. However, the map f_σ is continuous in all its argument and hence $|f_\sigma(\underline{p}', \omega')| > 0$ for all $\underline{p}' \in P^n(\bar{s})$ and ω' in an open neighborhood of ω . Thus, the set Ω_σ^n is open.

We now show that Ω_σ^n is dense. It suffices to prove that the Jacobian matrix $D_e f_\sigma(\underline{p}'; e, u, \delta)$ is a surjection for all $(\underline{p}', e, u, \delta) \in P'(\bar{s}) \times \Omega$ (and therefore, in $P^n(\bar{s}) \times \Omega$). If this is the case, by Assumption A1, $\dim P^n(\bar{s}) < H$ and there are more equations than unknowns in $f_\sigma(\underline{p}', \omega) = 0$. Hence, by the preimage and the transversality theorems, there is a dense subset E_σ^n of E (so Ω_σ^n of Ω) where $f_\sigma(\underline{p}', \omega) = 0$ has no solution in $P^n(\bar{s})$.

When $\underline{p}' \in P^n(\bar{s})$, the vectors p'_{k,\bar{s}^1} , $k = 1, 2$, are linearly independent. Therefore, we can find an appropriate perturbation $\Delta e_{\bar{s}}^1$ of $e_{\bar{s}}^1$ such that $p'_{1,\bar{s}^1} \Delta e_{\bar{s}}^1 = 1$, while $p'_{2,\bar{s}^1} \Delta e_{\bar{s}}^1 = 0$. We show that $D_{e_{\bar{s}}^1} f_\sigma$, the directional derivative of f_σ in the direction identified by the perturbation $\Delta e_{\bar{s}}^1$, is a surjection. First, observe that this perturbation does not affect $w_{2,\bar{s}^1}^{h1}(\underline{p}_2', \omega)$ and, hence,

that $D_{\vec{e}_{\bar{s}}^1} f_\sigma = D_{\vec{e}_{\bar{s}}^1} w_{1,\bar{s}^1}^{h1}$. Second, differentiate the first-order conditions to problem (5) for $k = 1$, drop h and k , and get

$$(7a) \quad H_{s^a} \Delta x_{s^a} - p'_{s^a} \Delta \lambda = 0,$$

$$(7b) \quad \sum_a \delta^a \sum_{s^a} \pi(s^a | s_0) p'_{s^a} \Delta x_{s^a} = \delta \pi(\bar{s}^1 | s_0),$$

where H_{s^a} is the invertible Hessian at x_{s^a} , a negative definite matrix, and the superscript T stands for transposed. Let $Q_{s^a} = p'_{s^a} H_{k,s^a}^{-1} p'_{k,s^a}^T < 0$ and $Q = \sum_{a,s^a} \delta^a \pi(s^a | s_0) Q_{s^a} < 0$. We get

$$\Delta \lambda = \frac{\delta \pi(\bar{s}^1 | s_0)}{Q}.$$

Differentiating the map $w_{1,\bar{s}^1}(\underline{p}'_1, \omega)$, we obtain

$$\Delta w_{\bar{s}^1} = (Q_{\bar{s}^1} \Delta \lambda - 1) \frac{1}{p'_{1,1,\bar{s}^1}} = \left(\frac{\delta \pi(\bar{s}^1 | s_0) Q_{\bar{s}^1}}{Q} - 1 \right) \frac{1}{p'_{1,1,\bar{s}^1}} < 0.$$

The argument is concluded by observing that $D_{\vec{e}_{\bar{s}}^1} f_\sigma^{h'} = 0$ for all h, h' with $h \neq h'$.

The microeconomics of the result is clear. The normality of all expenditures $p'_{s^a} x_{s^a}$, a by-product of separability of utility, implies that all of them move proportionally to, but less than, a lifetime wealth change. Thus, the change in $w_{\bar{s}^1}$ induced by the perturbation $\Delta e_{\bar{s}}^1$ is negative, since $w_{\bar{s}^1} = p'_{\bar{s}^1} / p'_{1,\bar{s}^1} (x_{\bar{s}^1}^1 - e_{\bar{s}}^1)$.

4.4.2. $G > 1$

When $G > 1$, multiplier inequalities (4) must be taken into account when defining the set of prices $P'(\sigma; \omega)$. Such inequalities can be generated by price differences across k trees arising at nodes s^a with $a > 1$. Relative to the case when $G = 1$, things are then considerably complicated by stationarity: endowment or utility perturbations at one node reverberate across the trees, possibly rendering the perturbations ineffective, that is, the derivative of the wealth differentials is zero. Fortunately, we show below that when perturbations are ineffective it is because the price differences are only “nominal,” that is, only due to labeling, and do not translate to differences in multipliers. To this extent it will be essential to identify equivalence classes of Arrow prices which determine $P'(\sigma, \omega)$, making the effectiveness of the available perturbations immediately apparent and $P'(\sigma, \omega)$ (locally) ω -independent. We then show how to bypass the possibility that the derivative of the wealth differentials is zero by introducing a nesting technique.

Creating Equivalence Classes of Arrow Prices: This section formalizes the necessary condition for $\underline{p}' \in P'(\sigma; \omega)$ by (a) identifying equivalence classes of Arrow prices which determine $\underline{p}' \in P'(\sigma; \omega)$, irrespective of ω , and (b) creating compact sets of prices where the condition holds.

(a) *Identifying equivalence classes of Arrow prices.* We start by making a simple, preliminary observation. Recall that by assumption, all individuals of all ages have identical wealths at (k, \bar{s}^1) , that is, $w_{1,\bar{s}^1}^{ha} = w_{2,\bar{s}^2}^{ha} \equiv w^{ha}$ for all h, a . If for all possible specifications of the economy ω' and such wealth w^{ha} , the (optimal) marginal utilities $\lambda_{1,\bar{s}^1}^{ha*}(\underline{p}'_k, \omega', \bar{w}^{ha*})$ in problems (6) computed at \underline{p}' are k -invariant for all h and $a^* \geq 1$, and $p'_{1,\bar{s}^1}/p'_{1,1,\bar{s}^1} = p'_{2,\bar{s}^2}/p'_{1,2,\bar{s}^1}$, then $\underline{p}' \notin P'(\sigma, \omega)$. Thus, a necessary condition for $\underline{p}' \in P'(\sigma, \omega)$ is that either $p'_{1,\bar{s}^1}/p'_{1,1,\bar{s}^1} \neq p'_{2,\bar{s}^2}/p'_{1,2,\bar{s}^1}$ or $\lambda_{1,\bar{s}^1}^{ha*}(\underline{p}'_1, \omega, \bar{w}^{ha*}) \neq \lambda_{2,\bar{s}^1}^{ha*}(\underline{p}'_2, \omega, \bar{w}^{ha*})$ for some $a^* \geq 1$, and some ω, \bar{w}^{ha*} . In the absence of restrictions on the fundamentals, this necessary condition translates into $p'_{1,s^a}/p'_{1,1,\bar{s}^1} \neq p'_{2,s^a}/p'_{1,2,\bar{s}^1}$ for some $s^a \geq \bar{s}^1$.

However, in our economies, the cardinality indices u^h are state and age invariant, endowments are stationary, and, by assumption, wealth is k -invariant. The pair of price processes $\underline{p}' = (\underline{p}'_1, \underline{p}'_2)$ can take different values at pairs of identical nodes $s^a \geq \bar{s}^1$ on the two trees, but they will still generate the same values for $\lambda_{k,\bar{s}^1}^{ha}(\underline{p}'_k, \omega', \bar{w}^{ha})$ for all h, a , and (ω', w^{ha}) if for each period $a \geq 1$, $s^a \geq \bar{s}^1$, and price realization p , the (discounted) probabilities that $p'_{k,s^a}/p'_{1,k,\bar{s}^1} = p$ are independent of k and the overall wealth of individuals (h, a^*) is k -invariant. The next example makes this point transparent.

EXAMPLE: Consider an economy with $G = 3$ and $S = \{\alpha, \beta\}$, where $\pi(s|s') = \frac{1}{2}$ for all s, s' , $s_{0k} = \alpha$ for all k , and $\bar{s} = \alpha$, $\delta = 1$, and wealth is k -invariant. Pick \underline{p}' such that the following statements hold:

- p'_{k,s^a} is k -invariant for $a \leq 2$.
- $p'_{1,(s^2,\alpha)} = p'_{2,(s^2,\beta)} = p^1$ while $p'_{1,(s^2,\beta)} = p'_{2,(s^2,\alpha)} = p^2$ for $s^2 = (\alpha, \alpha, \alpha), (\alpha, \beta, \beta)$ and vice versa.
- $p'_{1,(s^2,\alpha)} = p'_{2,(s^2,\beta)} = p^2$ while $p'_{1,(s^2,\beta)} = p'_{2,(s^2,\alpha)} = p^1$ for $s^2 = (\alpha, \beta, \alpha), (\alpha, \beta, \beta)$.

If $p^1 \neq p^2$, then $\underline{p}'_1 \neq \underline{p}'_2$. However, this difference is just a matter of relabeling the states and has no real consequences. Indeed, by wealth k -invariance and k -invariance of p'_{k,s^a} , $a \leq 2$, the multipliers $\lambda_{k,\bar{s}^1}^{ha*}$ associated to problems (6) for individuals of age $a^* \geq 2$ are k -invariant. Furthermore, by the definition of \underline{p}' and the assumptions on endowments and on conditional probabilities, $\mathbb{E}_{s_{0k}}^{\delta}(\underline{p}'_k e^h)$ is k -invariant. The latter implies that, for all u^h satisfying the maintained assumptions, the optimal solutions \underline{x}_k^h , $k = 1, 2$, to (5) satisfy the following statements:

- x_{k,s^a}^{ha} is k -invariant for $a \leq 2$.

- $x_{1,(s^2,\alpha)}^{h3} = x_{2,(s^2,\beta)}^{h3}$ while $x_{1,(s^2,\beta)}^{h3} = x_{2,(s^2,\alpha)}^{h3}$ for $s^2 = (\alpha, \alpha, \alpha), (\alpha, \alpha, \beta)$.
- $x_{1,(s^2,\alpha)}^{h3} = x_{2,(s^2,\beta)}^{h3}$ while $x_{1,(s^2,\beta)}^{h3} = x_{2,(s^2,\alpha)}^{h3}$ for $s^2 = (\alpha, \beta, \alpha), (\alpha, \beta, \beta)$.

It follows that λ_k^{h0} and w_k^{h1} are k -invariant, so that the wealth invariance condition is satisfied. Finally, $\lambda_{k,\bar{s}^1}^{h1}$ is k -invariant. Thus, \underline{p}' is not an element of $P'(\sigma; \omega)$ for any ω .

Next, we make the observations of this example general. By strict concavity of u^h , if $p'_{1,\hat{s}^a}/p'_{1,1,\bar{s}^1} = p'_{2,\hat{s}^a}/p'_{1,2,\bar{s}^1}$ for two distinct histories $\hat{s}^a \succeq \bar{s}^1$ and $\bar{s}^{a'} \succeq \bar{s}^1$ at ages a and a' , then at the optimal solution to (6), $x_{k,\hat{s}^a}^{h(a^*+a-1)} = x_{k,\bar{s}^{a'}}^{h(a^*+a'-1)}$. This allows us to rewrite both problems (6) by expressing prices in terms of their distinct realizations rather than in terms of their realizations at each (k, s^a) , $s^a \succeq \bar{s}^1$.

For $\underline{p}' \in P'(\sigma, \omega)$, let

$$\mathbb{P}^1 = \left\{ p \in \mathbb{R}_{++}^C : \frac{p'_{k,s^a}}{p'_{1,k,\bar{s}^1}} = p, \text{ for some } k, s^a \succeq \bar{s}^1 \right\}$$

with cardinality $\mathbb{P}^1 \leq 2^{\sum_{a=0}^{G-1} S^a}$ and generic element $p(\ell)$, where \mathbb{P}^1 denotes also the set of price indices ℓ .

For $\ell \in \mathbb{P}^1$ and $a \geq 1$, we define sets of histories of length a and their probability weights as

$$S_k^a(\ell) = \left\{ s^a \succeq \bar{s}^1 : \frac{p'_{k,s^a}}{p'_{1,k,\bar{s}^1}} = p(\ell) \right\}, \quad \Pi[S_k^a(\ell)] = \sum_{s^a \in S_k^a(\ell)} \pi(s^a | \bar{s}^1),$$

where $\Pi[S_k^a(\ell)] = 0$ if $S_k^a(\ell) = \emptyset$. To make the programming problems dependent only on the distinct realizations of the price processes $p'_{k,s^a}/p'_{1,k,\bar{s}^1}$, $s^a \succeq \bar{s}^1$, define

$$(8) \quad \Pi_k(a^*, \ell) = \sum_{a=1}^{G+1-a^*} \delta^{a-1} \Pi[S_k^a(\ell)], \quad a^* \geq 1.$$

To make transparent the overall endowment value (as well as the effectiveness of endowment perturbation) on the two subtrees (k, s^a) , $s^a \succeq \bar{s}^1$, define

$$p_{k+}(a, s) = \sum_{s^{a-1}:(s^{a-1}, s) \succeq \bar{s}^1} \pi(s^{a-1}, s | \bar{s}^1) \frac{p'_{k,(s^{a-1}, s)}}{p'_{k,1,\bar{s}^1}},$$

with $p_{k+}(a, s) = 0$ if $a = 0$ or $a = 1$ and $s \neq \bar{s}$. By definition of $\Pi_k(a^*, \ell)$ and $p_{k+}(a, s)$, and by strict concavity of u^h , problems (6), $1 \leq a^* \leq G$, can be equiv-

alently written as

$$(9) \quad \max_{\ell \in \mathbb{P}^1} \sum \Pi_k(a^*, \ell) u^h(x(\ell)) \quad \text{s.t.}$$

$$\sum_{\ell \in \mathbb{P}} \Pi_k(a^*, \ell) p(\ell) x(\ell) - \sum_{a=a^*}^G \delta^{a-a^*} \sum_s p_{k+}(a, s) e_s^{ha} = w_{k, \bar{s}^1}^{ha^*}.$$

A simple inspection of problems (9) delivers the necessary condition for $\underline{p}' \in P'(\sigma, \omega)$. Consider a pair \underline{p}' such that $p_{k+}(a, s)$ are k -invariant for all (a, s) , and $\Pi[S_k^a(\ell)]$ are k -invariant for all ℓ and $a \geq 1$. Then, by (8), also $\Pi_k(a^*, \ell)$ are k -invariant for all ℓ and $a^* \geq 1$, and then the two problems (9) are identical at \underline{p}' and so are their optimal solutions for all (h, a^*) . Thus, $\lambda_{k, \bar{s}^1}^{ha^*}(\underline{p}'_k, \omega, w^{ha^*})$ are k -invariant. Hence, the necessary condition for $\underline{p}' \in P'(\sigma, \omega)$ simply is

$$(\text{NC}) \quad \|((\Pi[S_1^a(\ell)])_{a, \ell}, (p_{1+}(a, s))_{a, s}) - ((\Pi[S_2^a(\ell)])_{a, \ell}, (p_{2+}(a, s))_{a, s})\| \neq 0.$$

Since $\pi(s^a | \bar{s}^1) = \pi(s^a | \bar{s})$ for all $s^a \succeq \bar{s}^1$, inequalities (NC) are independent of s_{0k} and ω , and we let $P'(\bar{s})$ denote the set of \underline{p}' satisfying condition (NC).

However, since we need to perturb the map f_σ , we need to make sure that at $\underline{p}' \in P'(\bar{s})$, $\Pi_1(1, \ell) \neq \Pi_2(1, \ell)$; otherwise, individuals of age 0 at s_{0k} may be solving identical programming problems. Thus, consider the set $P'(\bar{s}, \delta)$ of prices that satisfy the δ -dependent conditions:

$$(10) \quad \|((\Pi_1(1, \ell))_\ell, (p_{1+}(a, s))_{a, s}) - ((\Pi_2(1, \ell))_\ell, (p_{2+}(a, s))_{a, s})\| \neq 0.$$

Obviously, $P'(\bar{s}, \delta) \subset P'(\bar{s})$. We show below that, generically in δ , if the inequality $\Pi[S_1^a(\ell)] \neq \Pi[S_2^a(\ell)]$ holds true for some ℓ , then $\Pi_{1+}(1, \ell) \neq \Pi_{2+}(1, \ell)$. The latter has two desirable implications: $P'(\bar{s}, \delta) = P'(\bar{s})$ and, therefore, $P'(\bar{s}, \delta)$ is ω -invariant in the generic set of common discount factors.

LEMMA 5: *There exists an open and dense subset Ω' of Ω such that for all $\omega \in \Omega'$, $P'(\bar{s}, \delta) = P'(\bar{s})$.*

See the Appendix for the proof.

(b) *Creating compact sets of prices.* In the analysis for $G = 1$, we proved that Ω^* is residual by defining compact regions of the price domain $P^n(\bar{s}) \subset P'(\bar{s})$. For $\underline{p}' \in P^n(\bar{s})$, the effective difference between \underline{p}'_1 and \underline{p}'_2 is sizeable, a necessary condition for establishing openness of the sets Ω_σ^n . We have to repeat that maneuver. The notion of effective price difference is embedded in the definition of the set $P'(\bar{s})$ and it is precisely defined by (NC) or, equivalently, (10). What does sizeable mean? The details are in the proof

of the next lemma, but here is the precise idea. Since we are limiting attention to the set Ω' , for $\underline{p}' \in P'(\bar{s})$, either $p_{1+}(a, s) \neq p_{2+}(a, s)$ for some a, s or $\Pi_1(1, \ell) \neq \Pi_2(1, \ell)$ for some ℓ . However, the values $\Pi_k(1, \ell)$ depend on the distinct realizations of $p'_{k,s,a}$, $s^a \geq \bar{s}^1$, but not on the values $p(\ell)$ of these realizations. Thus, differences in probabilities $\Pi_k(1, \ell)$ may coexist with distinct, but arbitrarily close, values $p(\ell)$ of price realizations. Obviously, if this is the case and if $p_{1+}(a, s) = p_{2+}(a, s)$ for all (a, s) , individuals of age $a = 1$ face at \bar{s}^1 arbitrarily close price systems on the two trees. Thus, for $\omega \in \Omega'$, we say that in a subset of $P'(\bar{s})$, the difference between \underline{p}'_1 and \underline{p}'_2 is sizeable if the values of either $\|(p_{1+}(a, s))_{a,s} - (p_{2+}(a, s))_{a,s}\|$ or $\|p(\ell) - p(\ell')\|$ for some pair $\ell \neq \ell'$ are uniformly bounded away from zero by some positive constant.

LEMMA 6: *There exists a countable collection $\{P^n(\bar{s})\}_{n=1}^{+\infty}$ of compact subsets of $P'(\bar{s})$ such that (i) if $\underline{p}' \in P^n$, the effective difference between \underline{p}'_1 and \underline{p}'_2 is sizeable, (ii) $P^n(\bar{s}) \subset P^{n+1}(\bar{s})$, and (iii) $\bigcup_n P^n(\bar{s}) = \text{cl}(P'(\bar{s}))$.*

See the [Appendix](#) for the proof.

The Nesting Technique: A direct use of the transversality theorem to obtain $f_\sigma \neq 0$ requires the functions $f_\sigma^h(\underline{p}', \omega)$ to have nonzero derivatives with respect to ω^h for all h and for all $\underline{p}' \in P'(\bar{s})$. The analysis would be relatively straightforward if we could find perturbations that disturb the optimal solution on one tree without affecting it on the second. Indeed, this was the essence of the argument for $G = 1$. Unfortunately, for $G \geq 2$ these perturbations are not available in some regions of $P'(\bar{s})$ and in these regions it can be $D_{\omega^h} f_\sigma^h = 0$. However, we are still able to show that $f_\sigma^h \neq 0$ on $P'(\bar{s})$ for a generic set of parameters and for some h . The argument is based on a nesting technique. We first lay out its general mathematical structure and later we will apply it to our problem.

Let Ω'' be an open subset of Ω' and let \mathfrak{F} be a finite family of real-valued maps f_j^h with domain $P'(\bar{s}) \times \Omega''$, with $h \in H$ and $j = 1, \dots, J$, where J denotes also the set of indices. The maps f_j^h are assumed to be continuous differentiable and to satisfy $D_{\omega^{h'}} f_j^h(\underline{p}', \omega) = 0$ for all j and $h' \neq h$. Let

$$E_j^h = \{(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega'': D_{\omega} f_j^h(\underline{p}', \omega) \neq 0 \text{ or } f_j^h(\underline{p}', \omega) \neq 0\}$$

and

$$N_j^h = \{(\underline{p}', \omega) \in E_j^h : f_j^h(\underline{p}', \omega) \neq 0\} \quad \text{for all } j$$

and define the following two conditions on the maps f_j^h :

CONDITION UNIVERSAL: For all h , $P'(\bar{s}) \times \Omega'' \subset E_1^h \cup E_J^h$.

CONDITION NESTING: For all h , $N_j^h \cap N_{j'}^h \subset \bigcup_{j' \leq j} E_{j'-1}^h$, for all $j > 1$.

We call the first Condition **Universal** because the family satisfies a nonzero property for all prices and parameters: at any (\underline{p}', ω) we can extract from the family \mathfrak{F} an auxiliary system of H maps, one for each h , which are nonzero or have nonzero derivative. We call the second Condition **Nesting** because it allows us to nest the f_1 functions into a cascade of auxiliary systems of maps in \mathfrak{F} . The following result suffices for our analysis.

LEMMA 7: *If \mathfrak{F} satisfies Conditions Universal and Nesting, then there exists a dense set $\bar{\Omega} \subset \Omega''$ such that for all $(\underline{p}', \omega) \in P'(\bar{s}) \times \bar{\Omega}$, there exists h with $f_1^h(\underline{p}', \omega) \neq 0$.*

For the proof see the Appendix.

We sketch the reasoning. What is $\bar{\Omega}$? Consider the set of maps g that assign to each individual h a function $f_{g(h)}^h$ from \mathfrak{F} . Let $f_g = (f_{g(h)}^h)_{h \in H}$,

$$E_g = \bigcap_h E_{g(h)}^h, \quad \text{and} \quad N_g = \bigcup_h N_{g(h)}^h.$$

A straightforward application of the transversality theorem implies that for each g there exists a dense subset Ω_g of Ω'' such that for all $(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega_g$, either $(\underline{p}', \omega) \notin E_g$ or $f_g(\underline{p}', \omega) \neq 0$; equivalently, $(\underline{p}', \omega) \in N_g$. We then let $\bar{\Omega} = \bigcap_g \Omega_g$. Notice that by construction $P'(\bar{s}) \times \bar{\Omega} \cap E_g = N_g$ for all g . If $(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega_g$, $f_g(\underline{p}', \omega) \neq 0$ only if $(\underline{p}', \omega) \in E_g$, but apparently nothing excludes the possibility that $(\underline{p}', \omega) \notin E_g$. So why does $\bar{\Omega}$ work? Here, the nesting technique kicks in. Condition **Universal** states that $(\underline{p}', \omega) \in E_g$ for some g with $g(h) = 1$ or $g(h) = J$ for all h . Let $H_g = \{h : f_{g(h)}^h(\underline{p}', \omega) \neq 0\}$ and suppose that $g(h) = J$ for all $h \in H_g$; otherwise, the argument is concluded. Condition **Nesting** now states that $(\underline{p}', \omega) \in E_{g'}$ and, hence, in $N_{g'}$, for some g' with $g'(h) = g(h)$, for $h \in H \setminus H_g$, while $g'(h) < J$ for $h \in H_g$ and $(\underline{p}', \omega) \in N_J^h \cap N_{g'(h)}^h$. Once again Condition **Nesting** can be applied to move to an auxiliary system $f_{g'}$ with $g' < g$, and iterating finitely many times, eventually we reach the desired conclusion that $(\underline{p}', \omega) \in N_{g^*}$ with $g^*(h) = 1$ for some $h \in H_{g^*}$.

We are now ready to prove Proposition 4 for $G > 1$. First, we exploit the equivalence classes of Arrow prices to prove that the following family of maps \mathfrak{F} satisfies Conditions **Universal** and **Nesting**. Define \mathfrak{F} to be the following family of maps (remember that $N_0 = \sum_{a=0}^G S^a$):

- $f_1^h(\underline{p}', \omega) = f_\sigma^h(\underline{p}', \omega)$.
- $f_{s^a}^h(\underline{p}', \omega) = x_{s^a_0}^{h0}(\underline{p}'_1, \omega) - x_{s^a_2}^{ha}(\underline{p}'_2, \omega)$, $s^a \in \bigcup_{a=0}^G S^a$.

$$\bullet \quad f_{2+N_0}^h(\underline{p}', \omega) = \lambda_1^h(\underline{p}'_1, \omega) - \lambda_2^h(\underline{p}'_2, \omega).$$

Put the set of nodes $s^a \in \bigcup_{a=0}^G S^a$ in a one-to-one correspondence with $\{2, \dots, 1 + N_0\}$ and index them by j . All maps are continuously differentiable and obviously satisfy the condition $D_\omega f_j^h(p', \omega) = 0$ for all j and $h' \neq h$. The following lemma proves that our choice of \mathfrak{F} also satisfies Conditions **Universal** and **Nesting** on a dense subset Ω'' of Ω' .

LEMMA 8: *There exists an open and dense set $\Omega'' \subset \Omega'$ such that the family of maps $\mathfrak{F} = (f_j^h)_{j \geq 1}$ satisfies Conditions **Universal** and **Nesting**.*

For the proof see the [Appendix](#).

To gain intuition on this issue, notice three aspects of our choice. First, the construction of $P'(\bar{s})$ (i.e., of the equivalence classes of Arrow price pairs) implies that on $P'(\bar{s}) \times \Omega'$, either $D_\omega f_1^h \neq 0$ or $D_\omega f_{2+N_0}^h \neq 0$; that is, it implies Condition **Universal**. Second, whenever $f_{2+N_0}^h \neq 0$, endowment or utility perturbations yield different changes in optimal consumption bundles across the two trees; that is, if $f_{2+N_0}^h \neq 0$, then either $D_\omega f_{s^a}^h \neq 0$ or $f_{s^a}^h \neq 0$, $s^a \in \bigcup_{a=0}^G S^a$, and Condition **Nesting** holds true for $j > 1$. Third, whenever $f_{s^a}^h \neq 0$, $s^a \in \bigcup_{a=0}^G S^a$, the optimal bundle $x_{s_0 1}^h(\underline{p}'_1, \omega)$ does not appear on the second tree, thereby allowing for perturbations of utilities that affect optimal solutions on the first, but not on the second tree. These perturbations are powerful enough to show that when $f_{s^a}^h \neq 0$, $s^a \in \bigcup_{a=0}^G S^a$, and $f_{2+N_0}^h \neq 0$, $D_\omega f_1^h \neq 0$, that is, Condition **Nesting** holds true also for $j = 1$.

Next, for given n , Ω_σ^n denotes the subset of Ω'' where the system of equations $f_\sigma(\underline{p}', \omega) = 0$ does not have a solution in $P^n(\bar{s})$, the set introduced in Lemma 6. By the same argument provided for $G = 1$, Ω_σ^n is open. Again, let $\Omega^* = \bigcap_{n>0} \bigcap_\sigma \Omega_\sigma^n$. If the sets Ω_σ^n are also dense, Ω^* is a residual set where $f_\sigma(\underline{p}', \omega) \neq 0$. However, $P^n(\bar{s}) \times \bar{\Omega} \subset P'(\bar{s}) \times \bar{\Omega}$ and, hence, by Lemma 7, $f_\sigma(\underline{p}', \omega) \neq 0$ for all $(\underline{p}', \omega) \in P^n(\bar{s}) \times \bar{\Omega}$. Thus, $\bar{\Omega} \subset \Omega_\sigma^n$, thereby concluding the argument.

5. EXTENSIONS

In computational applications, utility functions are parametrically given, most frequently in the constant relative risk aversion (CRRA) class. As already mentioned, when $G = 1$, Proposition 4 can be immediately established by perturbing only the endowments, therefore covering this class. However, when our density result is based on local utility perturbations, it does not immediately cover those economies, as perturbations now have to be parametric and cannot alter the utility functional form.

On the other hand, many of our simplifying assumptions on the economic environments can be dropped either without altering the results or by actually sharpening them. For example, the simple demographic structure can be

generalized going from a constant population process to any exogenous time-homogeneous finite Markov chain only with an increased notational burden. Utilities can be assumed to be state or age dependent, actually facilitating our proofs. Beliefs can differ across agents. We also have assumed that financial assets are short-lived and in zero net supply, that there is no production, and that all risk is aggregate. Our result clearly survives all such extensions.

Long-Lived Assets: We can add long-lived assets in positive supply (i.e., Lucas' trees), storage, and even production to our model as done in Rios Rull (1996) or Kubler and Polemarchakis (2004) without altering any of our results. Obviously now beginning-of-period financial wealth w has to be defined to include the value of long-lived assets held by the individuals and the value of stored commodities. The latter implies that the wealth of the individuals depends on all current prices, their portfolios of long-lived assets, and the stored amount of commodities. The state spaces of Markov and recursive equilibria have to be expanded, since they must now include portfolios of long-lived assets and amounts of stored commodities. If there is production, the capital distribution across firms also needs to be included as an endogenous state. Now, at a critical Markov pair, not only wealths, but also portfolios of long-lived assets, stored commodities, and capital must be invariant. Hence, if anything, by adding equations, these extensions can potentially weaken the degree of heterogeneity needed to rule out the existence of critical pairs and critical initial pairs of Markov states.

Idiosyncratic Risk: Recall that the issue here is not whether we can include idiosyncratic risk in our model, rather whether this inclusion can be used to substantially reduce the ex ante heterogeneity in Assumption A1. The results will depend on how one models idiosyncratic shocks. We just give a hint of how to carry out the analysis for what is the hardest case for our approach.

For each type h , there is a large number of individuals subject to individual shocks that affect endowments (such as unemployment, accidental loss risk, and so on). In each period the set of states of uncertainty is $S \times \Sigma$, with Σ denoting the set of individual states, and individual risks are independent and identically distributed. Ex ante identical individuals of age 0 enter the economy under different uninsured realizations of individual risk: this is the key feature that can be exploited to weaken Assumption A1. The maneuver comes, however, at a cost. To perturb independently the various functions of identical individuals born at different personal states, we need a richer set of perturbations: utilities have to be age and (aggregate) state dependent.

Competitive—and, therefore, Markov and recursive—equilibrium prices are affected by the realizations of the aggregate, but not of the individual states. Thus, the only exogenous variable entering the definition of a Markov state is $s \in S$. A Markov (or recursive) state must now specify, for each type and age, a distribution of wealth. The state space includes vectors of the form $w_{\sigma^a}^{ha}$, $\sigma^a \in \Sigma^a$. The definitions of confounding and nonconfounding equilibria, and critical and initial critical pairs are identical. However, the density argument in

Proposition 4 changes considerably and is more demanding. The essence is to perturb independently individuals indexed by the same h , but by different σ at birth. Since these individuals were endowed at birth with different endowments, their overall wealth and, hence, their consumption plans will typically be different. With state and age dependent utility perturbations, this is sufficient to establish that generically the system of $H\Sigma$ equations $f_{(s_{01}, s_{02}, \bar{s})} = 0$ cannot have a solution. In other words, Assumption A1 can be weakened:

$$\text{ASSUMPTION A1': } H\Sigma > 2[C \sum_{a=0}^G S^a - 1].$$

Therefore, even $H = 1$ is compatible with the existence of a recursive equilibrium if Σ is large enough, as we wanted to show.

APPENDIX

PROOF OF PROPOSITION 2: (i) Pick an economy $\omega \in \Omega$ and any family of subsets $O_s \subset W_{s,\omega}$, $s \in S$. Consider the family of competitive equilibrium processes $\tilde{\xi}$. They are parametrized by initial states (s_0, w_0, ξ_0) . Hereafter, let Z_0 denote the projection of the initial states of the family of competitive equilibrium processes onto the set of variables (s, w, p, λ) , that is, onto the (smaller) state space of the simple Markov equilibrium. For all $\bar{z} = (\bar{s}, \bar{w}, \bar{p}, \bar{\lambda}) \in Z_0$ with $\bar{w} \in O_{\bar{s}}$, select a unique competitive equilibrium, thereby describing a family of competitive equilibria $\tilde{\xi}_{\bar{z}}$ parametrized by such elements $\bar{z} \in Z_0$. We call such \bar{z} an initial condition. The construction of the simple Markov equilibrium is based on an observation that we state under the form of a separate claim. For each history s^t , $t \geq 1$, the competitive equilibrium realization at s^t , $\tilde{\xi}_{\bar{z},s^t} = [x, \theta, \lambda, \psi]_{\bar{z},s^t}$, and the corresponding financial wealth distribution, $w_{\bar{z},s^t} = d_{s^t} \theta_{\bar{z},s^{t-1}}$, define a Markov state $[s_t, (w, p, \lambda)_{\bar{z},s^t}]$. We call two competitive equilibria realizations $\tilde{\xi}_{z,\hat{s}^t}$ and $\tilde{\xi}_{\bar{z},\hat{s}^t}$ *Markov equivalent* if they generate the same Markov states (at \hat{s}^t and \bar{s}^t). Given two equilibrium processes $\tilde{\xi}_z$ and $\tilde{\xi}_{\bar{z}}$, and a pair of histories $\bar{s}^t, \hat{s}^t \in \tilde{S}$, $t' > 0$, such that $\tilde{\xi}_{\bar{z},\hat{s}^t}$ and $\tilde{\xi}_{z,\hat{s}^t}$ are Markov equivalent, we define a binary operation $[\hat{s}^t \lambda_{\bar{s}^t}] : \Xi^{\tilde{S}} \times \Xi^{\tilde{S}} \rightarrow \Xi^{\tilde{S}}$ called *grafting* and denote its result by $\tilde{\xi}_z[\hat{s}^t \lambda_{\bar{s}^t}] \tilde{\xi}_{\bar{z}}$. It is the process defined as

$$\{\tilde{\xi}_z[\hat{s}^t \lambda_{\bar{s}^t}] \tilde{\xi}_{\bar{z}}\}_{s^{t^*}} = \begin{cases} \tilde{\xi}_{z,s^{t^*}} & \text{for } s^{t^*} \not\sim \hat{s}^t, \\ \tilde{\xi}_{\bar{z},\bar{s}^{t'+\tau}} & \text{for } s^{t^*} = \hat{s}^{t+\tau}, \tau \geq 0. \end{cases}$$

Notice that the grafting operation $[\hat{s}^t \lambda_{\bar{s}^t}]$ can be applied to the same competitive equilibrium $\tilde{\xi}_z$ at two distinct Markov equivalent histories.

CLAIM 9: *For each pair of (not necessarily distinct) competitive equilibria $\tilde{\xi}_z$ and $\tilde{\xi}_{\bar{z}}$ with Markov equivalent realizations at \hat{s}^t and \bar{s}^t , the grafted process $\tilde{\xi}_z[\hat{s}^t \lambda_{\bar{s}^t}] \tilde{\xi}_{\bar{z}}$ is a competitive equilibrium process that starts from the initial condition $z \in Z_0$.*

The proof of Claim 9 is put off to the end of this argument. Claim 9 implies that we can apply the operator $[\cdot_{\tilde{s}^t} \wedge \cdot_{\tilde{s}'^t}]$ countably many times and still obtain a competitive equilibrium. This is what we do to construct a simple Markov equilibrium. We build the functions T and $\xi: Z \rightarrow \Xi$, and define the state space Z recursively as follows.

Start from $t = 0$. Drop the subscript ω , consider the set Z_0 and the selection of equilibrium processes $\xi_{\bar{z}}$, $\bar{z} \in Z_0$, $\bar{w} \in O_{\bar{s}}$, and start defining the endogenous map ξ :

- (i) Set $\xi(\bar{z}) \equiv \xi_{\bar{z}}$ for all $\bar{z} \in Z_0$.

Moving to $t = 1$, the competitive equilibrium $\tilde{\xi}_{\bar{z}}$ determines $\xi_{\bar{z}, s^1}$ and $w_{\bar{z}, s^1}$, $s^1 = (\bar{s}, s)$, for all $\bar{z} \in Z^0$. Therefore, it describes uniquely on Z_0 the continuation of \bar{z} , that is, the S Markov states $z' = (s, \zeta_{\bar{z}, s^1})$, $s \in S$, and $\zeta = (w, p, \lambda)$, immediately following \bar{z} .

- (ii) Set $T_s(\bar{z}) = (s, \zeta_{\bar{z}, s^1})$ for $\bar{z} \in Z_0$ and $Z_1 = \bigcup_{s \in S} \{T_s(Z_0)\}$.

A predecessor of $z' = (s', \zeta') \in Z_1$ is $z \in Z_0$ such that $z' = T_{s'}(z)$; $z_- \subset Z_0$ denotes the set of predecessors of z' . Predecessors may not be unique, since the same (endogenous) state z' can be generated by the equilibrium processes of different initial conditions \bar{z} . Partition the sets Z_1 into (Z_{1a}, Z_{1b}, Z_{1c}) , three disjoint and exhaustive subsets defined as follows:

(a) $Z_{1a} = Z_0 \cap Z_1$ is the set of states that are both initial and successors to the initial conditions; (b) $Z_{1,b} = \{z \in Z_1 \setminus Z_0 : \#z_- = 1\}$ is the set of states with a unique predecessor, but that are not initial conditions; and (c) $Z_{1c} = \{z \in Z_1 \setminus Z_0 : \#z_- > 1\}$ is the set of states with multiple predecessors, but that are not initial conditions.

If (a), then the state \bar{z} is both initial condition—and $\tilde{\xi}_{\bar{z}}$ is the equilibrium associated to it—as well as a successor of an initial condition z , and ξ_z is the competitive equilibrium associated to it. Now, with $s^0 = \bar{s}$ and $s^1 = (s, \bar{s})$, ξ_{z, s^1} and $\xi_{\bar{z}, s^0}$ are Markov equivalent states. Apply $[\cdot_{\tilde{s}^t} \wedge \cdot_{\tilde{s}'^t}]$ as $\tilde{\xi}_z [\cdot_{s^1} \wedge \cdot_{s^0}] \tilde{\xi}_{\bar{z}} = \tilde{\xi}_z^*$. From Claim 9, $\tilde{\xi}_z^*$ is a new competitive equilibrium starting at z and, by construction, $\tilde{\xi}_z^*$ has the same continuation path at s^1 of $\tilde{\xi}_{\bar{z}}$ at $s^0 = \bar{s}$.

- (iii) Thus set $T_s(\bar{z}) = (s, \zeta_{\bar{z}, (\bar{s}, s)}^*)$ and $\xi(z) = \xi_{\bar{z}, (\bar{s}, s)}^*$ for $z \in Z_{1a}$.

If (b), $\tilde{\xi}_{\bar{z}}$ will be the unique competitive equilibrium that has generated $z' \in Z_{1b}$ at $s^1 = (\bar{s}, s')$.

- (iv) Then set $T_s(z') = (s', \zeta_{\bar{z}, (s^1, s)})$ and $\xi(z') = \xi_{\bar{z}, s^1}$ for $\zeta \in Z_{1b}$.

If (c), there may be multiple competitive values for the current endogenous variables at $z = (s, \zeta) \in Z_{1c}$ as well as multiple continuations $(s, \zeta_{z, (s^1, s')})$, $s' \in S$, depending on which competitive process $\tilde{\xi}_{\bar{z}}$, $\bar{z} \in z_-$, we follow. In such a case, we first select arbitrarily one predecessor, $\phi(z) = (s^*, \zeta^*) \in z_-$. Then, for all $z' \in z_-$, we define the new competitive equilibrium $\tilde{\xi}_{z'} [\cdot_{s', s} \wedge \cdot_{s^*, s}] \tilde{\xi}_{\phi(z)}$. Again by Claim 9, this grafting operation uniquely defines the continuation of the Markov state z as $(\hat{s}, \zeta_{\phi(z), (s^{*1}, \hat{s})})$, $\hat{s} \in S$ and $s^{*1} = (s^*, s)$, determined by the competitive equilibrium $\tilde{\xi}_{\phi(z)}$.

- (v) Thus, we set $T_{\hat{s}}(z) = \zeta_{\phi(z), (s^{*1}, \hat{s})}$ and $\xi(z) = \xi_{\phi(z), s^{*1}}$ for $z \in Z_{1c}$.

By construction, the maps ξ and T are well defined functions on $Z_0 \cup Z_1$: each Markov state $z \in Z_0 \cup Z_1$ selects a unique competitive equilibrium and, hence, a unique continuation.

Let $Z_2 = \bigcup_{s \in S} T_s(Z_1)$ be the sets of endogenous Markov states obtained by using operations (iii)–(v). Since each Markov state $z \in Z_0 \cup Z_1$ selects a unique competitive equilibrium and a unique continuation, moving to $t = 3$ we can treat $Z_0 \cup Z_1$ as Z_0 , and Z_2 as Z_1 , and repeat operations (iii)–(v) thereby creating sets Z_3 and extending through the same operations the transition T and the maps ξ to $Z_2 \setminus (Z_0 \cup Z_1)$.

Apply recursively this procedure, thereby creating sets Z_t , $t = 0, 1, \dots$, and defining the maps T and ξ over $\bigcup_{t \geq 0} Z_t$. This describes a Markov equilibrium of $\omega \in \Omega$. Notice that the space of Markov states is $Z = \bigcup_{t \geq 0} Z_t$. Obviously, $Z_0 \subset Z$, and hence $O_s \subset W_{s,\omega}^Z$, $s \in S$, and we have proven Proposition 2(i). Furthermore, by construction, $\bar{z} = (\bar{s}, \bar{w}, \bar{p}, \bar{\lambda})$ is an initial Markov state only if $\bar{z} \in Z_0$ and $\bar{z} \notin Z_t$ for all $t > 1$. Then $\bar{w} \in O_s$ (Proposition 2(ii)) and $(\bar{s}, \bar{w}, \bar{p}, \bar{\lambda}) \in Z_0$ for a unique choice of $(\bar{p}, \bar{\lambda})$, the competitive equilibrium value taken at the initial state \bar{z} , that is, Proposition 2(iii).

PROOF OF CLAIM 9: The argument is simple, so we just sketch it. By construction, $\tilde{\xi}_z^* \equiv \tilde{\xi}_z[\hat{s}^t \wedge \hat{s}''] \tilde{\xi}_{\bar{z}}$ satisfies the market clearing conditions. Thus, it suffices to check that $\tilde{\xi}_z^*$ satisfies the first-order conditions of the individual programming problems. However, since $\tilde{\xi}_z^*$ coincides at s^t with either $\tilde{\xi}_z$ or with $\tilde{\xi}_{\bar{z}}$, it satisfies the optimality conditions $Du^h(x^{ha}) - \lambda^{ha} p = 0$ and the budget constraints for all s^t, h, a . For the same reason, $\tilde{\xi}_z^*$ satisfies as well the no arbitrage equations

$$(11) \quad -\lambda_{s^t}^{ha} q_{s^t} + \delta \sum_s \delta \pi(s|s_t) \lambda_{(s^t, s)}^{h(a+1)} d_s = 0, \quad a < G$$

for $s^t \neq \hat{s}^t$ and $s^t \neq \hat{s}^{t-1}$, the immediate predecessor of \hat{s}^t , since then only variables defined by $\tilde{\xi}_z$ or $\tilde{\xi}_{\bar{z}}$ appear. Hence, it only remains to be checked that (11) is satisfied at \hat{s}^t and \hat{s}^{t-1} . The definition of Markov equivalent realization implies that $(\bar{s}', \lambda_{\bar{z}, \bar{s}'}) = (\hat{s}_t, \lambda_{z, \hat{s}^t})$, while the construction of $\tilde{\xi}_z^*$ implies that $\lambda_{z, \hat{s}^t}^* = \lambda_{\bar{z}, \bar{s}''} = \lambda_{z, \hat{s}^t}$. By direct inspection, the latter implies that both equations are satisfied, thereby concluding the argument. $Q.E.D.$

PROOF OF PROPOSITION 3: For given ω and $\bar{s} \in S$, let

$$\bar{W}_{\bar{s}, \omega}^1 = \left\{ \bar{w}^1 = (\bar{w}^{h1})_{h \in H} : (\bar{w}^1, (\bar{w}^a)_{a>1}) \in \overset{\circ}{W}_{\bar{s}, \omega} \text{ for some } (\bar{w}^{ha})_{h, a>1} \right\}.$$

Given the properties of $\overset{\circ}{W}_{\bar{s}, \omega}$, $\bar{W}_{\bar{s}, \omega}^1$ is an open and bounded subset of \mathbb{R}^H if $G > 1$ and it coincides with $\overset{\circ}{W}_{\bar{s}, \omega}$ otherwise. Let $\bar{s}^1 = (s_{01}, \bar{s})$ (i.e., pick $s_{01} \in S$

and the corresponding finite tree). For $\bar{w}^1 \in \bar{W}_{\bar{s}, \omega}^1$, $f_{(s_{01}, \bar{s}), \bar{w}^1} : P(\omega) \rightarrow \mathbb{R}^H$, where $\dim P(\omega) < H$ by Assumption A1, and there are fewer equations than unknowns. Furthermore, the map $f_{(s_{01}, \bar{s})}$ is independent of \bar{w}^{ha} , $a > 1$. If $G > 1$, the Jacobian matrix of the map $f_{(s_{01}, \bar{s})}(\cdot)$ with respect to $\bar{w}^1 \equiv (\bar{w}^{h1})_{h \in H}$ is equal to the identity matrix. Thus, for given $\omega \in \Omega$, the transversality and preimage theorems imply that for $\bar{w} \in W_{(s_{01}, \bar{s}), \omega}^*$, a full Lebesgue measure subset of $\bar{W}_{\bar{s}, \omega}$, $f_{(s_{01}, \bar{s})}(p, q, \omega, \bar{w}^{h1}) = 0$ does not have a solution in $P(\omega)$. Furthermore, since $P(\omega)$ is compact, the natural projection onto $\overset{\circ}{W}_{\bar{s}, \omega}$ restricted to $P(\omega)$ is proper, and $W_{(s_{01}, \bar{s}), \omega}^*$ is also open.

If $G = 1$, then $\sum_h \bar{w}^{h1} = 0$. Assumption A1 now reads $H > 2[C(1 + S) - 1]$, which implies $H - 1 > C(S + 1) - 1 = \dim P(\omega)$. Thus, drop the function $w_{(s_{01}, \bar{s})}^{h1}(\cdot) - \bar{w}^{h1}$ from the map $f_{(s_{01}, \bar{s})}$ for $h = 1$ and call $f_{(s_{01}, \bar{s})}^\lambda$ the map so obtained. The latter is independent of \bar{w}^{11} , while full-rank perturbations of it can be obtained by perturbing independently \bar{w}^{h1} , $h > 1$. Hence, by the same argument used above, for $\bar{w} \in W_{(s_{01}, \bar{s}), \omega}^*$, an open and full Lebesgue measure subset of $\bar{W}_{\bar{s}, \omega}$, $f_{(s_{01}, \bar{s})}^\lambda(\cdot) = 0$ (and therefore $f_{(s_{01}, \bar{s})}(\cdot) = 0$) does not have a solution in $P(\omega)$.

Now the set $W_{\bar{s}, \omega}^* = \bigcap_{s_{01}} W_{(s_{01}, \bar{s}), \omega}^*$ is an open and full Lebesgue measure subset of $\overset{\circ}{W}_{\bar{s}, \omega}$ that satisfies the required property. $Q.E.D.$

PROOF OF LEMMA 5: We want to show that there exists an open and dense set of parameters Ω' such that for $(e, u, \delta) \in \Omega'$, if $\underline{p}' \in P'(\bar{s})$, then $\underline{p}' \in P'(\bar{s}, \delta)$. Pick $\underline{p}' \in P'(\bar{s})$ and, to avoid trivialities, assume that

$$(12) \quad p_{1+}(a, s) = p_{2+}(a, s) \quad \text{for all } a, s.$$

Then there exists a and ℓ' such that

$$(13) \quad \Pi[S_1^a(\ell')] \neq \Pi[S_2^a(\ell')].$$

Write expand the conditions $\Pi_1(1, \ell) = \Pi_2(1, \ell)$. They are a polynomial in δ , namely,

$$(14) \quad \sum_{a \geq 1} \delta^{a-1} \phi(a, \ell) = 0,$$

where

$$\begin{aligned} \phi(a, \ell) &= \left[\sum_{s^a \in S_{+}(a, \ell)} \pi(s^a | \bar{s}^1) - \sum_{s^a \in S_{2+}(a, \ell)} \pi(s^a | \bar{s}^1) \right] \\ &= \Pi[S_1^a(\ell)] - \Pi[S_2^a(\ell)]. \end{aligned}$$

The coefficients $\phi = (\phi(a, \ell))_{a, \ell}$ are uniquely determined by the Arrow price pairs \underline{p}' through the determination of the sets $S_k^a(\ell)$. Thus, the function and the

notation $\phi_{\underline{p}'}$ are well defined. Condition (13) can be written as $\phi_{\underline{p}'}(a, \ell') \neq 0$. Since the trees are finite, there are only finitely many sets $S_k^a(\ell)$ or, equivalently, there are only finitely many coefficient values ϕ . Let Φ be their set. By definition of $p_{k+}(a, s)$, (12) implies that $p'_{k, \bar{s}^1}/p'_{1, k, \bar{s}^1}$ and $p'_{k, s^2}/p'_{1, k, \bar{s}^1}$ for all $s^2 > \bar{s}^1$ are k -invariant. Thus, $\phi(1, \ell) = \phi(2, \ell) = 0$ for all ℓ and $\phi \in \Phi$. Now, for $\bar{a} \geq 3$, let

$$\begin{aligned}\Phi(\bar{a}) = \{\phi \in \Phi : \phi(a, \ell) = 0, \text{ for all } \ell \text{ and} \\ a < \bar{a}, \phi(\bar{a}, \ell') \neq 0, \text{ for some } \ell'\}.\end{aligned}$$

If $\underline{p}' \in P'(\bar{s})$ and (12) holds, then $\phi_{\underline{p}'} \in \Phi(\bar{a})$, for some $\bar{a} \geq 3$. The equation $\underline{\Pi}_1(1, \ell') = \underline{\Pi}_2(1, \ell')$ now reads

$$\sum_{a \geq \bar{a}}^G \delta^{a-\bar{a}} \phi(a, \ell') = 0.$$

Since the coefficient for the degree-zero term is $\phi(\bar{a}, \ell') \neq 0$, the latter is a nonzero polynomial and by Theorem 14, in Zariski and Samuel (1960, Chap. I, p. 38) the set of zeros of the polynomial is closed and has measure zero in $(0, 1]$. Let Ω_ϕ be the complement of this set and let $\Omega(\bar{a}) = \bigcap_{\phi \in \Phi(\bar{a})} \Omega_\phi$. By construction, Ω_ϕ and, therefore, $\Omega(\bar{a})$ are open and dense. Finally set $\Omega' = \bigcap_{\bar{a} \geq 3} \Omega(\bar{a})$ as an open and dense subset set of Ω . Most importantly, if $\underline{p}' \in P'(\bar{s})$, then $\phi_{\underline{p}'} \in \Phi(\bar{a})$ for some \bar{a} ; then if $\omega = (e, u, \delta) \in \Omega'$, $\underline{\Pi}_{1+}(\ell') \neq \underline{\Pi}_{2+}(\ell')$ for some ℓ' or $\underline{p}' \in P'(\bar{s}, \delta)$. $Q.E.D.$

PROOF OF LEMMA 6: Each $\underline{p}' \in P'(\bar{s})$ uniquely determines the set of indices \mathbb{P}^1 , the sets of histories of length a , $S_k^a(\ell)$ for all $k, \ell \in \mathbb{P}^1$ and $a \geq 1$, and the values $p_{k+}(a, s)$ for all a, s . First, the possible configurations of the index set \mathbb{P}^1 and of histories $S_k^a(\ell)$ are finite. Second, if two pairs of Arrow prices determine the same $\mathbb{P}^1, S_k^a(\ell)$ for all $k, \ell \in \mathbb{P}^1$, and $a \geq 1$, they deliver identical values for $\underline{\Pi}[S_k^a(\ell)]$. Then partition $P'(\bar{s})$ into a collection of J disjoint and exhaustive subsets $P'_j(\bar{s})$, where J is the cardinality of the sets $(\mathbb{P}^1, (S_k^a(\ell))_{k, \ell \in \mathbb{P}^1, a \geq 1})$ generated by the elements of $P'(\bar{s})$ and j is their indices. Two distinct price pairs in $P'_j(\bar{s})$ determine the same sets of price indices \mathbb{P}^1 and of states $S_k^a(\ell)$, but distinct values $p(\ell)$. Therefore, without ambiguity denote with \mathbb{P}_j^1 the set of price indices generated by $\underline{p}' \in P'_j(\bar{s})$. Define $P_j^m(\bar{s})$ as the subset of $P'_j(\bar{s})$ that satisfies

$$\|p(\ell) - p(\ell')\| \geq \frac{1}{n} \quad \text{for all } \ell \text{ and } \ell' \in \mathbb{P}_j^1, j = 1, \dots, J.$$

Then define $P_0^m(\bar{s})$ as

$$\|(p_{1+}(a, s) - p_{2+}(a, s))_{a, s}\| \geq \frac{1}{n}.$$

Finally, let

$$P'^n(\bar{s}) = \bigcup_{j=0}^J P_j'^n(\bar{s}).$$

The sets $P_j'^n(\bar{s})$ are compact for all j and, therefore, so is $P'^n(\bar{s})$. Lemma 5 guarantees that for $\omega \in \Omega'$ and $\underline{p}' \in P'^n(\bar{s}) \subset P'(\bar{s})$, inequalities (10) are satisfied. Obviously $P'^n(\bar{s}) \subset P'^{n+1}(\bar{s})$ and $\bigcup_n P'^n(\bar{s}) = \text{cl}(P'(\bar{s}))$ by construction. *Q.E.D.*

Utility Perturbations: For the proofs that follow, we use utilities to perturb the equations $f_j^h = 0$. To do so, we use a locally finite, linear parametrization of the utility functions. Pick $N > 1$ distinct consumption bundles $x_j \in \mathbb{R}_{++}^C$, $j = 1, \dots, N$. We perturb the gradient of $u^h(\cdot)$ around the bundles x_j , $j = 1, \dots, N$. For each j , pick a pair of open balls $B_{\varepsilon_1}(x_j)$, $i = 1, 2$, centered around x_j and such that (i) $\varepsilon_2 > \varepsilon_1$ and (ii) $\bigcap_j \text{cl}(B_{\varepsilon_2}(x_j)) = \emptyset$. Then pick smooth “bump” functions Φ_j such that $\Phi_j(x; x_j) = 1$ for $x \in B_{\varepsilon_1}(x_j)$, and $\Phi_j(x; x_j) = 0$ for $x \notin \text{cl}(B_{\varepsilon_2}(x_j))$. For given arbitrary vectors $\Delta u = (\Delta u_j)_{j=1}^N \in \mathbb{R}^{CN}$ and scalar η , define the utility function

$$u_\eta^h(x, \Delta u) = u^h(x) + \eta \sum_{j=1}^N \Phi_j(x; x_j) \sum_c \Delta u_{j,c} x_c.$$

For any given Δu , we can pick η so close to zero that $u_\eta^h(\cdot)$ is arbitrarily close to $u^h(\cdot)$ in the C^2 -uniform convergence topology and it satisfies, therefore, all the maintained assumptions. We identify a utility perturbation with the (Gateaux) derivative of D (the derivative as a linear map of functions) at u^h in the direction $\sum_j \Phi_j(x; x_j) \Delta u_j$, that is, with a vector Δu . Which finite set of points x_j is used depends on \underline{p}' , so the derivative of u^h , hence of f , is not finitely parameterized.

PROOF OF LEMMA 7: As a preliminary step, to apply the infinite-dimensional version of transversality, we need the space of utilities to be Banach. This is done as follows. Let $\bar{X} \subset \mathbb{R}_{++}^C$ be the compact set defined in Section 4.4. Recall that, thereafter, we have identified Ω with $B_\varepsilon(\omega)$ for a given ω . By construction, when $(\psi, \omega') \in P \times B_\varepsilon(\omega)$, optimal consumptions are contained in \bar{X} . This is going to stay true when we move to P' and, in particular, to $P'(\bar{s})$. Since density is a local property, in what follows we can identify \mathcal{U} with the utilities u^h restricted to the compact domain \bar{X} , a complete subspace of $C^2(\bar{X}, \mathbb{R})$ when endowed with the uniform topology, and a Banach space. If a set is dense in Ω in this topology, it is also dense when Ω is endowed with the original, coarser topology. For simplicity, we keep the notation \mathcal{U} , Ω , Ω'' , and f_j^h unchanged.

Let $\mathcal{G} = \{g : H \rightarrow \{1, \dots, 2 + \sum_a S^a\}\}$ be the set of maps that assign for each h a map $f_{g(h)}^h \in \mathfrak{F}$. Then, for all $g \in \mathcal{G}$, $f_g = (f_{g(h)}^h)_{h \in H} : P'(\bar{s}) \times \Omega'' \rightarrow \mathbb{R}^H$ are continuously differentiable and E_g are (relatively) open sets of $P'(\bar{s}) \times \Omega''$. Let $\phi_g : E_g \rightarrow \mathbb{R}^H$ be the restriction of f_g to E_g and let $\Omega(g) = \text{proj}_{\Omega}(E_g)$. By construction, either $\phi_g \neq 0$ or $D\phi_g$ is onto \mathbb{R}^H , and since $\ker D\phi_g$ is finite codimensional in $\mathbb{R}^{\dim P'(\bar{s})} \times \Omega(g)$, ϕ_g is transversal to zero. Then, by the parametric transversality theorem (see Abraham, Marsden, and Ratiu (1988, Theorem 3.6.22)), $\phi_{g,\omega}$ is also transversal to zero for all $\omega \in \Omega^*(g)$, a dense subset of $\Omega(g)$. However, since by Assumption A1, $\dim P < H$, by the preimage theorem, $\phi_{g,\omega}^{-1}(0) = \emptyset$ for $\omega \in \Omega^*(g)$. Then $\Omega_g = \Omega^*(g) \cup \Omega'' \setminus \Omega(g)$ is a dense subset of Ω'' and so is $\overline{\Omega} = \bigcap_g \Omega_g$. Bear in mind that by construction $E_g \cap P'(\bar{s}) \times \overline{\Omega} = N_g$ for all g . We now show that for all $(\underline{p}', \omega) \in P'(\bar{s}) \times \overline{\Omega}$, there exists h such that $f_1^h(\underline{p}', \omega) \neq 0$. Pick any such (\underline{p}', ω) . By Condition Universal, $E_1^h \cup E_J^h = P'(\bar{s}) \times \Omega''$ for all h , and hence $(\underline{p}', \omega) \in E_g$, and, therefore, $(\underline{p}', \omega) \in N_g$, for some $g \in \mathcal{G}$ with $g(h) = 1$ or J for all h . Let $H_g = \{h : (\underline{p}', \omega) \in N_{g(h)}^h\}$. By construction of N_g and E_g , H_g is nonempty. If $g(h) = 1$ for some $h \in H_1$, then $f_1^h(\underline{p}', \omega) \neq 0$, a contradiction that concludes the argument. Otherwise, $g(h) = J$ for all $h \in H_g$. Since $(\underline{p}', \omega) \in N_J^h$, by Condition Nesting, $(\underline{p}', \omega) \in E_{j(h)}^h$ with $j(h) < J$ for all $h \in H_1$. Let $g^1 \in \mathcal{G}$ be defined as $g^1(h) = g(h)$ for $h \in H \setminus H_g$, while $g^1(h) = j(h)$ for $h \in H_g$. Since $\omega \in \overline{\Omega}$, $(\underline{p}', \omega) \in N_{g^1}$. By construction of H_g and g^1 , $f_{g^1(h)}^h(\underline{p}', \omega) = 0$ for all $h \in H \setminus H_g$. Since $(\underline{p}', \omega) \in N_{g^1}$, it must be that $H_{g^1} = \{h : f_{g^1(h)}^h(\underline{p}', \omega) \neq 0\} \subset H_g$ is nonempty. For $h \in H_{g^1}$, it is $(\underline{p}', \omega) \in N_{g^1(h)}^h \cap N_J^h$ and, therefore, Condition Nesting implies that $(\underline{p}', \omega) \in E_{j(h)}^h$ for some $j(h) < g^1(h)$. Iterating finitely many times, the argument concludes that $(\underline{p}', \omega) \in N_{g^*}$ for g^* with $g^*(h) = 1$ for some $h \in H_{g^*}$. $Q.E.D.$

PROOF OF LEMMA 8:

Preliminary Computations. In analogy with the notation in the text, for $\underline{p}' \in P'$, let

$$\mathbb{P} = \{\hat{p} \in \mathbb{R}_{++}^C : p'_{k,s^a} = \hat{p}, \text{ for some } k, s^a\}$$

with cardinality $\mathbb{P} \leq 2^{\sum_{a=0}^{G-1} S^a}$, generic element $\hat{p}(\ell)$, and \mathbb{P} denoting also the set of price indices ℓ . For $\ell \in \mathbb{P}$, we define sets of histories and probability weights

$$S_k(\ell) = \{s^a : p'_{k,s^a} = \hat{p}(\ell)\}, \quad \Pi_k(\ell) = \sum_{s^a \in S_k(\ell)} \delta^a \pi(s^a | s_{0k}),$$

where $\Pi_k(\ell) = 0$ if $S_k(\ell) = \emptyset$ and

$$p_k(a, s) = \delta^{a-1} \sum_{s^{a-1}: (s^{a-1}, s)} \pi(s^{a-1}, s | \bar{s}^1) p'_{k, (s^{a-1}, s)}.$$

Given when $p'_{1,1,\bar{s}^1} \neq p'_{1,1,\bar{s}^1}$, $\underline{p}' \in P'(\bar{s})$, $\hat{p}(\ell) \in \mathbb{P}$ generates two distinct values in the set \mathbb{P}_1 : $p(\ell^1) = (\hat{p}(\ell)) / p'_{1,1,\bar{s}^1}$ and $p(\ell^2) = (\hat{p}(\ell)) / p'_{1,2,\bar{s}^1}$. To keep the two index sets consistent, use the convention $\ell \in \mathbb{P} \cap \mathbb{P}^1$ if and only if $p(\ell) = (\hat{p}(\ell)) / p'_{1,k,\bar{s}^1}$ for some $k = 1, 2$.

Pick $\underline{p}' \in P'(\bar{s})$ and, with this notation at hand, rewrite the individual programming problem (5) as

$$(15) \quad \begin{aligned} \max_{\ell \in \mathbb{P}} & \sum \Pi_k(\ell) u^h(x_k(\ell)) \quad \text{s.t.} \\ & \sum_{\ell \in \mathbb{P}} \Pi_k(\ell) \hat{p}(\ell) x_k(\ell) = \sum_{a,s} p_k(a, s) e_s^{ha}. \end{aligned}$$

The first-order conditions associated to problem (15) are, for $\ell \in \mathbb{P}$,

$$\begin{aligned} Du^h(x_k(\ell)) - \lambda_k^h \hat{p}(\ell) &= 0, \\ \sum_{\ell \in \mathbb{P}} \Pi_k(\ell) \hat{p}(\ell) x_k(\ell) - \sum_{a,s} p_k(a, s) e_s^{ha} &= 0. \end{aligned}$$

Drop h . We perturb the utility function around $x_k(\ell)$ without disturbing it around $x_k(\ell')$, $\ell' \neq \ell$. We denote such a perturbation by $\Delta u(\ell)$ and denote the endowment perturbation by Δe_s^a , while $(\Delta x, \Delta \lambda, \Delta w)$ is their effect on the variables (x, λ, w) .

Differentiating the first-order conditions, we get

$$\begin{aligned} (l) \quad & H_k(\ell) \Delta x_k(\ell) - \hat{p}^T(\ell) \Delta \lambda_k - \Delta u(\ell) = 0, \\ (\text{bc}) \quad & \sum_{\ell \in \mathbb{P}} \Pi_k(\ell) \hat{p}(\ell) \Delta x_k(\ell) - \sum_{a,s} p_k(a, s) \Delta e_s^a = 0, \end{aligned}$$

where $H_k(\ell)$ is the invertible Hessian at $x_k(\ell)$ and the superscript T stands for transpose. While differentiating the map w_{k,\bar{s}^1} , recalling that $p(\ell) = (\hat{p}(\ell)) / p'_{1,k,\bar{s}^1}$, we get

$$(\text{W}) \quad \Delta w_{k,\bar{s}^1} = \sum_{\ell \in \mathbb{P}} \Pi_k(1, \ell) p(\ell) \Delta x_k(\ell) - \sum_{a>0, s} p_{k+}(a, s) \Delta e_s^a.$$

Define

$$Q_k(\ell) \equiv \hat{p}(\ell) H_k^{-1}(\ell) \hat{p}^T(\ell).$$

Since $H_k^{-1}(\ell)$ is a negative definite matrix, the terms $Q_k(\cdot)$ are negative, and so is $Q_k = \sum_{\ell \in \mathbb{P}} \Pi_k(\ell) Q_k(\ell)$. From equation (1), by performing elementary computations, we get

$$\Delta \lambda_k = \left\{ \sum_{a,s} \frac{p_k(a, s) \Delta e_s^a}{Q_k} - \sum_{\ell \in \mathbb{P}} \frac{\Pi_k(\ell) \hat{p}(\ell) H_k^{-1} \Delta u(\ell)}{Q_k} \right\},$$

and taking into account that $p(\ell) = (\hat{p}(\ell)) / p_{1,k,\bar{s}^1}$, we get

$$\begin{aligned} \Delta w_{k,\bar{s}^1} &= Q_{k+} \Delta \lambda_k + \sum_{\ell \in \mathbb{P}^1} \Pi_k(1, \ell) p(\ell) H_k^{-1}(\ell) \Delta u(\ell) \\ &\quad - \sum_{a>0, s} p_{k+}(a, s) \Delta e_s^a, \end{aligned}$$

where $Q_{k+} = \sum_{\ell \in \mathbb{P}^1} \Pi_{k+}(\ell) (Q_k(\ell)) / p'_{1,k,\bar{s}^1} < 0$.

Denote by λ_k the value $\lambda_k(\underline{p}_k', \omega)$. By the first-order conditions of the individual problems, if $\Pi_k(\ell) > 0$ for all k and if $\lambda_1 = \lambda_2$, then $x_1(\ell) = x_2(\ell)$ at the optimal solutions of the two programming problems. However, if $\lambda_1 \neq \lambda_2$, for any pair ℓ and ℓ_ℓ such that $\Pi_1(\ell) > 0$ and $\Pi_2(\ell_\ell) > 0$, $x_1(\ell) = x_2(\ell_\ell)$ if and only if $\hat{p}(\ell) = (\lambda_1/\lambda_2) \hat{p}(\ell_\ell)$. Hereafter, for each ℓ (with $\Pi_1(\ell) > 0$), we denote with ℓ_ℓ the index associated to $\hat{p}(\ell) = (\lambda_1/\lambda_2) \hat{p}(\ell_\ell)$ —and if $\lambda_1 = \lambda_2$, $\ell_\ell = \ell$. Also bear in mind that $H_1^{-1}(\ell) = H_2^{-1}(\ell_\ell)$. We denote by $\ell = 1$ the price equivalence class identified by p'_{1,s_01} and denote by ℓ_1 the price equivalence class associated to $p'_{2,s^a} = (\lambda_1/\lambda_2) p'_{1,s_01}$. Define $f_{\ell_1} = x_1(1) - x_2(\ell_1)$ and observe that f_{ℓ_1} is smooth in ω if $\Pi_2(\ell_1) > 0$ or, equivalently, if $\lambda_1(\underline{p}_1', \omega) \neq \lambda_2(\underline{p}_2', \omega)$, while $f_{\ell_1}(\underline{p}', \omega) \neq 0$ otherwise. Notice that $f_{\ell_1} \neq 0$ if and only if $f_j \neq 0$, $j = 2, \dots, \sum_{a=0}^G S^a$, and that $D_\omega f_j = D_\omega f_{\ell_1}$ for all $j = 2, \dots, 1 + \sum_{a=0}^G S^a$ such that $f_j = 0$. Therefore, the set E_2^h is just the set

$$E_2^h = \{(\underline{p}', \omega) : [f_{\ell_1}^h = 0 \text{ and } D_\omega f_{\ell_1} \neq 0] \text{ or } f_{\ell_1} \neq 0\}.$$

By taking into account that $\hat{p}(\ell) = (\lambda_1/\lambda_2) \hat{p}(\ell_\ell)$, the derivatives of the map f_{ℓ_1} are

$$\begin{aligned} (16) \quad D_{u(\ell)} f_{\ell_1} &= H_1^{-1}(1) \hat{p}^T(1) \hat{p}(\ell) H_1^{-1}(\ell) \\ &\quad \times \left[\left(\frac{\lambda_1}{\lambda_2} \right)^2 \frac{\Pi_2(\ell_\ell) \Delta u(\ell)}{Q_2} - \frac{\Pi_1(\ell) \Delta u(\ell)}{Q_1} \right] \end{aligned}$$

and

$$(17) \quad D_{e_s^a} f_{\ell_1} = H_1^{-1}(1) \hat{p}^T(1) \left[\left(\frac{\lambda_1}{\lambda_2} \right) \frac{p_2(a, s)}{Q_2} - \frac{p_1(a, s)}{Q_1} \right].$$

Hereafter, to simplify notation, we set $J = 2 + N_0$. We are now ready to establish Condition **Universal**. If, for $(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega'$, either $f_J^h(\underline{p}', \omega) \neq 0$ or $f_1^h(\underline{p}', \omega) \neq 0$, Condition **Universal** holds true. Hence, in the next claim we are concerned only with the complementary case.

CLAIM 10: Suppose that $(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega'$ and $f_1^h(\underline{p}', \omega) = 0$. If $(s_{0k}, p'_{k,s_{0k}})$ is one-to-one in k , then $D_{\omega^h} f_1^h(\underline{p}', \omega) \neq 0$; otherwise, and if $f_J^h(\underline{p}', \omega) = 0$, then either $D_{\omega^h} f_1^h(\underline{p}', \omega) \neq 0$ or $D_{\omega^h} f_J^h(\underline{p}', \omega) \neq 0$.

PROOF: Drop h . We argue by contradiction, assuming that $D_\omega f_1 = 0$. If $D_{\Delta e_s^a} f_1 = 0$ for all (a, s) , then

$$(18) \quad \frac{Q_{1+}}{Q_1} p_1(a, s) - p_{1+}(a, s) = \frac{Q_{2+}}{Q_2} p_2(a, s) - p_{2+}(a, s) \quad \text{for all } (a, s).$$

Equation (18) computed at $(0, s_{01})$ implies that $s_{01} = s_{02}$. Moreover, since by the adopted normalization $p_{1,k,s_0} = 1$, it also implies that p'_{k,s_0} and Q_{k+}/Q_k are k -invariant, a contradiction. Assume, therefore, that $(s_{0k}, p'_{k,s_{0k}})$ is k -invariant and that also, by contradiction, $D_{\omega^h} f_1 = D_{\omega^h} f_J = 0$. If $D_{\Delta e_s^a} f_J = 0$ for all (a, s) , then $(p_1(a, s))/Q_1 = (p_2(a, s))/Q_2$ for all (a, s) . Computed at $(0, s_0)$, the latter implies $Q_k = Q$ for $k = 1, 2$; computed at any other (a, s) implies that $p_k(a, s)$ is k -invariant and we set $p_k(a, s) = p(a, s)$. Thus, since $p_k(1, \bar{s}) = \delta\pi(\bar{s}^1|s_0)p'_{k,\bar{s}^1}$, it is $p'_{1,1,\bar{s}^1} = p'_{1,2,\bar{s}^1} \equiv p'_{1,\bar{s}^1}$. Then $D_{e_s^a} f_1 = 0$ or, equivalently, equation (18) reads

$$\frac{Q_{1+}}{Q} p(a, s) - p_{1+}(a, s) = \frac{Q_{2+}}{Q} p(a, s) - p_{2+}(a, s).$$

By taking into account that the first entry of $p_{k+}(1, \bar{s})$ is equal to 1, the last equation computed at $(1, \bar{s})$ implies that $Q_{k+} = Q_+$ and then that

$$(19) \quad p_{1+}(a, s) = p_{2+}(a, s) \quad \text{for all } a, s.$$

We turn next to utility perturbations. For each ℓ , pick the perturbation $\Delta u(\ell)$ and bear in mind that since λ_k is k -invariant, so is $x_k(\ell)$ k -invariant. By direct computation, if $D_{\Delta u(\ell)} f_1 = 0$ (and recalling that $p(\ell)H_k^{-1}(\ell)\Delta u(\ell) = (c_k(\ell))/p_{1,\bar{s}^1} = 1/p'_{1,\bar{s}^1}$), then

$$(20) \quad \frac{\Pi_1(1, \ell)}{p'_{1,\bar{s}^1}} - \frac{\Pi_1(\ell)Q_+}{Q} = \frac{\Pi_2(1, \ell)}{p'_{1,\bar{s}^1}} - \frac{\Pi_2(\ell)Q_+}{Q} \quad \text{for all } \ell.$$

Similarly, if $D_{\Delta u(\ell)} f_J = 0$ for all ℓ , then $(\Pi_1(\ell))/Q = (\Pi_2(\ell))/Q$ for all ℓ . These equations and equation (20) immediately imply that

$$\Pi_1(\ell) = \Pi_2(\ell) \quad \text{and} \quad \Pi_{1+}(\ell) = \Pi_{2+}(\ell) \quad \text{for all } \ell.$$

The latter together with (19) contradict the assumption $\underline{p}' \in P'(\bar{s})$. $\quad Q.E.D.$

The next lemma shows that if $\lambda_1^h \neq \lambda_2^h$, then either $f_j^h \neq 0$ or $D_{\omega^h} f_j^h \neq 0$ for all $j = 2, \dots, J-1$. In other words, it shows that $N_J^h \subset \bigcap_{2 \leq j \leq J-1} E_j^h$. Since, both $N_J \cap N_j \subset N_J$ and $\bigcap_{2 \leq j \leq J-1} E_j^h \subset \bigcup_{j' < j} E_{j'}^h$ for all $j > 1$, the claim shows that Condition Nesting holds true for all $j > 1$.

CLAIM 11: Suppose that $(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega'$ is such that (i) $f_1^h(\underline{p}', \omega) = 0$ and (ii) $f_J^h(\underline{p}', \omega) \neq 0$. If $f_j^h(\underline{p}', \omega) = 0$, then $D_{\omega^h} f_j^h(\underline{p}', \omega) \neq 0$ for all $1 < j < J$.

PROOF: Drop h . Arguing by contradiction, assume that both $D_{\omega} f_j = 0$ and $f_j(\underline{p}', \omega) = 0$ for some $1 < j < J$. Then, by equation (17), $D_e f_{\ell_1} = 0$ reads

$$\left(\frac{\lambda_1}{\lambda_2} \right) \frac{p_2(a, s)}{Q_2} = \frac{p_1(a, s)}{Q_1} \quad \text{for all } (a, s).$$

Computing the latter at $(0, s_0)$ (and since $p_{1,k,s_0} = 1$), we get that

$$\left(\frac{\lambda_1}{\lambda_2} \right) \frac{1}{Q_2} = \frac{1}{Q_1}.$$

Next we move to utility perturbations. Then by equation (16), $D_{\Delta u(\ell')} f_{\ell_1} = 0$ reads

$$\left(\frac{\lambda_1}{\lambda_2} \right) \Pi_2(\ell_\ell) = \Pi_1(\ell) \quad \text{for all } \ell.$$

Summing across ℓ and noticing the k -invariance of $\sum_\ell \Pi_k(\ell)$, we get $\lambda_1 = \lambda_2$, an immediate contradiction with Claim 11(ii). $\quad Q.E.D.$

The final claim establishes Condition Nesting for $j = 1$, thereby concluding the argument.

CLAIM 12: There exists an open and dense subset Ω'' of Ω' such that if $(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega''$ and $f_j^h(\underline{p}', \omega) \neq 0$ for all $j \geq 2$, then $D_{\omega^h} f_1^h(\underline{p}', \omega) \neq 0$.

PROOF: Drop h and, assume that $(s_{0k}, p'_{k,s_{0k}}) \equiv (s_0, p_{s_0})$ is k -invariant; otherwise Claim 10 implies the thesis. Since $f_j^h(\underline{p}', \omega) \neq 0$ for all $j \geq 2$, it is $\Pi_2(\ell_1) = 0$. Furthermore, since $p'_{k,s_{0k}} = p_{s_0}$ for all k , if $\Pi_1(\ell) = 0$ for all $\ell \neq 1$, there must be at least two distinct indexes $\ell \neq \ell_1$ and $\ell' \neq \ell_1$ with both $\Pi_2(\ell') > 0$ and $\Pi_2(\ell) > 0$; otherwise, $p'_{k,s^a} = p'_{1,s_0}$ for all k and s^a , contradicting $\underline{p}' \in P'(\bar{s})$. Second, given s_{0k} is k -invariant and given $\underline{p}' = (\underline{p}_1', \underline{p}_2')$, then

$D_\omega f_1((\underline{p}'_1, \underline{p}'_2), \omega) = -D_\omega f_1((\underline{p}'_2, \underline{p}'_1), \omega)$. Therefore, up to a relabeling of the two trees, there is no loss of generality in assuming that $p'_{1,s^a} \neq p'_{1,s_0}$ for some s^a .

We now argue by contradiction, that is, we assume that $D_{\Delta s^a} f_1 = 0$. If $D_{\Delta s^a} f_1 = 0$ for all a, s , then equation (18) holds true and p'_{1,k,\bar{s}^1} is k -invariant, thereby implying as already argued the k -invariance of Q_{k+}/Q_k .

Thus, if $D_{\Delta u(\ell)} f_1 = 0$, then

$$\left[\frac{\Pi_1(1, \ell)}{p'_{1,\bar{s}^1}} - \frac{Q_{1+}}{Q_1} \Pi_1(\ell) \right] = \left(\frac{\lambda_1}{\lambda_2} \right) \left[\frac{\Pi_2(1, \ell_\ell)}{p'_{1,\bar{s}^1}} - \frac{Q_{1+}}{Q_1} \Pi_2(\ell_\ell) \right],$$

and summing across ℓ and observing that $\sum \Pi_k(\ell) \equiv \Pi$ and $\sum \Pi_k(1, \ell) \equiv \Pi_+$ are k -invariant, it must be that $\Pi_+/\Pi = p'_{1,\bar{s}^1} Q_{k+}/Q_k$ since $\lambda_1 \neq \lambda_2$. Therefore,

$$\left[\frac{\Pi_1(1, \ell)}{\Pi_+} - \frac{\Pi_1(\ell)}{\Pi} \right] = \left(\frac{\lambda_1}{\lambda_2} \right) \left[\frac{\Pi_2(1, \ell_\ell)}{\Pi_+} - \frac{\Pi_2(\ell_\ell)}{\Pi} \right] \quad \text{for all } \ell.$$

However, since $\Pi_2(\ell_1) = \Pi_2(1, \ell_1) = 0$, it is

$$(21) \quad \Pi \Pi_1(1, 1) - \Pi_1(1) \Pi_+ = 0.$$

Since

$$\Pi = \sum_{j=0}^G \delta^j = \frac{1 - \delta^{G+1}}{1 - \delta},$$

while

$$\Pi_+ = \sum_{j=0}^{G-1} \delta^j = \frac{1 - \delta^G}{1 - \delta}$$

(21) is a polynomial equation of degree $2G$ in δ of the form

$$(1 - \delta^{G+1})(a_0 + a_1 \delta + \cdots + a_G \delta^{G-1}) - (1 - \delta^G)(b_0 + b_1 \delta + \cdots + b_G \delta^G) \equiv \sum_{j=0}^{2G} t_j \delta^j = 0,$$

where

$$a_n = \sum_{s^{n+1} \in S_1(n+1, 1)} \pi(s^{n+1} | \bar{s}^1)$$

and

$$b_n = \sum_{s^n \in \bar{S}_1(n, 1)} \pi(s^n | s_0)$$

for

$$\bar{S}_1(n, 1) = \{s^a : p_{1,s^a} = p_{1,s_0}, \text{ for } a = n\}.$$

We want to show that this polynomial is nonzero, that is, that $t_j \neq 0$ for some j . By trivial computations,

$$t_j = a_j - b_j \quad \text{for } j = 0, \dots, G-1, t_G = b_0 - b_G,$$

while

$$t_{j+(G+1)} = b_{j+1} - a_j \quad \text{for } j = 0, \dots, G-1.$$

Thus, if $t_j = 0$, for all j , then $b_j = b_{j+1}$, $j \geq 1$. Since, by construction, $b_0 = 1$ and $t_G = b_0 - b_G$, then $t_j = 0$ for all j if and only if $b_j = a_j = 1$ for all j . Obviously, the latter contradicts the assumption $p'_{1,s^a} \neq p'_{1,s_0}$ for some s^a . Since the polynomial is nonzero, by Theorem 14 in Zariski and Samuel (1960, Chap. I, p. 38), the set of zeros of the polynomial is closed and has measure zero in $[0, 1]$. Most importantly, there are only finitely many pairs of nonempty and exhaustive subsets of the tree. Each of these pairs uniquely determines the vectors of coefficients t of the polynomial above and we only consider pairs that deliver a vector t with $t_j \neq 0$ for some j . The union of the zeros of such polynomials intersected with $(0, 1]$ is the finite union of closed and measure zero sets. Thus, its complement Δ is open and of full measure in $(0, 1]$. Let $\Omega'' = \Omega' \cap [E \times \mathcal{U} \times \Delta]$. By construction, for all $(\underline{p}', \omega) \in P'(\bar{s}) \times \Omega''$, no equation (21) is satisfied or, equivalently, $D_\omega f_1 \neq 0$, concluding the proof. *Q.E.D.*

REFERENCES

- ABRAHAM, R., J. MARSDEN, AND T. RATIU (1988): *Manifolds, Tensor Analysis, and Applications*. New York: Springer-Verlag. [340]
- ALLAIS, M. (1947): *Economie et Intérêt*. Paris: Imprimerie Nationale. [309]
- BALASKO, Y., AND K. SHELL (1980): "The Overlapping-Generations Model: I. The Case of Pure Exchange Without Money," *Journal of Economic Theory*, 23, 281–306. [315]
- CASS, D., R. GREEN, AND S. SPEAR (1992): "Stationary Equilibria With Incomplete Markets and Overlapping Generations," *International Economic Review*, 33, 495–512. [309]
- CITANNA, A., AND P. SICONOLFI (2007): "Recursive Equilibrium in Stochastic OLG Economies: Incomplete Markets," Mimeo, Columbia University. [311]
- (2008): "On the Nonexistence of Recursive Equilibrium in Stochastic OLG Economies," *Economic Theory*, 37, 417–437. [312]
- CONSTANTINIDES, G., J. DONALDSON, AND R. MEHRA (2002): "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle," *Quarterly Journal of Economics*, 117, 269–276. [310]
- DUFFIE, D., J. GEANAKOPLOS, A. MAS-COLELL, AND A. MCLENNAN (1994): "Stationary Markov Equilibria," *Econometrica*, 62, 745–781. [309, 310, 318]
- GEANAKOPLOS, J., M. MAGILL, AND M. QUINZII (2004): "Demography and the Long-Run Predictability of the Stock Market," *Brookings Papers on Economic Activity*, 1, 241–307. [310]

- GOTTARDI, P. (1996): "Stationary Monetary Equilibria in OLG Models With Incomplete Markets," *Journal of Economic Theory*, 71, 75–89. [309]
- KUBLER, F., AND H. POLEMARCHAKIS (2004): "Stationary Markov Equilibria for Overlapping Generations," *Economic Theory*, 24, 623–643. [310,312,316,333]
- KUBLER, F., AND K. SCHMEDDERS (2005): "Approximate versus Exact Equilibria in Dynamic Economies," *Econometrica*, 73, 1205–1235. [312]
- LJUNGQVIST, L., AND T. SARGENT (2000): *Recursive Macroeconomic Theory*. Cambridge, MA: MIT Press. [310]
- MAS-COLELL, A., AND J. NACHBAR (1991): "On the Finiteness of the Number of Critical Equilibria, With an Application to Random Selections," *Journal of Mathematical Economics*, 20, 397–409. [311]
- RIOS RULL, J. V. (1996): "Life-Cycle Economies and Aggregate Fluctuations," *Review of Economic Studies*, 63, 465–489. [310,311,316,333]
- SAMUELSON, P. (1958): "An Exact Consumption-Loan Model of Interest Without the Social Contrivance of Money," *Journal of Political Economy*, 66, 467–482. [309]
- SPEAR, S. (1985): "Rational Expectations Equilibria in the OLG Model," *Journal of Economic Theory*, 35, 251–275. [309]
- STOKEY, N., AND R. LUCAS (1989): *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press. [310]
- STORESLETTEN, S., C. TELMER, AND A. YARON (2004): "Cyclical Dynamics in Idiosyncratic Labor-Market Risk," *Journal of Political Economy*, 112, 695–717. [310]
- ZARISKI, O., AND P. SAMUEL (1960): *Commutative Algebra*, 1. Princeton, NJ: Van Nostrand. [338, 346]

*Dept. of Economics and Finance, HEC-Paris, 1 Rue de la Libération, Jouy-en-Josas, 78351 Paris, France; citanna@hec.fr
and*

*Columbia University, Graduate School of Business, 3022 Broadway, New York,
NY 10027-6902, U.S.A.; ps17@columbia.edu.*

Manuscript received June, 2007; final revision received June, 2009.

NOTES AND COMMENTS MULTIPLE TEMPTATIONS

BY JOHN E. STOVALL¹

We use a preference-over-menus framework to model a decision maker who is affected by multiple temptations. Our two main axioms on preference—exclusion and inclusion—identify when the agent would want to restrict his choice set and when he would want to expand his choice set. An agent who is tempted would want to restrict his choice set by *excluding* the normatively worst alternative of that choice set. Simultaneously, he would want to expand his choice set by *including* a normatively superior alternative. Our representation identifies the agent's normative preference and temptations, and suggests the agent is uncertain which of these temptations will affect him. We provide examples to illustrate how our model improves on those of Gul and Pesendorfer (2001) and Dekel, Lipman, and Rustichini (2009).

KEYWORDS: Temptation, self-control, exclusion, inclusion.

1. INTRODUCTION

WE USE A PREFERENCE-OVER-MENUS FRAMEWORK to model a decision maker who is affected by multiple temptations. We regard temptation as a craving or desire that is different from the agent's normative preference (i.e., his view of how he *should* choose between alternatives, absent temptation). When the agent cannot simultaneously satisfy a temptation and his normative desire, he is conflicted. He wants to avoid such conflict *and* make normatively good choices.

Our two main axioms on preference—exclusion and inclusion—identify when the agent would want to restrict his choice set and when he would want to expand his choice set. An agent who is tempted would want to restrict his choice set by *excluding* the normatively worst alternative of that choice set, thus avoiding a potential conflict between a temptation and his normative preference. At the same time, he would want to expand his choice set by *including* an alternative that is normatively superior, thus potentially helping him make a normatively good choice.

Using the exclusion and inclusion axioms among others, Theorem 1 characterizes the representation

$$(1) \quad V_T(x) = \sum_{i=1}^I q_i \left\{ \max_{\beta \in x} [u(\beta) + v_i(\beta)] - \max_{\beta \in x} v_i(\beta) \right\},$$

¹I would like to thank Val Lambson, Eddie Dekel, Bart Lipman, and numerous seminar audiences for their comments. A co-editor and three anonymous referees provided very useful comments. I especially thank my advisor, Larry Epstein, for his guidance. The results presented here were originally distributed in a paper titled “Temptation and Self-Control as Duals.”

where $q_i > 0$ for all i , $\sum_i q_i = 1$, and u and each v_i are von Neumann–Morgenstern expected-utility functions. For any singleton menu, $V_T(\{\beta\}) = u(\beta)$. Hence u represents the agent’s normative preference. We interpret i to be a subjective state to which the agent assigns probability q_i . In state i , v_i is the temptation that affects the agent. He compromises between his normative preference and temptation preference and chooses the alternative that maximizes $u + v_i$. However, he experiences the disutility $\max_{\beta \in x} v_i(\beta)$, which is the foregone utility from the most tempting alternative. Thus V_T takes an expected-utility form, where $\max_{\beta \in x}[u(\beta) + v_i(\beta)] - \max_{\beta \in x} v_i(\beta)$ is the utility attained in state i , suggesting the agent is uncertain which of his temptations will affect him.

This work is most closely related to the seminal paper by [Gul and Pesendorfer \(2001\)](#) and a recent paper by [Dekel, Lipman, and Rustichini \(2009\)](#) (henceforth GP and DLR, respectively). Conceptually, our model can be viewed as a compromise between these two as the set of preferences we consider is a special case of DLR’s and a generalization of GP’s. In Section 3, we discuss the relationship between these models. Through examples, we argue for the relaxation of GP’s model and the strengthening of DLR’s model.

2. THE MODEL

The agent in our model chooses between menus, which are sets of lotteries. It is understood, though unmodeled, that he will later choose an alternative from the menu he chooses now. Thus we think of a menu as being the agent’s future choice set. We assume that when choosing a menu, the agent is in a “cool” state (meaning he is not tempted by any alternatives when considering a menu), but that he anticipates being in a “hot” state at the time of choosing an alternative.² The agent’s normative preference is identified with his preference over singleton menus.

Formally, let Δ denote the set of probability distributions over a finite set of prizes, and call $\beta \in \Delta$ a lottery. Let X denote the set of closed nonempty subsets of Δ and call $x \in X$ a menu. We endow X with the Hausdorff topology and define the mixture operation

$$\lambda x + (1 - \lambda)y \equiv \{\lambda\beta + (1 - \lambda)\beta' : \beta \in x, \beta' \in y\}$$

for $\lambda \in [0, 1]$. Our primitive is a binary relation \succeq over X which represents the agent’s preference. Consider the following axioms.

AXIOM 1—Weak Order: *\succeq is complete and transitive.*

AXIOM 2—Continuity: *The sets $\{x : x \succeq y\}$ and $\{x : y \succeq x\}$ are closed.*

²See [Noor \(2007\)](#) for a model that relaxes this assumption and allows an agent to be tempted at the time of choosing menus.

AXIOM 3—Independence: *If $x \succ y$, then for every $z \in X$ and $\lambda \in (0, 1]$,*

$$\lambda x + (1 - \lambda)z \succ \lambda y + (1 - \lambda)z.$$

These are straightforward extensions of the standard expected-utility axioms. See [Dekel, Lipman, and Rustichini \(2001\)](#) and GP for discussion of these axioms.

The next axiom was introduced by DLR and requires the following definition. First, for any menu x , let $\text{conv}(x)$ denote its convex hull.

DEFINITION 1: $x' \subset \text{conv}(x)$ is *critical for x* if for all y such that $x' \subset \text{conv}(y) \subset \text{conv}(x)$, we have $y \sim x$.

Observe that if x' is critical for x , then $x' \sim x$. We think of a critical set as stripping away the irrelevant alternatives of a menu. That is, suppose x' is critical for x and let $\beta \in x \setminus x'$. Then since x' is critical, we have $x' \cup \{\beta\} \sim x$. That is, adding β to x' does not affect the agent's ranking of x' . Since $\beta \in x$, we conclude then that β is not important to the decision maker when evaluating x .

AXIOM 4—Finiteness: *Every menu has a finite critical subset.*

Following DLR, we assume finiteness to simplify the analysis. See DLR for more discussion of this axiom.

The next two axioms are meant to capture the effects of temptation on preference.³

AXIOM 5—Exclusion: *If $\{\beta\} \succeq \{\alpha\}$ for every $\beta \in x$, then $x \succeq x \cup \{\alpha\}$.*

If every alternative in x normatively dominates α , then Axiom 5 (exclusion) states that the agent would prefer to exclude α from his menu. We can think of α as a bad alternative being added to a good menu x . If α tempts the agent, then this temptation will conflict with his normative preference. If the agent thinks he might choose α , then he would be choosing α over an alternative that is normatively superior. Thus adding α can only make the menu less desirable.

AXIOM 6—Inclusion: *If $\{\alpha\} \succeq \{\beta\}$ for every $\beta \in x$, then $x \cup \{\alpha\} \succeq x$.*

If α normatively dominates every alternative in x , then Axiom 6 (inclusion) states that the agent would prefer to include α in his menu. We can think of α as a good alternative being added to a bad menu x . If α tempts the agent,

³Inclusion was proposed independently by [Nehring \(2006\)](#) under the name “singleton monotonicity.” [Chandrasekher \(2009\)](#) considered an axiom labeled A1 that is similar to, though distinct from, exclusion.

there is no conflict with his normative preference. If the agent thinks he might choose α , then he would be choosing α over an alternative that is normatively inferior. Thus adding α can only make the menu more desirable.

Our utility representation takes the form of equation (1), which we call a *temptation representation*.

THEOREM 1: *The preference \succeq satisfies weak order, continuity, independence, finiteness, exclusion, and inclusion if and only if \succeq has a temptation representation.*

The proof is given in Appendix B.

REMARK 1: Finiteness is independent of our other axioms. As an example, preferences represented by

$$V(x) = \int \left\{ \max_{\beta \in x} [u(\beta) + v(\beta)] - \max_{\beta \in x} v(\beta) \right\} \mu(dv),$$

where μ is a measure (with possibly infinite support) over the set of von Neumann–Morgenstern expected-utility functions, would satisfy all our axioms but finiteness.⁴ This is in contrast to GP’s model. Though GP did not assume finiteness, it is implied by their main axiom, set betweenness.

REMARK 2: Though exclusion and inclusion are necessary, the proof of sufficiency does not actually use their full force. Specifically, we could replace exclusion with DLR’s axiom, desire for commitment. Alternatively, because of the symmetry of the representation and the axioms, we could replace inclusion with an axiom symmetric to desire for commitment, one which we call desire for better alternatives. However, it is not possible to weaken both exclusion and inclusion, as shown by a counterexample in Appendix C.

3. RELATED LITERATURE

3.1. Gul and Pesendorfer (2001)

GP were the first to use a preference-over-menus framework to model temptation. Their key axiom is set betweenness.

AXIOM 7—Set Betweenness: *If $x \succeq y$, then $x \succeq x \cup y \succeq y$.*

⁴In recent work, Dekel and Lipman (2007) adapted our proof of Theorem 1 to characterize such preferences. Besides dropping finiteness, their set of axioms differs from ours in two ways. First, they add a continuity axiom called Lipschitz continuity. See Dekel, Lipman, Rustichini, and Sarver (2007) for a discussion of this axiom. Second, they replace exclusion and inclusion with an axiom they call weak set betweenness. Lemma 1 shows that weak set betweenness is equivalent to exclusion and inclusion, given weak order and continuity.

GP's main theorem states that \succeq satisfies weak order, continuity, independence, and set betweenness if and only if \succeq has the following representation.

DEFINITION 2: A *self-control representation* is a function V_{GP} such that

$$(2) \quad V_{\text{GP}}(x) = \max_{\beta \in x} [u(\beta) + v(\beta)] - \max_{\beta \in x} v(\beta),$$

where u and v are von Neumann–Morgenstern expected-utility functions.

Observe that a self-control representation is a temptation representation where $I = 1$. Thus GP's model seems to not allow uncertainty about temptation. The following example, borrowed from DLR, uses uncertainty about temptation to explain a violation of set betweenness. This example, however, is consistent with exclusion and inclusion.

EXAMPLE 1: Suppose an agent is on a diet and has three snacks he could eat: broccoli (b), chocolate cake (c), and potato chips (p). Broccoli is the healthiest snack while chocolate cake and potato chips are equally unhealthy. Hence

$$\{b\} \succ \{c\} \sim \{p\}.$$

The agent thinks he will experience either a salt craving or a sugar craving. If he has a salt craving, then he is better off not having potato chips as an option. Hence

$$\{b\} \succ \{b, p\} \quad \text{and} \quad \{b, c\} \succ \{b, c, p\}.$$

If he has a sugar craving, then he is better off not having chocolate cake as an option. Hence

$$\{b\} \succ \{b, c\} \quad \text{and} \quad \{b, p\} \succ \{b, c, p\}.$$

This violates set betweenness since $\{b, c\} \cup \{b, p\} \not\succeq \{b, c\}, \{b, p\}$. However, it is consistent with exclusion and inclusion.

GP argued that, for preferences with a self-control representation, certain behavior (described below) reveals that an agent anticipates exerting self-control. However this interpretation is not justified for preferences with a temptation representation. Uncertainty about temptation is key to this difference.⁵

⁵Dekel and Lipman (2007) argued that GP's definition can have interpretations other than self-control, even for preferences with a self-control representation. Roughly, they showed that a self-control representation can be interpreted alternatively as allowing uncertainty in temptation.

DEFINITION 3: \succeq has *self-control at z* if there exists x, y such that $z = x \cup y$ and $x \succ x \cup y \succ y$.

As GP (2001, pp. 1410–1411) explained, “[$x \succ x \cup y$] captures the fact that [y] entails greater temptation than [x] while [$x \cup y \succ y$] captures the fact that the agent resists this temptation.”⁶ The intuition motivating this definition is that there are two requirements for an agent to exert self-control: First, he must be tempted; second, he must resist that temptation. GP supported their interpretation of this behavior with the following theorem.⁷ First, for any continuous $f: \Delta \rightarrow \mathbb{R}$, define

$$c(x, f) \equiv \arg \max_{\beta \in x} f(\beta).$$

THEOREM 2—Gul and Pesendorfer (2001, Theorem 2): Suppose \succeq has a self-control representation given by (2). The following statements are equivalent:

- (i) \succeq has self-control at x .
- (ii) $c(x, u + v) \cap c(x, v) = \emptyset$.

Property (ii) implies that the agent resists temptation since his anticipated choice is not a most tempting alternative.

Consider the following generalization of property (ii) for a temptation representation:

$$(*) \quad \text{There exists } i \text{ such that } c(x, u + v_i) \cap c(x, v_i) = \emptyset.$$

This statement captures the intuition behind GP’s definition of self-control since the agent would anticipate some state where he is tempted but resists the temptation. However, as the following example shows, GP’s definition of self-control does *not* characterize (*) for preferences with a temptation representation.

EXAMPLE 2: Suppose the agent has the following normative and temptation utilities for two alternatives, α and ω :

	u	v_1	v_2
α	2	2	0
ω	0	1	3

⁶Notation has been changed to be consistent with ours.

⁷This is not the full statement of GP’s theorem. We have omitted a portion that is not related to the present discussion.

Suppose also that he ranks menus by the temptation representation

$$V(x) = \sum_{i=1}^2 \frac{1}{2} \left\{ \max_{\beta \in x} [u(\beta) + v_i(\beta)] - \max_{\beta \in x} v_i(\beta) \right\}.$$

Then $\{\alpha\} \succ \{\alpha, \omega\} \succ \{\omega\}$. However, in no state does the agent anticipate exerting self-control. In state 1, the agent is most tempted by α (since $v_1(\alpha) > v_1(\omega)$), but he also expects to choose α (since $u(\alpha) + v_1(\alpha) > u(\omega) + v_1(\omega)$). Similarly in state 2, the agent is most tempted by ω , and he expects to choose ω . So even though the agent has “self-control” (according to GP’s definition) at $\{\alpha, \omega\}$, condition $(*)$ does not hold. Though $\{\alpha\} \succ \{\alpha, \omega\}$ captures the fact that ω is tempting in some state and $\{\alpha, \omega\} \succ \{\omega\}$ captures the fact that the agent expects to choose α in some state, *those two states are not the same*.

Hence GP’s definition is not enough to characterize $(*)$ for a temptation representation. In fact, Example 2 shows a little bit more. If \succeq has a temptation representation and $\{\alpha\} \succ \{\omega\}$, then there are only three possible orderings of $\{\alpha\}$, $\{\omega\}$, and $\{\alpha, \omega\}$:

- (i) $\{\alpha\} \sim \{\alpha, \omega\} \succ \{\omega\}$.
- (ii) $\{\alpha\} \succ \{\alpha, \omega\} \sim \{\omega\}$.
- (iii) $\{\alpha\} \succ \{\alpha, \omega\} \succ \{\omega\}$.

One can show that $\{\alpha\} \sim \{\alpha, \omega\}$ in ranking (i) implies

$$\alpha \in c(\{\alpha, \omega\}, u + v_i) \cap c(\{\alpha, \omega\}, v_i)$$

for every i .⁸ Similarly, $\{\alpha, \omega\} \sim \{\omega\}$ in ranking (ii) implies

$$\omega \in c(\{\alpha, \omega\}, u + v_i) \cap c(\{\alpha, \omega\}, v_i)$$

for every i . Hence, (iii) is the only possible ranking where $(*)$ might hold. Therefore, Example 2 shows that *for an arbitrary set*, there is no behavior that would reveal that an agent anticipates exerting self-control.

3.2. Dekel, Lipman, and Rustichini (2009)

Using Example 1 as part of their motivation to weaken set betweenness, DLR proposed their own axiom to capture temptation.

AXIOM 8—Desire for Commitment (DFC): *For every x , there exists $\alpha \in x$ such that $\{\alpha\} \succeq x$.*

⁸If not, then there exists i such that $v_i(\omega) > v_i(\alpha)$, which implies $u(\alpha) > u(\alpha) + v_i(\alpha) - v_i(\omega)$, which implies $V_T(\{\alpha\}) > V_T(\{\alpha, \omega\})$, a contradiction.

DLR showed that \succeq satisfies weak order, continuity, independence, finiteness, DFC, and a technical axiom called approximate improvements are chosen if and only if \succeq has the following representation.

DEFINITION 4: A *DLR temptation representation* is a function V_{DLR} such that

$$V_{\text{DLR}}(x) = \sum_{i=1}^I q_i \left\{ \max_{\beta \in x} \left[u(\beta) + \sum_{j \in J_i} v_j(\beta) \right] - \sum_{j \in J_i} \max_{\beta \in x} v_j(\beta) \right\},$$

where $q_i > 0$ for every i , $\sum_i q_i = 1$, and where u and each v_j are von Neumann–Morgenstern expected-utility functions.

Observe that if J_i is a singleton for every i , then this is a temptation representation. Hence, this is a generalization of a temptation representation where the agent can be affected by more than one temptation in each state.

We argue that a DLR temptation representation permits preferences which are not explained by temptation. Consider the following example.

EXAMPLE 3: Suppose an agent has two snacks he could eat: chocolate cake (c) and potato chips (p). Assume the ranking of menus

$$\{c\} \sim \{p\} \succ \{c, p\}.$$

This ranking is consistent with DFC but not with inclusion.⁹ It implies that chocolate cake adds some psychic cost to the agent when coupled with potato chips and vice versa. But the agent considers chocolate cake and potato chips to be normatively equivalent. Thus temptation is not an explanation for this since *there is no conflict with his normative preference*.¹⁰

This example suggests a need for a stronger set of axioms. DLR proposed their own strengthening of DFC.

AXIOM 9—Weak Set Betweenness: If $\{\alpha\} \succeq \{\beta\}$ for all $\alpha \in x$ and $\beta \in y$, then $x \succeq x \cup y \succeq y$.

In an earlier version of their paper, DLR conjectured that weak order, continuity, independence, finiteness, and weak set betweenness characterize a temptation representation. How does weak set betweenness relate to exclusion and

⁹This ranking is also consistent with DLR's other axiom, approximate improvements are chosen.

¹⁰We do not mean to say that preferences like Example 3 are unreasonable, only that temptation alone is not a good explanation for them. Such preferences may be explained by, for example, regret (e.g., Sarver (2008)) or perfectionism (e.g., Kopylov (2009)).

inclusion? It should be obvious that weak set betweenness implies exclusion and inclusion.¹¹ The following lemma shows that in the presence of our other axioms, the other direction holds as well.

LEMMA 1: *If \succeq satisfies weak order, continuity, exclusion, and inclusion, then \succeq satisfies weak set betweenness.*

PROOF: We show that the conclusion holds for finite menus. The result then follows from continuity and the fact that, in the Hausdorff topology, any menu is the limit of a sequence of finite menus (see GP, Lemma 0).

Let $x = \{\alpha_1, \dots, \alpha_M\}$ and $y = \{\beta_1, \dots, \beta_N\}$ satisfy

$$\{\alpha_1\} \succeq \{\alpha_2\} \succeq \dots \succeq \{\alpha_M\} \succeq \{\beta_1\} \succeq \{\beta_2\} \succeq \dots \succeq \{\beta_N\}.$$

By repeatedly applying exclusion, we obtain

$$\{\alpha_1, \dots, \alpha_M\} \succeq \{\alpha_1, \dots, \alpha_M\} \cup \{\beta_1, \dots, \beta_N\}$$

or $x \succeq x \cup y$. By repeatedly applying inclusion, we obtain

$$\{\alpha_1, \dots, \alpha_M\} \cup \{\beta_1, \dots, \beta_N\} \succeq \{\beta_1, \dots, \beta_N\}$$

or $x \cup y \succeq y$.

Q.E.D.

Hence Theorem 1 is a proof of DLR's conjecture. However, we prefer exclusion and inclusion over weak set betweenness because they are more basic axioms. This is desirable because it makes the assumptions on behavior more transparent.

APPENDIX A: NOTATION

Let $K \geq 3$ denote the number of outcomes or prizes. (Results are simple if $K = 2$.) Let $\mathbf{0}$ and $\mathbf{1}$ denote K -vectors of zeros and ones, respectively. We will use u, w_i, v_j, \dots to denote von Neumann–Morgenstern expected-utility functions as well as K -vectors of payoffs of pure outcomes, so that $u(\beta) = \beta \cdot u, \dots$. For any $f \in \mathbb{R}^K$, define

$$c(x, f) \equiv \arg \max_{\beta \in x} \beta \cdot f$$

and

$$H^f \equiv \{g \in \mathbb{R}^K : g \cdot f = 0\}.$$

¹¹In the statement of weak set betweenness, take y as a singleton to get exclusion and take x as a singleton to get inclusion.

In particular, $H^1 = \{g \in \mathbb{R}^K : g \cdot \mathbf{1} = 0\}$ is the set of vectors whose coordinates sum to zero. Let Δ° denote the relative interior of Δ , that is,

$$\Delta^\circ \equiv \{\beta = (\beta_k) \in \Delta : 0 < \beta_k < 1, k = 1, \dots, K\}.$$

APPENDIX B: PROOF OF THEOREM 1

We prove the sufficiency part of the theorem in Section B.1 followed by the necessity part in Section B.2.

B.1. Sufficiency of Axioms

The proof that the axioms are sufficient will proceed as follows. In Section B.1.1, we present some useful results for linear functionals. In Section B.1.2, we define a key intermediate representation, the finite additive expected-utility (EU) representation. We then prove some results for this representation and our axioms. In Section B.1.3, we use these results to finish the proof.

B.1.1. Some Results for Linear Functionals

We present the first two lemmas of this section without proof.

LEMMA 2: Suppose $u \in H^1$. Suppose x is a sphere in $H^u \cap \Delta^\circ$ and $f, g \in H^1 \setminus \{\mathbf{0}\}$. Write $f = au + \tilde{f}$ and $g = bu + \tilde{g}$, where $\tilde{f}, \tilde{g} \in H^u \cap H^1$. Then $c(x, f) = c(x, g)$ if and only if \tilde{f} is a positive scalar multiple of \tilde{g} . Furthermore, if $\tilde{f} \neq \mathbf{0}$, then $c(x, f)$ is a singleton.

We will use Lemma 2 often, especially the cases (i) $u = \mathbf{0}$, (ii) $\tilde{f} = g$, and (iii) $f, g \in H^u$.

DEFINITION 5: A set of vectors \mathcal{F} is *not redundant* if for every $f, g \in \mathcal{F}$, f is not a positive scalar multiple of g .

LEMMA 3: Suppose $u \in H^1$, and let $\{f_i\}_{i=1}^I$ and $\{g_j\}_{j=1}^J$ be two finite non-redundant sets of vectors in $H^u \cap H^1 \setminus \{\mathbf{0}\}$. Then

$$\sum_{i=1}^I \max_{\beta \in x} \beta \cdot f_i = \sum_{j=1}^J \max_{\beta \in x} \beta \cdot g_j \quad \text{for all closed } x \subset H^u \cap \Delta,$$

if and only if $I = J$ and, without loss of generality, $f_i = g_i$ for every i .

The following lemma generalizes Lemma 3 by allowing for redundancies and zero vectors.

LEMMA 4: Suppose $u \in H^1$, and let $\{f_i\}_{i=1}^I$ and $\{g_j\}_{j=1}^J$ be two finite sets of vectors in $H^u \cap H^1$. Then

$$\sum_{i=1}^I \max_{\beta \in x} \beta \cdot f_i = \sum_{j=1}^J \max_{\beta \in x} \beta \cdot g_j \quad \text{for all closed } x \subset H^u \cap \Delta,$$

if and only if there exist (i) $I' \subset \{1, \dots, I\}$ and $J' \subset \{1, \dots, J\}$, (ii) I_1, \dots, I_N , a partition of I' , and J_1, \dots, J_M , a partition of J' , (iii) positive scalars $\{p_i\}_{i \in I'}$ and $\{q_j\}_{j \in J'}$, where $\sum_{i \in I_n} p_i = \sum_{j \in J_n} q_j = 1$ for every $n \in \{1, \dots, N\}$, and (iv) $\{h_1, \dots, h_N\} \subset H^u \cap H^1 \setminus \{\mathbf{0}\}$ not redundant, such that

$$f_i = \mathbf{0} \quad \forall i \notin I'$$

and

$$g_j = \mathbf{0} \quad \forall j \notin J',$$

and such that for every n ,

$$f_i = p_i h_n \quad \forall i \in I_n$$

and

$$g_j = q_j h_n \quad \forall j \in J_n.$$

PROOF: The “if” part is straightforward, so we just prove the “only if” part. Define $I' \equiv \{i : f_i \neq \mathbf{0}\}$ and $J' \equiv \{j : g_j \neq \mathbf{0}\}$. Partition I' into I_1, \dots, I_N , where $i, i' \in I_n$ if and only if f_i is a positive scalar multiple of $f_{i'}$. Similarly, partition J' into J_1, \dots, J_M , where $j, j' \in J_m$ if and only if g_j is a positive scalar multiple of $g_{j'}$. Then we have

$$\sum_{n=1}^N \max_{\beta \in x} \beta \cdot \left(\sum_{i \in I_n} f_i \right) = \sum_{m=1}^M \max_{\beta \in x} \beta \cdot \left(\sum_{j \in J_m} g_j \right)$$

for all closed $x \subset H^u \cap \Delta,$

where $\{\sum_{i \in I_n} f_i\}_{n=1}^N$ and $\{\sum_{j \in J_m} g_j\}_{m=1}^M$ are finite sets of vectors in $H^u \cap H^1 \setminus \{\mathbf{0}\}$, each of which is not redundant. Lemma 3 then implies that $N = M$, and that $\sum_{i \in I_n} f_i = \sum_{j \in J_n} g_j$ for every n . Define $h_n \equiv \sum_{i \in I_n} f_i$ for every n . Observe that $h_n \in H^1 \cap H^u \setminus \{\mathbf{0}\}$ for every n and that $\{h_1, \dots, h_N\}$ is not redundant.

For every $i \in I'$, let $n(i)$ denote the n such that $i \in I_n$. Observe that for every $i \in I'$, f_i is a positive scalar multiple of $h_{n(i)}$. So for every $i \in I'$, define $p_i > 0$ by the equation $f_i = p_i h_{n(i)}$. Similarly, for every $j \in J'$, let $n(j)$ denote the n such that $j \in J_n$. For every $j \in J'$, define $q_j > 0$ by the equation $g_j = q_j h_{n(j)}$. Hence $\sum_{i \in I_n} p_i = \sum_{j \in J_n} q_j = 1$ for every n . $Q.E.D.$

B.1.2. Finite Additive EU Representation and Some Preliminary Results

In their appendix, DLR proved the following theorem.

THEOREM 3—Dekel, Lipman, and Rustichini (2009, Theorem 6): *The preference \succeq satisfies weak order, continuity, independence, and finiteness if and only if \succeq has the representation*

$$(3) \quad V(x) = \sum_{i=1}^I \max_{\beta \in x} w_i(\beta) - \sum_{j=1}^J \max_{\beta \in x} v_j(\beta),$$

where each w_i and each v_j is a von Neumann–Morgenstern expected-utility function.

A representation of the form given in (3) is called a *finite additive EU representation*. It is a modified version of the set of preferences studied by Dekel, Lipman, and Rustichini (2001). For such a representation, define

$$(4) \quad u \equiv \sum_i w_i - \sum_j v_j,$$

which represents preference over singleton menus.

DEFINITION 6: A finite additive EU representation given by (3) is in *reduced form* if $\{w_1, \dots, w_I, v_1, \dots, v_J\} \subset H^1 \setminus \{\mathbf{0}\}$ and is not redundant.

LEMMA 5: *If \succeq has a finite additive EU representation, then it has a reduced form finite additive EU representation.*

The proof of Lemma 5 is straightforward, so we omit it. The following lemma shows some implications of exclusion and inclusion.¹²

LEMMA 6: *Suppose \succeq has a reduced form finite additive EU representation, and satisfies exclusion and inclusion. Then for every i , w_i is not a positive scalar multiple of $-u$ and for every j , v_j is not a positive scalar multiple of u .*

PROOF: We prove only the first part, which uses exclusion. The proof of the second part is similar and uses inclusion.

By way of contradiction, suppose i^* is such that $w_{i^*} = -au$ for some $a > 0$. Let x be a sphere in Δ° . Since $w_{i^*} \in H^1 \setminus \{\mathbf{0}\}$, Lemma 2 implies $c(x, w_{i^*}) = \{\beta_{i^*}\}$

¹²Lemmas 6 and 7 can each be proven with the weaker axioms DFC and desire for better alternatives replacing exclusion and inclusion. See Appendix C for a formal statement of desire for better alternatives. Thus Lemma 11 is the only result in the proof where exclusion or inclusion is needed, and only one of these is needed, not both. See Remark 2.

is a singleton and that $\{\beta_{i^*}\} = c(x, -u)$ or $\{\beta_{i^*}\} = \arg \min_{\beta \in x} \beta \cdot u$. Since v_j is not a positive scalar multiple of w_{i^*} and nonzero for every j , Lemma 2 also implies

$$\max_{\beta \in x} \beta \cdot v_j > \beta_{i^*} \cdot v_j \quad \forall j.$$

Since $\beta_{i^*} \in \Delta^\circ$ and $u \in H^1$, there exists $\varepsilon > 0$ such that $\beta^* \equiv \beta_{i^*} - \varepsilon u \in \Delta$ and satisfies

$$(5) \quad \max_{\beta \in x} \beta \cdot v_j > \beta^* \cdot v_j \quad \forall j.$$

Observe that $\beta^* \cdot w_{i^*} > \beta_{i^*} \cdot w_{i^*}$ since $(-\varepsilon u) \cdot w_{i^*} = a\varepsilon u \cdot u > 0$. Also observe that $\beta^* \cdot u < \beta_{i^*} \cdot u$, which implies that $\{\beta\} \succ \{\beta^*\}$ for every $\beta \in x$.

Now consider the menu $x \cup \{\beta^*\}$. Inequality (5) implies

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot v_j = \max_{\beta \in x} \beta \cdot v_j \quad \forall j.$$

We also have

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i \geq \max_{\beta \in x} \beta \cdot w_i \quad \forall i,$$

with a strict inequality for i^* . Therefore, $V(x \cup \{\beta^*\}) > V(x)$. But this contradicts exclusion since $\{\beta\} \succ \{\beta^*\}$ for every $\beta \in x$. *Q.E.D.*

Next we introduce a key axiom that will be useful for proving an intermediate result later (Lemma 9).

DEFINITION 7: A menu x is *constant* if for every $\beta, \beta' \in x$, $\{\beta\} \sim \{\beta'\}$.

AXIOM 10—Constant Menus Are Not Tempted (CMNT): *For every constant menu x , for every $\alpha \in x$, $\{\alpha\} \sim x$.*

Intuitively, CMNT states that a constant menu cannot tempt since there can be no conflict between an agent's normative preference and a temptation. The following lemma shows that CMNT is implied by our axioms.

LEMMA 7: *Suppose \succeq satisfies weak order, continuity, exclusion, and inclusion. Then \succeq satisfies CMNT.*

The proof is similar to that of Lemma 1, so we omit it. The next lemma takes care of a trivial case.

LEMMA 8: *Suppose \succeq has a reduced form finite additive EU representation and satisfies CMNT. If $u = \mathbf{0}$, then $I = J = 0$.*

PROOF: If $u = \mathbf{0}$, then $\{\alpha\} \sim \{\beta\}$ for every $\alpha, \beta \in \Delta$. Hence, for every $x \in X$, x is a constant menu. CMNT then implies $V(x) = 0$ for every $x \in X$ or

$$\sum_i \max_{\beta \in x} \beta \cdot w_i = \sum_j \max_{\beta \in x} \beta \cdot v_j \quad \forall x \in X.$$

But since $\{w_1, \dots, w_I\} \subset H^1 \setminus \{\mathbf{0}\}$ is not redundant and $\{v_1, \dots, v_J\} \subset H^1 \setminus \{\mathbf{0}\}$ is not redundant, Lemma 3 implies $I = J$ and $w_i = v_i$ for every i . But since $\{w_1, \dots, w_I, v_1, \dots, v_J\}$ is not redundant, it must be that $I = J = 0$. *Q.E.D.*

The following lemma shows the implications of CMNT.

LEMMA 9: *Suppose \succeq has a reduced form finite additive EU representation and satisfies CMNT. Then there are (i) scalars a_1, \dots, a_I and b_1, \dots, b_J , where $\sum_{i=1}^I a_i - \sum_{j=1}^J b_j = 1$, (ii) $I' \subset \{1, \dots, I\}$ and $J' \subset \{1, \dots, J\}$, (iii) I_1, \dots, I_N , a partition of I' , and J_1, \dots, J_N , a partition of J' , (iv) positive scalars $\{p_i\}_{i \in I'}$ and $\{q_j\}_{j \in J'}$, where $\sum_{i \in I_n} p_i = \sum_{j \in J_n} q_j = 1$ for every $n \in \{1, \dots, N\}$, and (v) $\{f_1, \dots, f_N\} \subset H^u \cap H^1 \setminus \{\mathbf{0}\}$ not redundant, such that*

$$w_i = a_i u \quad \forall i \notin I'$$

and

$$v_j = b_j u \quad \forall j \notin J',$$

and such that for every $n \in \{1, \dots, N\}$,

$$w_i = a_i u + p_i f_n \quad \forall i \in I_n$$

and

$$v_j = b_j u + q_j f_n \quad \forall j \in J_n.$$

PROOF: If $u = \mathbf{0}$, then Lemma 8 implies that the result is trivial. So assume $u \neq \mathbf{0}$.

Observe that $u \in H^1$. Observe also that for every i , there is $\tilde{w}_i \in H^u \cap H^1$ and scalar a_i such that

$$(6) \quad w_i = a_i u + \tilde{w}_i,$$

and for every j , there is $\tilde{v}_j \in H^u \cap H^1$ and scalar b_j such that

$$(7) \quad v_j = b_j u + \tilde{v}_j.$$

Then $u = (\sum_i a_i - \sum_j b_j)u + \sum_i \tilde{w}_i - \sum_j \tilde{v}_j$. Since $u \neq \mathbf{0}$ and $\tilde{w}_i, \tilde{v}_j \in H^u$ for every i and j , this means $\sum_i a_i - \sum_j b_j = 1$.

Let x be any constant menu. Set $\bar{u} \equiv \beta \cdot u$ for any $\beta \in x$. Then CMNT implies that

$$\bar{u} = \sum_{i=1}^I \max_{\beta \in x} \beta \cdot w_i - \sum_{j=1}^J \max_{\beta \in x} \beta \cdot v_j.$$

Hence, using (6) and (7),

$$\begin{aligned} \bar{u} &= \sum_{i=1}^I \max_{\beta \in x} [a_i \bar{u} + \beta \cdot \tilde{w}_i] - \sum_{j=1}^J \max_{\beta \in x} [b_j \bar{u} + \beta \cdot \tilde{v}_j] \\ &= \bar{u} + \sum_{i=1}^I \max_{\beta \in x} \beta \cdot \tilde{w}_i - \sum_{j=1}^J \max_{\beta \in x} \beta \cdot \tilde{v}_j, \end{aligned}$$

which implies $\sum_I \max_{\beta \in x} \beta \cdot \tilde{w}_i = \sum_J \max_{\beta \in x} \beta \cdot \tilde{v}_j$ for any constant menu x . By Lemma 4, there are (i) $I' \subset \{1, \dots, I\}$ and $J' \subset \{1, \dots, J\}$, (ii) I_1, \dots, I_N , a partition of I' , and J_1, \dots, J_N , a partition of J' , (iii) positive scalars $\{p_i\}_{i \in I'}$ and $\{q_j\}_{j \in J'}$, where $\sum_{i \in I_n} p_i = \sum_{j \in J_n} q_j = 1$ for every $n \in \{1, \dots, N\}$, and (iv) $f_1, \dots, f_N \in H^1 \cap H^u \setminus \{\mathbf{0}\}$, such that

$$\tilde{w}_i = \mathbf{0} \quad \forall i \notin I'$$

and

$$\tilde{v}_j = \mathbf{0} \quad \forall j \notin J',$$

and such that for every n ,

$$\tilde{w}_i = p_i f_n \quad \forall i \in I_n$$

and

$$\tilde{v}_j = q_j f_n \quad \forall j \in J_n.$$

Inserting these into (6) and (7) gives us our result. *Q.E.D.*

B.1.3. Finishing the Proof of Sufficiency

We now show that our axioms are sufficient for a temptation representation.

Let \succeq satisfy weak order, continuity, independence, finiteness, exclusion and inclusion. By Theorem 3 and Lemma 5, \succeq has a reduced form finite additive EU representation V of the form given in (3). Define u by equation (4). By Lemma 7, \succeq satisfies CMNT. We apply Lemma 9 to get (i) scalars a_1, \dots, a_I and b_1, \dots, b_J , where $\sum_{i=1}^I a_i - \sum_{j=1}^J b_j = 1$, (ii) $I' \subset \{1, \dots, I\}$ and

$J' \subset \{1, \dots, J\}$, (iii) I_1, \dots, I_N , a partition of I' , and J_1, \dots, J_N , a partition of J' , (iv) positive scalars $\{p_i\}_{i \in I'}$ and $\{q_j\}_{j \in J'}$, where $\sum_{i \in I_n} p_i = \sum_{j \in J_n} q_j = 1$ for every $n \in \{1, \dots, N\}$, and (v) $\{f_1, \dots, f_N\} \subset H^u \cap H^1 \setminus \{\mathbf{0}\}$ not redundant, such that

$$w_i = a_i u \quad \forall i \notin I'$$

and

$$v_j = b_j u \quad \forall j \notin J',$$

and such that for every $n \in \{1, \dots, N\}$,

$$(8) \quad w_i = a_i u + p_i f_n \quad \forall i \in I_n$$

and

$$(9) \quad v_j = b_j u + q_j f_n \quad \forall j \in J_n.$$

To interpret this geometrically, for every n , $\pm u$ and f_n define a half-plane, and we can think of I_n and J_n as the subsets of the w_i 's and v_j 's that lie in that half-plane. Since u and f_n are orthogonal, we can think of the ratio a_i/p_i as describing the angle that w_i makes with u and f_n . Similarly, we can think of the ratio b_j/q_j as describing the angle that v_j makes with u and f_n . Figure 1 illustrates.

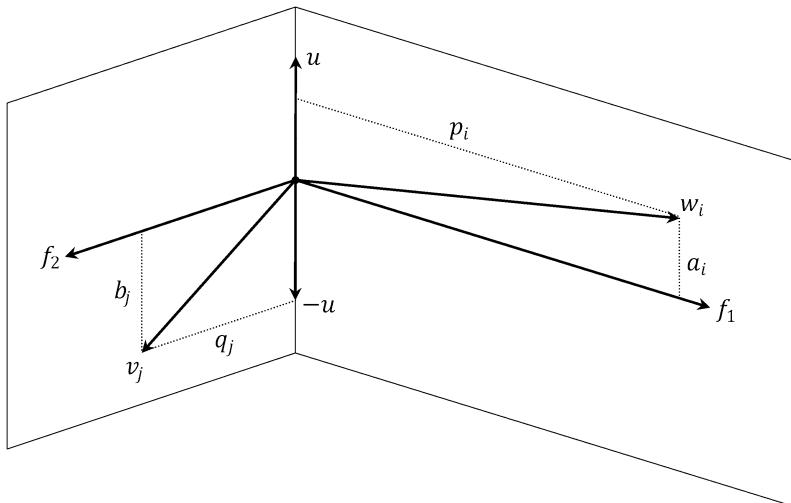


FIGURE 1.—Half-planes for f_1 and f_2 . So $i \in I_1$ and $w_i = a_i u + p_i f_1$. Also, $j \in J_2$ and $v_j = b_j u + q_j f_2$.

The following notation will be useful. For $i \in I'$, let $n(i)$ denote the n such that $i \in I_n$. Similarly, for $j \in J'$, let $n(j)$ denote the n such that $j \in J_n$. For any $n \in \{1, \dots, N\}$ and for any $\theta \in \mathbb{R}$, define the sets

$$I_n(\theta) \equiv \left\{ i \in I_n : \frac{a_i}{p_i} \leq \theta \right\}$$

and

$$J_n(\theta) \equiv \left\{ j \in J_n : \frac{b_j}{q_j} < \theta \right\}.$$

Thus $I_n(\theta)$ is the set of the w_i 's that make an angle with f_n weakly less than $\arctan(\theta \|u\|/\|f_n\|)$. Similarly, $J_n(\theta)$ is the set of v_j 's that make an angle with f_n strictly less than $\arctan(\theta \|u\|/\|f_n\|)$.

Before proceeding, we outline the rest of the proof to help the reader follow the argument. Using the fact that $\{w_1, \dots, w_I, v_1, \dots, v_J\}$ is not redundant, Lemma 10 shows that we can slightly perturb θ without changing the sets $I_n(\theta)$ and $J_n(\theta)$. Lemma 11 is the key step in the proof. It shows the relationship between the p_i 's and q_j 's for the expected-utility functions in a given half-plane. This is proved by contradiction using Lemma 10 to construct a lottery and menu that would otherwise violate exclusion.¹³ Lemma 12 then shows that we can take any w_i and v_j in the same half-plane and write

$$d_{ij}w_i = c_{ij}u + e_{ij}v_j,$$

where c_{ij} , d_{ij} , and e_{ij} have appropriate properties. Substituting this into a finite additive EU representation gives the result.

LEMMA 10: *For every $n \in \{1, \dots, N\}$ and for every $\theta^* \in \mathbb{R}$, there exists an interval $[\underline{\theta}, \bar{\theta}] \ni \theta^*$, where $\underline{\theta} < \bar{\theta}$, such that*

$$I_n(\theta) = I_n(\theta^*) \quad \forall \theta \in [\underline{\theta}, \bar{\theta}]$$

and

$$J_n(\theta) = J_n(\theta^*) \quad \forall \theta \in [\underline{\theta}, \bar{\theta}].$$

Figure 2 illustrates Lemma 10.

PROOF: Fix n and θ^* . Since $\{w_1, \dots, w_I, v_1, \dots, v_J\}$ is not redundant, $a_i/p_i \neq b_j/q_j$ for every $i \in I_n$ and $j \in J_n$. Hence, it cannot be that $a_i/p_i = b_j/q_j = \theta^*$ for some $i \in I_n$ and $j \in J_n$. So we consider two overlapping cases.

¹³Alternatively, one could prove Lemma 11 by constructing a lottery and menu that would otherwise violate inclusion.

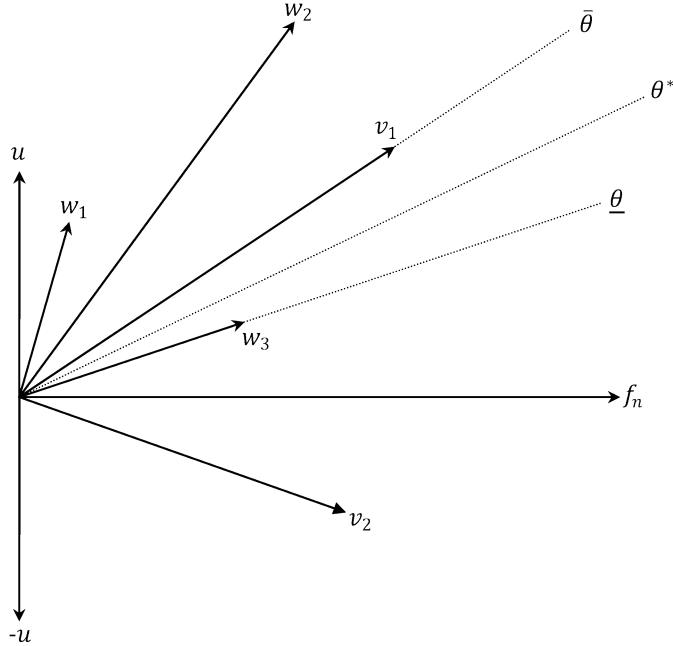


FIGURE 2.— $\bar{\theta} = \min_{j \in J_n \setminus J_n(\theta^*)} b_j/q_j$ and $\underline{\theta} = \max_{i \in I_n(\theta^*)} a_i/p_i$. It is easy to see that $I_n(\theta) = I_n(\theta^*)$ and $J_n(\theta) = J_n(\theta^*)$ for any $\theta \in [\underline{\theta}, \bar{\theta}]$.

Case 1— $a_i/p_i \neq \theta^* \forall i \in I_n$. By definition, $I_n(\theta^*) = \{i \in I_n : a_i/p_i \leq \theta^*\}$. But since $a_i/p_i \neq \theta^*$ for every $i \in I_n$, then $I_n(\theta^*) = \{i \in I_n : a_i/p_i < \theta^*\}$. So there exists $\varepsilon > 0$ such that $I_n(\theta^* - \varepsilon) = I_n(\theta^*)$ and $J_n(\theta^* - \varepsilon) = J_n(\theta^*)$. Set $\underline{\theta} \equiv \theta^* - \varepsilon$ and $\bar{\theta} \equiv \theta^*$.

Case 2— $b_j/q_j \neq \theta^* \forall j \in J_n$. Observe that $J_n \setminus J_n(\theta^*) = \{j \in J_n : b_j/q_j \geq \theta^*\}$. But since $b_j/q_j \neq \theta^*$ for every $j \in J_n$, then $J_n \setminus J_n(\theta^*) = \{j \in J_n : b_j/q_j > \theta^*\}$. So there exists $\varepsilon > 0$ such that $J_n \setminus J_n(\theta^* + \varepsilon) = J_n \setminus J_n(\theta^*)$ and $I_n \setminus I_n(\theta^* + \varepsilon) = I_n \setminus I_n(\theta^*)$, which implies $J_n(\theta^* + \varepsilon) = J_n(\theta^*)$ and $I_n(\theta^* + \varepsilon) = I_n(\theta^*)$. Set $\underline{\theta} \equiv \theta^*$ and $\bar{\theta} \equiv \theta^* + \varepsilon$. *Q.E.D.*

Our next lemma shows the relationship between the w_i 's and v_j 's in one of the half-planes.

LEMMA 11: *For every $n \in \{1, \dots, N\}$ and for every $\theta \in \mathbb{R}$, we have*

$$\sum_{i \in I_n(\theta)} p_i - \sum_{j \in J_n(\theta)} q_j \leq 0.$$

PROOF: If $u = \mathbf{0}$, then $I = J = \emptyset$ by Lemma 8 and the result is vacuous. So assume $u \neq \mathbf{0}$.

Now suppose the result is not true, that is, there exist n^* and θ^* such that

$$(10) \quad \sum_{i \in I_{n^*}(\theta^*)} p_i - \sum_{j \in J_{n^*}(\theta^*)} q_j > 0.$$

By Lemma 10, there exists a nonsingleton interval $[\underline{\theta}, \bar{\theta}] \ni \theta^*$ such that

$$(11) \quad I_{n^*}(\theta) = I_{n^*}(\theta^*) \quad \forall \theta \in [\underline{\theta}, \bar{\theta}]$$

and

$$(12) \quad J_{n^*}(\theta) = J_{n^*}(\theta^*) \quad \forall \theta \in [\underline{\theta}, \bar{\theta}].$$

In the remainder of the proof, we construct a menu $x \cup \{\beta^*\}$ and lottery β^{**} that violate exclusion given (10). We show this by determining where each expected-utility function (i.e., $w_1, \dots, w_I, v_1, \dots, v_J$) attains its maximum on $x \cup \{\beta^*\}$ and then again on $x \cup \{\beta^*\} \cup \{\beta^{**}\}$. Figure 3 collects these results. As Figure 3 shows, the only functions whose maximum changes when β^{**} is added to the menu are those associated with $I_{n^*}(\theta^*)$ and $J_{n^*}(\theta^*)$. This allows us to use (10) to show that $V(x \cup \{\beta^*\}) < V(x \cup \{\beta^*\} \cup \{\beta^{**}\})$, violating exclusion.

Let x be a sphere in $H^u \cap \Delta^\circ$. Observe that x is a constant menu. Since $f_{n^*} \in H^u \cap H^1 \setminus \{\mathbf{0}\}$, Lemma 2 implies that $c(x, f_{n^*}) = \{\beta_{n^*}\}$ is a singleton and that

$$(13) \quad \max_{\beta \in x} \beta \cdot w_i = \beta_{n^*} \cdot w_i \quad \forall i \in I_{n^*}.$$

Also, since $\{f_1, \dots, f_N\} \subset H^u \cap H^1 \setminus \{\mathbf{0}\}$ is not redundant, Lemma 2 implies

$$(14) \quad \max_{\beta \in x} \beta \cdot w_i > \beta_{n^*} \cdot w_i \quad \forall i \in I' \setminus I_{n^*}.$$

	$x \cup \{\beta^*\}$	$x \cup \{\beta^*\} \cup \{\beta^{**}\}$
$\{1, \dots, I\} \setminus I'$	x	x
$\{1, \dots, J\} \setminus J'$	β^*	β^*
$I' \setminus I_{n^*}$	x	x
$J' \setminus J_{n^*}$	x	x
$I_{n^*} \setminus I_{n^*}(\theta^*)$	x	x
$J_{n^*} \setminus J_{n^*}(\theta^*)$	x	x
$I_{n^*}(\theta^*)$	β^*	β^{**}
$J_{n^*}(\theta^*)$	β^*	β^{**}

FIGURE 3.—Where the maximum for the respective expected-utility function is attained. For example, the third row and first column show $\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i = \max_{\beta \in x} \beta \cdot w_i$ for $i \in I' \setminus I_{n^*}$.

Similarly, we can show

$$(15) \quad \max_{\beta \in x} \beta \cdot v_j = \beta_{n^*} \cdot v_j \quad \forall j \in J_{n^*}$$

and

$$(16) \quad \max_{\beta \in x} \beta \cdot v_j > \beta_{n^*} \cdot v_j \quad \forall j \in J' \setminus J_{n^*}.$$

Observe that $(\underline{\theta}/(f_{n^*} \cdot f_{n^*}))f_{n^*} - (1/(u \cdot u))u \in H^1$ since $u, f_{n^*} \in H^1$. Hence, $\beta_{n^*} \in \Delta^\circ$, and inequalities (14) and (16) imply that there exists $\varepsilon > 0$ such that

$$\beta^* \equiv \beta_{n^*} + \varepsilon \left(\frac{\underline{\theta}}{f_{n^*} \cdot f_{n^*}} f_{n^*} - \frac{1}{u \cdot u} u \right) \in \Delta^\circ,$$

$$(17) \quad \max_{\beta \in x} \beta \cdot w_i > \beta^* \cdot w_i \quad \forall i \in I' \setminus I_{n^*},$$

and

$$(18) \quad \max_{\beta \in x} \beta \cdot v_j > \beta^* \cdot v_j \quad \forall j \in J' \setminus J_{n^*}.$$

Observe that $\beta^* \cdot u = \beta_{n^*} \cdot u - \varepsilon$, which implies $\beta^* \cdot u = \min_{\beta \in x \cup \{\beta^*\}} \beta \cdot u$ since x is constant. Geometrically, β^* is a lottery that is an ε move from β_{n^*} in the direction of $(\underline{\theta}/(f_{n^*} \cdot f_{n^*}))f_{n^*} - (1/(u \cdot u))u$.

Consider the menu $x \cup \{\beta^*\}$. To fill in the column “ $x \cup \{\beta^*\}$ ” in Figure 3, we must determine the maximizers over $x \cup \{\beta^*\}$ for the expected-utility functions $\{w_1, \dots, w_I, v_1, \dots, v_J\}$. We show this only for the expected-utility functions associated with I_{n^*} and J_{n^*} , and leave the rest to the reader.

Using (8), observe that $\beta^* \cdot w_i = \beta_{n^*} \cdot w_i + \varepsilon p_i (\underline{\theta} - a_i/p_i)$ for $i \in I_{n^*}$. Since $\varepsilon > 0$ and $p_i > 0$, this implies $\beta^* \cdot w_i \geq \beta_{n^*} \cdot w_i$ if and only if $i \in I_{n^*}(\underline{\theta})$. Hence equations (11) and (13) imply

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i = \beta^* \cdot w_i \quad \forall i \in I_{n^*}(\theta^*)$$

and

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i = \max_{\beta \in x} \beta \cdot w_i \quad \forall i \in I_{n^*} \setminus I_{n^*}(\theta^*).$$

Similarly, using equations (12) and (15), we obtain

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot v_j = \beta^* \cdot v_j \quad \forall j \in J_{n^*}(\theta^*)$$

and

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot v_j = \max_{\beta \in x} \beta \cdot v_j \quad \forall j \in J_{n^*} \setminus J_{n^*}(\theta^*).$$

Now we construct the lottery β^{**} that will lead to a contradiction of exclusion. The idea is that we will take the lottery β^* and move a small distance in the direction of f_{n^*} . Since f_{n^*} and u are orthogonal, this will not change the commitment utility u . However, we will use inequality (10) to show that adding this lottery to our menu x must increase the utility of the menu.

Equations (17) and (18) imply

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i > \beta^* \cdot w_i \quad \forall i \in I' \setminus I_{n^*}$$

and

$$\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot v_j > \beta^* \cdot v_j \quad \forall j \in J' \setminus J_{n^*}.$$

Since $f_{n^*} \in H^1$ and $\beta^* \in \Delta^\circ$, there exists $\hat{\varepsilon}' > 0$ such that $\beta^* + \hat{\varepsilon}' f_{n^*} \in \Delta$ and

$$(19) \quad \max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i > (\beta^* + \hat{\varepsilon}' f_{n^*}) \cdot w_i \quad \forall i \in I' \setminus I_{n^*}$$

and

$$(20) \quad \max_{\beta \in x \cup \{\beta^*\}} \beta \cdot v_j > (\beta^* + \hat{\varepsilon}' f_{n^*}) \cdot v_j \quad \forall j \in J' \setminus J_{n^*}.$$

Define $\hat{\varepsilon}'' \equiv \varepsilon((\bar{\theta} - \underline{\theta})/(f_{n^*} \cdot f_{n^*})) > 0$. Observe that by the definition of β^* ,

$$\beta^* + \hat{\varepsilon}'' f_{n^*} = \beta_{n^*} + \varepsilon \left(\frac{\bar{\theta}}{f_{n^*} \cdot f_{n^*}} f_{n^*} - \frac{1}{u \cdot u} u \right).$$

Hence, using (8), $(\beta^* + \hat{\varepsilon}'' f_{n^*}) \cdot w_i = \beta_{n^*} \cdot w_i + \varepsilon p_i(\bar{\theta} - a_i/p_i)$ for $i \in I_{n^*}$. This implies $(\beta^* + \hat{\varepsilon}'' f_{n^*}) \cdot w_i \leq \beta_{n^*} \cdot w_i$ if $i \in I_{n^*} \setminus I_{n^*}(\bar{\theta})$. Hence by (11),

$$(21) \quad \beta_{n^*} \cdot w_i \geq (\beta^* + \hat{\varepsilon}'' f_{n^*}) \cdot w_i \quad \forall i \in I_{n^*} \setminus I_{n^*}(\theta^*).$$

Similarly, using (12), we obtain

$$(22) \quad \beta_{n^*} \cdot v_j > (\beta^* + \hat{\varepsilon}'' f_{n^*}) \cdot v_j \quad \forall j \in J_{n^*} \setminus J_{n^*}(\theta^*).$$

Set $\hat{\varepsilon} \equiv \min\{\hat{\varepsilon}', \hat{\varepsilon}''\}$ and $\beta^{**} \equiv \beta^* + \hat{\varepsilon} f_{n^*}$. Hence $\beta^{**} \in \Delta$, so consider the menu $x \cup \{\beta^*\} \cup \{\beta^{**}\}$. First, observe that $\beta^{**} \cdot u = \beta^* \cdot u = \min_{\beta \in x \cup \{\beta^*\}} \beta \cdot u$. Hence, $\{\beta\} \succeq \{\beta^{**}\}$ for every $\beta \in x \cup \{\beta^*\}$. To fill in the column “ $x \cup \{\beta^*\} \cup \{\beta^{**}\}$ ” in Figure 3, we must determine the maximizers over $x \cup \{\beta^*\} \cup \{\beta^{**}\}$ for the expected-utility functions $\{w_1, \dots, w_I, v_1, \dots, v_J\}$. Again, we show this only for the expected-utility functions associated with I_{n^*} and J_{n^*} , and leave the rest to the reader. (Inequalities (19) and (20) essentially give us the results for $I' \setminus I_{n^*}$ and $J' \setminus J_{n^*}$.)

Inequality (21) implies

$$\beta_{n^*} \cdot w_i \geq \beta^{**} \cdot w_i \quad \forall i \in I_{n^*} \setminus I_{n^*}(\theta^*),$$

which implies

$$\max_{\beta \in x \cup \{\beta^*\} \cup \{\beta^{**}\}} \beta \cdot w_i = \max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i \quad \forall i \in I_{n^*} \setminus I_{n^*}(\theta^*).$$

Inequality (22) implies

$$\beta_{n^*} \cdot v_j > \beta^{**} \cdot v_j \quad \forall j \in J_{n^*} \setminus J_{n^*}(\theta^*),$$

which implies

$$\max_{\beta \in x \cup \{\beta^*\} \cup \{\beta^{**}\}} \beta \cdot v_j = \max_{\beta \in x \cup \{\beta^*\}} \beta \cdot v_j \quad \forall j \in J_{n^*} \setminus J_{n^*}(\theta^*).$$

Using (8), observe that $\beta^{**} \cdot w_i = \beta^* \cdot w_i + \hat{\varepsilon} p_i f_{n^*} \cdot f_{n^*}$ for $i \in I_{n^*}$. But $\hat{\varepsilon} p_i f_{n^*} \cdot f_{n^*} > 0$, so this implies $\beta^{**} \cdot w_i > \beta^* \cdot w_i$ for every $i \in I_{n^*}$. Hence, this holds for every $i \in I_{n^*}(\theta^*)$ or

$$\beta^{**} \cdot w_i > \beta^* \cdot w_i \quad \forall i \in I_{n^*}(\theta^*),$$

which implies

$$\max_{\beta \in x \cup \{\beta^*\} \cup \{\beta^{**}\}} \beta \cdot w_i = \beta^{**} \cdot w_i \quad \forall i \in I_{n^*}(\theta^*),$$

since $\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot w_i = \beta^* \cdot w_i$ for $i \in I_{n^*}(\theta^*)$. Similarly,

$$\beta^{**} \cdot v_j > \beta^* \cdot v_j \quad \forall j \in J_{n^*}(\theta^*),$$

which implies

$$\max_{\beta \in x \cup \{\beta^*\} \cup \{\beta^{**}\}} \beta \cdot v_j = \beta^{**} \cdot v_j \quad \forall j \in J_{n^*}(\theta^*),$$

since $\max_{\beta \in x \cup \{\beta^*\}} \beta \cdot v_j = \beta^* \cdot v_j$ for $j \in J_{n^*}(\theta^*)$.

As is evident from viewing Figure 3, the only expected-utility functions that increase by going from $x \cup \{\beta^*\}$ to $x \cup \{\beta^*\} \cup \{\beta^{**}\}$ are those associated with the sets $I_{n^*}(\theta^*)$ and $J_{n^*}(\theta^*)$. Hence,

$$\begin{aligned} V(x \cup \{\beta^*\} \cup \{\beta^{**}\}) - V(x \cup \{\beta^*\}) \\ = \sum_{i \in I_{n^*}(\theta^*)} (\beta^{**} - \beta^*) \cdot w_i - \sum_{j \in J_{n^*}(\theta^*)} (\beta^{**} - \beta^*) \cdot v_j \\ = \sum_{i \in I_{n^*}(\theta^*)} \hat{\varepsilon} f_{n^*} \cdot w_i - \sum_{j \in J_{n^*}(\theta^*)} \hat{\varepsilon} f_{n^*} \cdot v_j \end{aligned}$$

$$\begin{aligned}
&= \sum_{i \in I_{n^*}(\theta^*)} \hat{\epsilon} p_i f_{n^*} \cdot f_{n^*} - \sum_{j \in J_{n^*}(\theta^*)} \hat{\epsilon} q_j f_{n^*} \cdot f_{n^*} \\
&= \hat{\epsilon} (f_{n^*} \cdot f_{n^*}) \left(\sum_{i \in I_{n^*}(\theta^*)} p_i - \sum_{j \in J_{n^*}(\theta^*)} q_j \right) \\
&> 0
\end{aligned}$$

since $\hat{\epsilon}(f_{n^*} \cdot f_{n^*}) > 0$ and $\sum_{i \in I_{n^*}(\theta^*)} p_i - \sum_{j \in J_{n^*}(\theta^*)} q_j > 0$ by (10). This implies $x \cup \{\beta^*\} \cup \{\beta^{**}\} \succ x \cup \{\beta^*\}$, which contradicts exclusion since $\{\beta\} \succeq \{\beta^{**}\}$ for every $\beta \in x \cup \{\beta^*\}$. *Q.E.D.*

LEMMA 12: *For every n , there exist nonnegative scalars $\{c_{ij}\}_{i \in I_n, j \in J_n}$, $\{d_{ij}\}_{i \in I_n, j \in J_n}$, and $\{e_{ij}\}_{i \in I_n, j \in J_n}$, where $\sum_{j \in J_n} d_{ij} = 1$ for all $i \in I_n$ and $\sum_{i \in I_n} e_{ij} = 1$ for all $j \in J_n$, such that*

$$d_{ij}w_i = c_{ij}u + e_{ij}v_j \quad \forall i \in I_n, \quad \forall j \in J_n.$$

PROOF: The proof is by construction. Fix n . For every $i \in I_n$ and $j \in J_n$, define

$$\begin{aligned}
\hat{L}_{ij} &\equiv \min \left\{ \sum_{i' \in I_n : a_{i'}/p_{i'} \geq a_i/p_i} p_{i'}, \sum_{j' \in J_n : b_{j'}/q_{j'} \geq b_j/q_j} q_{j'} \right\} \\
&\quad - \max \left\{ \sum_{i' \in I_n : a_{i'}/p_{i'} > a_i/p_i} p_{i'}, \sum_{j' \in J_n : b_{j'}/q_{j'} > b_j/q_j} q_{j'} \right\}
\end{aligned}$$

and

$$L_{ij} \equiv \max\{0, \hat{L}_{ij}\}.$$

Hence, $L_{ij} \geq 0$ for every $i \in I_n$ and $j \in J_n$. One can show that for every $i \in I_n$, $\sum_{j \in J_n} L_{ij} = p_i$, and that for every $j \in J_n$, $\sum_{i \in I_n} L_{ij} = q_j$.

Now for every $i \in I_n$ and $j \in J_n$, define $d_{ij} \equiv L_{ij}/p_i$, $e_{ij} \equiv L_{ij}/q_j$, and $c_{ij} \equiv a_i d_{ij} - b_j e_{ij}$. (These are well defined since p_i and q_j are strictly positive.)

Fix $i \in I_n$ and $j \in J_n$. Observe that $d_{ij} \geq 0$, $e_{ij} \geq 0$, $\sum_{j' \in J_n} d_{ij'} = 1$, and $\sum_{i' \in I_n} e_{i'j} = 1$. Using (8) and (9), one can show that $d_{ij}w_i = c_{ij}u + e_{ij}v_j$.

Now we show that $c_{ij} \geq 0$. If $L_{ij} = 0$, then $c_{ij} = 0$. So suppose $L_{ij} > 0$. Observe that

$$c_{ij} = \left(\frac{a_i}{p_i} - \frac{b_j}{q_j} \right) L_{ij}.$$

Hence, $c_{ij} \geq 0$ if and only if $a_i/p_i \geq b_j/q_j$. Since $L_{ij} > 0$, it must be that

$$\sum_{j' \in J_n : b_{j'}/q_{j'} \geq b_j/q_j} q_{j'} > \sum_{i' \in I_n : a_{i'}/p_{i'} > a_i/p_i} p_{i'}.$$

But Lemma 11 implies

$$\sum_{i' \in I_n : a_{i'}/p_{i'} > b_j/q_j} p_{i'} \geq \sum_{j' \in J_n : b_{j'}/q_{j'} \geq b_j/q_j} q_{j'}.$$

Together, these imply

$$\sum_{i' \in I_n : a_{i'}/p_{i'} > b_j/q_j} p_{i'} > \sum_{i' \in I_n : a_{i'}/p_{i'} > a_i/p_i} p_{i'}$$

which can only be true if $a_i/p_i > b_j/q_j$. *Q.E.D.*

Now we finish the proof of sufficiency. Observe that we can write V as

$$\begin{aligned} V(x) = & \sum_{n=1}^N \left\{ \sum_{i \in I_n} \max_{\beta \in x} \beta \cdot w_i - \sum_{j \in J_n} \max_{\beta \in x} \beta \cdot v_j \right\} \\ & + \sum_{i \notin I'} \max_{\beta \in x} \beta \cdot w_i - \sum_{j \notin J'} \max_{\beta \in x} \beta \cdot v_j. \end{aligned}$$

Using Lemma 12, we get

$$\begin{aligned} V(x) = & \sum_{n=1}^N \left\{ \sum_{i \in I_n} \sum_{j \in J_n} \left(\max_{\beta \in x} \beta \cdot d_{ij} w_i - \max_{\beta \in x} \beta \cdot e_{ij} v_j \right) \right\} \\ & + \sum_{i \notin I'} \max_{\beta \in x} \beta \cdot w_i - \sum_{j \notin J'} \max_{\beta \in x} \beta \cdot v_j \\ = & \sum_{n=1}^N \left\{ \sum_{i \in I_n} \sum_{j \in J_n} \left(\max_{\beta \in x} [\beta \cdot c_{ij} u + \beta \cdot e_{ij} v_j] - \max_{\beta \in x} \beta \cdot e_{ij} v_j \right) \right\} \\ & + \sum_{i \notin I'} \max_{\beta \in x} \beta \cdot w_i - \sum_{j \notin J'} \max_{\beta \in x} \beta \cdot v_j. \end{aligned}$$

If $c_{ij} = 0$, then $d_{ij} = e_{ij} = 0$ since w_i and v_j are not redundant. So let M denote the number of (i, j) pairs in $I' \times J'$ such that $n(i) = n(j)$ and $c_{ij} \neq 0$. Let $m(i, j)$ denote a distinct element of $\{1, \dots, M\}$. Define $\hat{v}_{m(i,j)} \equiv (e_{ij}/c_{ij})v_j$ and

$\hat{c}_{m(i,j)} \equiv c_{ij}$. Hence we can write

$$\begin{aligned} V(x) &= \sum_{m=1}^M \hat{c}_m \left\{ \max_{\beta \in x} [\beta \cdot u + \beta \cdot \hat{v}_m] - \max_{\beta \in x} \beta \cdot \hat{v}_m \right\} \\ &\quad + \sum_{i \notin I'} \max_{\beta \in x} \beta \cdot w_i - \sum_{j \notin J'} \max_{\beta \in x} \beta \cdot v_j. \end{aligned}$$

For $i \notin I'$, recall that Lemma 6 and the fact that $w_i \neq \mathbf{0}$ imply that $w_i = a_i u$, where $a_i > 0$. So for $i \notin I'$, define $\tilde{v}_i \equiv \mathbf{0}$. Similarly, for $j \notin J'$, Lemma 6 and the fact that $v_j \neq \mathbf{0}$ imply that $v_j = -b_j u$, where $b_j > 0$. So for $j \notin J'$, define $\bar{v}_j \equiv -u$. Hence

$$\begin{aligned} (23) \quad V(x) &= \sum_{m=1}^M \hat{c}_m \left\{ \max_{\beta \in x} [\beta \cdot u + \beta \cdot \hat{v}_m] - \max_{\beta \in x} \beta \cdot \hat{v}_m \right\} \\ &\quad + \sum_{i \notin I'} a_i \left\{ \max_{\beta \in x} [\beta \cdot u + \beta \cdot \tilde{v}_i] - \max_{\beta \in x} \beta \cdot \tilde{v}_i \right\} \\ &\quad + \sum_{j \notin J'} b_j \left\{ \max_{\beta \in x} [\beta \cdot u + \beta \cdot \bar{v}_j] - \max_{\beta \in x} \beta \cdot \bar{v}_j \right\}. \end{aligned}$$

Finally, observe that

$$\begin{aligned} u &= \sum_{i \in I'} w_i - \sum_{j \in J'} v_j + \sum_{i \notin I'} w_i - \sum_{j \notin J'} v_j \\ &= \sum_{n=1}^N \left\{ \sum_{i \in I_n} \sum_{j \in J_n} (d_{ij} w_i - e_{ij} v_j) \right\} + \sum_{i \notin I'} w_i - \sum_{j \notin J'} v_j \\ &= \sum_{n=1}^N \left\{ \sum_{i \in I_n} \sum_{j \in J_n} c_{ij} u \right\} + \sum_{i \notin I'} a_i u + \sum_{j \notin J'} b_j u \\ &= \left(\sum_{m=1}^M \hat{c}_m + \sum_{i \notin I'} a_i + \sum_{j \notin J'} b_j \right) u, \end{aligned}$$

which implies $\sum_{m=1}^M \hat{c}_m + \sum_{i \notin I'} a_i + \sum_{j \notin J'} b_j = 1$. So $\{\hat{c}_1, \dots, \hat{c}_m\} \cup \{a_i\}_{i \notin I'} \cup \{b_j\}_{j \notin J'}$ is a set of positive scalars that sum to 1. Hence (23) is a temptation representation.

B.2. Theorem 1: Necessity of Axioms

We show only the necessity of exclusion; the proof for inclusion is similar. The necessity of the other axioms is immediate from DLR Theorem 6.

Suppose \succeq has a temptation representation of the form in (1). Let $x \in X$ and $\alpha \in \Delta$ be such that $\{\beta\} \succeq \{\alpha\}$ for every $\beta \in x$. Then $\min_{\beta \in x} u(\beta) \geq u(\alpha)$. We will show that for every i ,

$$\begin{aligned} & \max_{\beta \in x} [u(\beta) + v_i(\beta)] - \max_{\beta \in x} v_i(\beta) \\ & \geq \max_{\beta \in x \cup \{\alpha\}} [u(\beta) + v_i(\beta)] - \max_{\beta \in x \cup \{\alpha\}} v_i(\beta), \end{aligned}$$

thus proving that $V(x) \geq V(x \cup \{\alpha\})$. Fix i .

Case 1— $u(\alpha) + v_i(\alpha) \geq \max_{\beta \in x} [u(\beta) + v_i(\beta)]$. Since $\min_{\beta \in x} u(\beta) \geq u(\alpha)$, it must be that $v_i(\alpha) \geq \max_{\beta \in x} v_i(\beta)$. Hence

$$\begin{aligned} & \max_{\beta \in x \cup \{\alpha\}} [u(\beta) + v_i(\beta)] - \max_{\beta \in x \cup \{\alpha\}} v_i(\beta) \\ & = u(\alpha) \leq \min_{\beta \in x} u(\beta) \leq \max_{\beta \in x} [u(\beta) + v_i(\beta)] - \max_{\beta \in x} v_i(\beta), \end{aligned}$$

where the second inequality can easily be verified by making the substitution $w_i \equiv -u - v_i$.

Case 2— $\max_{\beta \in x} [u(\beta) + v_i(\beta)] > u(\alpha) + v_i(\alpha)$. Then

$$\begin{aligned} & \max_{\beta \in x \cup \{\alpha\}} [u(\beta) + v_i(\beta)] - \max_{\beta \in x \cup \{\alpha\}} v_i(\beta) \\ & = \max_{\beta \in x} [u(\beta) + v_i(\beta)] - \max_{\beta \in x \cup \{\alpha\}} v_i(\beta) \\ & \leq \max_{\beta \in x} [u(\beta) + v_i(\beta)] - \max_{\beta \in x} v_i(\beta). \end{aligned} \quad Q.E.D.$$

APPENDIX C: COUNTEREXAMPLE

Here we provide an example of preferences that satisfy DFC and desire for better alternatives (and all our other axioms) but not exclusion or inclusion. Since exclusion and inclusion are necessary, this proves that such preferences do not have a temptation representation. We thank an anonymous referee for providing this example.

First we state formally the axiom desire for better alternatives.

AXIOM 11—Desire for Better Alternatives: *For every x , there exists $\alpha \in x$ such that $x \succeq \{\alpha\}$.*

Fix any vectors $w_1, f \in H^1 \setminus \{\mathbf{0}\}$ such that $f \perp w_1$. Set $v_1 \equiv -w_1$, $v_2 \equiv f + aw_1$, and $w_2 \equiv f - aw_1$ for $0 < a < 1/2$. Let preferences be represented by

$$V(x) = \max_{\beta \in x} w_1(\beta) + \max_{\beta \in x} w_2(\beta) - \max_{\beta \in x} v_1(\beta) - \max_{\beta \in x} v_2(\beta),$$

which is a finite additive EU representation, and hence satisfies weak order, continuity, independence, and finiteness. Set

$$u \equiv w_1 + w_2 - v_1 - v_2 = (2 - 2a)w_1.$$

Then we can rewrite V as

$$\begin{aligned} V(x) &= \frac{1}{2 - 2a} \max_{\beta \in x} u(\beta) + \frac{1 - 2a}{2 - 2a} \left\{ \max_{\beta \in x} [u(\beta) + \hat{v}_1(\beta) + \hat{v}_2(\beta)] \right. \\ &\quad \left. - \max_{\beta \in x} \hat{v}_1(\beta) - \max_{\beta \in x} \hat{v}_2(\beta) \right\}, \end{aligned}$$

where $\hat{v}_1 = ((2 - 2a)/(1 - 2a))v_1$ and $\hat{v}_2 = ((2 - 2a)/(1 - 2a))v_2$, which is a DLR temptation representation. Hence, this preference satisfies DFC. Using the symmetry of the representation, one could similarly show that $-V$ can also be written as a DLR temptation representation. Hence the preference represented by $-V$ satisfies DFC. But this implies that the preference represented by V satisfies desire for better alternatives. (This preference also satisfies DLR's other axiom—approximate improvements are chosen—as well as an axiom symmetric to this, so even adding these axioms would be insufficient to get a temptation representation.)

Now we show that this preference violates exclusion. (The following is similar to the method used to prove Lemma 11 in the proof of sufficiency for Theorem 1.) Fix any $\beta \in \Delta^\circ$. Let $\varepsilon > 0$ be such that $\alpha \equiv \beta + \varepsilon u \in \Delta$. Observe then that $\alpha \cdot w_1 > \beta \cdot w_1$, $\beta \cdot w_2 > \alpha \cdot w_2$, $\beta \cdot v_1 > \alpha \cdot v_1$, and $\alpha \cdot v_2 > \beta \cdot v_2$. Let $\varepsilon' > 0$ be such that $\beta' \equiv \beta + \varepsilon' f \in \Delta$ and $\alpha \cdot v_2 > \beta' \cdot v_2$. Then $V(\{\alpha, \beta, \beta'\}) > V(\{\alpha, \beta\})$ since w_2 is the only vector whose maximum changes when β' is added to $\{\alpha, \beta\}$. But this violates exclusion since $\beta' \cdot u = \beta \cdot u < \alpha \cdot u$.

A violation of inclusion could be constructed in a similar manner.

REFERENCES

- CHANDRASEKHER, M. (2009): “A Theory of Local Menu Preferences,” Working Paper, Arizona State University. [351]
- DEKEL, E., AND B. LIPMAN (2007): “Self-Control and Random Strotz Representations,” Working Paper, Boston University. [352,353]
- DEKEL, E., B. LIPMAN, AND A. RUSTICHINI (2001): “Representing Preferences With a Unique Subjective State Space,” *Econometrica*, 69, 891–934. [351,360]
- (2009): “Temptation-Driven Preferences,” *Review of Economic Studies*, 76, 937–971. [350,355,360]
- DEKEL, E., B. LIPMAN, A. RUSTICHINI, AND T. SARVER (2007): “Representing Preferences With a Unique Subjective State Space: Corrigendum,” *Econometrica*, 75, 591–600. [352]
- GUL, F., AND W. PESENDORFER (2001): “Temptation and Self-Control,” *Econometrica*, 69, 1403–1435. [350,352,354]
- KOPYLOV, I. (2009): “Perfectionism and Choice,” Working Paper, UC Irvine. [356]
- NEHRING, K. (2006): “Self-Control Through Second-Order Preferences,” Working Paper, UC Davis. [351]

- NOOR, J. (2007): "Commitment and Self-Control," *Journal of Economic Theory*, 135, 1–34. [350]
SARVER, T. (2008): "Anticipating Regret: Why Fewer Options May Be Better," *Econometrica*, 76, 263–305. [356]

Dept. of Economics, Harkness Hall, University of Rochester, Rochester, NY 14627, U.S.A.; jstovall@mail.rochester.edu.

Manuscript received August, 2008; final revision received September, 2009.

EVALUATING MARGINAL POLICY CHANGES AND THE AVERAGE EFFECT OF TREATMENT FOR INDIVIDUALS AT THE MARGIN

BY PEDRO CARNEIRO, JAMES J. HECKMAN, AND EDWARD VYTLACIL¹

This paper develops methods for evaluating marginal policy changes. We characterize how the effects of marginal policy changes depend on the direction of the policy change, and show that marginal policy effects are fundamentally easier to identify and to estimate than conventional treatment parameters. We develop the connection between marginal policy effects and the average effect of treatment for persons on the margin of indifference between participation in treatment and nonparticipation, and use this connection to analyze both parameters. We apply our analysis to estimate the effect of marginal changes in tuition on the return to going to college.

KEYWORDS: Marginal treatment effect, effects of marginal policy changes, marginal policy relevant treatment effect, average marginal treatment effect.

1. INTRODUCTION

THE POLICY RELEVANT TREATMENT EFFECT (PRTE) is the mean effect of changing from a baseline policy to an alternative policy that provides different incentives to participate in treatment (Heckman and Vytlacil (2001b, 2005)). Identification and estimation of the PRTE are generally difficult tasks. Identification of the PRTE typically requires large support conditions. \sqrt{N} -consistent estimation of the PRTE parameter is generally not possible.

In many cases, proposed policy reforms are incremental in nature and a marginal version of the PRTE (MPRTE) is all that is required to answer questions of economic interest. This paper develops the MPRTE and establishes how the MPRTE depends on the direction of a proposed marginal policy change. We establish that the support conditions required for identifying the MPRTE are very weak. The essential requirement is availability of a continuous instrument. The MPRTE parameter can be represented as a weighted average derivative with weights determined by the marginal policy of interest. The parameter is \sqrt{N} -estimable under standard regularity conditions. Thus, the MPRTE parameter is fundamentally easier to identify and estimate than the PRTE parameter.

We connect the MPRTE to the average marginal treatment effect (AMTE): the mean benefit of treatment for people at the margin of indifference between

¹This research was supported by NIH R01-HD32058-3, NSF SES-0832845, NSF SES-024158, NSF SES-05-51089, ESRC RES-000-22-2542, the Geary Institute at University College Dublin, the American Bar Foundation, the Pritzker Consortium on Early Childhood Development, the Leverhulme Trust, and ESRC (RES-589-28-0001) through the funding of the Centre for Microdata Methods and Practice. The research was conducted in part while Edward Vytlacil was a Visiting Professor at Hitotsubashi University. We thank the editor, two anonymous referees, Hidehiko Ichimura, Richard Robb, Daniel Schmierer, and Azeem Shaikh for very helpful comments. We would like to thank Erica Blom and Sukjin Han for research assistance.

participation in treatment and nonparticipation. AMTE is compared to marginal cost in an aggregate benefit–cost analysis of a program. We establish a correspondence between MPRTE parameters and AMTE parameters, showing that the effect of a marginal policy change in a particular direction is the same as the average effect of treatment for those at the margin of indifference. We use this correspondence to produce new insights about the AMTE parameter.

The paper proceeds as follows. Section 2 presents the nonparametric selection model that underlies our analysis. Section 3 reviews the analysis of the PRTE by Heckman and Vytlacil (2005) and presents a new interpretation of the PRTE as a function from the space of all possible policies to the space of effects of policies. This interpretation is key to the analysis in Section 4, which introduces and analyzes the MPRTE. We discuss identification and estimation of the MPRTE in Sections 5 and 6, respectively. We define and analyze the AMTE in Section 7. Section 8 presents an empirical application of our analysis. Section 9 concludes.

2. NONPARAMETRIC SELECTION MODEL AND OUR ASSUMPTIONS

Assume that there are two potential outcomes (Y_0, Y_1) and a binary treatment choice indicator D . Outcome Y is written in switching regression form $Y = DY_1 + (1 - D)Y_0$, where Y_1 is the potential outcome that is observed if the agent chooses treatment 1 and Y_0 is the potential outcome that is observed if the agent chooses treatment 0. $Y_1 - Y_0$ is the individual level treatment effect. We keep implicit the conditioning on observed variables X that determine Y_1 or Y_0 , and maintain the assumption that D does not determine X (see Heckman and Vytlacil (2005)).

Program participation is voluntary. To link this framework to standard choice models, we characterize the decision rule for program participation by a latent index model:

$$(2.1) \quad D = \mathbf{1}[\mu(Z) - V \geq 0],$$

where $\mathbf{1}[\cdot]$ is the indicator function taking the value 1 if its argument is true and the value 0 otherwise. From the point of view of the econometrician, Z is observed and V is unobserved.

We maintain the following assumptions:

ASSUMPTION A-1: (Y_0, Y_1, V) is independent of Z .

ASSUMPTION A-2: The distribution of V is absolutely continuous with respect to Lebesgue measure.

ASSUMPTION A-3: The distribution of $\mu_D(Z)$ is absolutely continuous with respect to Lebesgue measure.

ASSUMPTION A-4: $\sup_v E(|Y_1||V = v) < \infty$ and $\sup_v E(|Y_0||V = v) < \infty$.

ASSUMPTION A-5: $0 < \Pr(D = 1) < 1$.

These conditions are discussed in Heckman and Vytlacil (2005, 2007). Under them, Vytlacil (2002) establishes the equivalence between the nonparametric latent index model (2.1) and the monotonicity assumption used by Imbens and Angrist (1994). A necessary condition for A-3 is that Z contains a continuous variable (i.e., that there is a continuous instrument for D). We use this assumption to analyze marginal policy changes and to identify marginal policy treatment effects. These conditions should be interpreted as conditional on X , which we have kept implicit. For example, A-1 should be interpreted as saying that (Y_0, Y_1, V) is independent of Z conditional on X , while there can be dependence between (Y_0, Y_1, V) and X .

Define $P(Z)$ as the probability of receiving treatment given Z : $P(Z) \equiv \Pr(D = 1|Z) = F_V(\mu(Z))$, where $F_V(\cdot)$ is the distribution function of V . We sometimes denote $P(Z)$ by P , suppressing the Z argument. We also use U , defined by $U = F_V(V)$ so that U is distributed unit uniform. In this notation, applying a monotonic transformation to both sides of the argument in equation (2.1) allows us to write that equation as $D = \mathbf{1}[P(Z) \geq U]$.

The marginal treatment effect (MTE) plays a fundamental role in our analysis. MTE is defined as $MTE(u) \equiv E(Y_1 - Y_0|U = u)$, that is, the expected treatment effect conditional on the unobservables that determine participation. For values of u close to zero, $MTE(u)$ is the expected effect of treatment on individuals who have unobservables that make them most likely to participate in treatment and who would participate even if the mean scale utility $\mu(Z)$ were small. See Heckman and Vytlacil (2005, 2007) for more discussion and interpretation of the MTE.

In some of our examples, we consider the following special case of our general model:

ASSUMPTION B-1: Suppose $\mu(Z) = Z\gamma$ and suppose that F_V is strictly increasing. Suppose that the k th component of Z , $Z^{[k]}$, has a strictly positive coefficient, $\gamma^{[k]} > 0$. Let $\tilde{Z}\tilde{\gamma} = Z\gamma - Z^{[k]}\gamma^{[k]}$. Suppose that the distribution of $Z^{[k]}\gamma^{[k]}$ conditional on $\tilde{Z}\tilde{\gamma}$ has a density with respect to Lebesgue measure.

3. THE POLICY RELEVANT TREATMENT EFFECT

We first present the PRTE of Heckman and Vytlacil (2001b, 2005) to motivate our marginal version of it. We then reformulate the PRTE as a function of the proposed policy change. This enables us to define a sequence of PRTEs corresponding to a sequence of proposed policy changes.

3.1. Review of PRTE

Following Heckman and Vytlacil (2001b, 2005), consider a class of policies that affect P —the probability of participation in a program—but that do not affect potential outcomes or unobservables related to the selection process, (Y_1, Y_0, U) .² An example from the literature on the economic returns to schooling would be policies that change tuition or distance to school but that do not directly affect potential wages (Card (2001)). We ignore general equilibrium effects.

Let D^* be the treatment choice that would be made after the policy change. Let P^* be the corresponding probability that $D^* = 1$ after the policy change. D^* is defined by $D^* = \mathbf{1}[P^* \geq U]$. Let $Y^* = D^* Y_1 + (1 - D^*) Y_0$ be the outcome under the alternative policy. Following Heckman and Vytlacil (2005), the mean effect of going from a baseline policy to an alternative policy per net person shifted is the PRTE, defined when $E(D) \neq E(D^*)$ as

$$(3.1) \quad \begin{aligned} & \frac{E(Y|\text{alternative policy}) - E(Y|\text{baseline policy})}{E(D|\text{alternative policy}) - E(D|\text{baseline policy})} \\ &= \frac{E(Y^*) - E(Y)}{E(D^*) - E(D)} = \int_0^1 \text{MTE}(u) \omega_{\text{PRTE}}(u) du, \end{aligned}$$

where

$$(3.2) \quad \omega_{\text{PRTE}}(u) = \frac{F_P(u) - F_{P^*}(u)}{E_{F_{P^*}}(P) - E_{F_P}(P)}.$$

The condition $E(D) \neq E(D^*)$ is consistent with the program having a non-monotonic effect on participation as long as the fraction switching into treatment is not exactly offset by the fraction switching out of treatment. The PRTE parameter gives the normalized effect of a change from a baseline policy to an alternative policy and depends on the alternative being considered.³ Heckman and Vytlacil (2005) show that the PRTE can be identified under strong support conditions. In Section 5, we establish that the marginal version of the PRTE parameter can be identified under much weaker conditions than are required to identify the PRTE. In Section 6, we establish that, unlike the PRTE parameter, the MPRTE parameter is generally estimable at a \sqrt{N} rate.

We define PRTE as the average effect per net person shifted into treatment. With this definition, no normalization is required when taking limits to define

²This restriction can be relaxed to a weaker policy invariance for the distribution of (Y_1, Y_0, U) ; see Heckman and Vytlacil (2005, 2007).

³The PRTE can be interpreted as an economically more explicit version of Stock's (1989) nonparametric policy analysis parameter for a class of policy interventions with explicit agent preferences where the policies evaluated operate solely on agent choice sets.

the MPRTE, and we obtain an equivalence between the MPRTE and the average effect for individuals at the margin of indifference (AMTE). Under this definition, the MPRTE corresponds to a weighted average derivative and can be estimated by a weighted average derivative estimator.

Alternatively, we could follow Heckman and Vytlacil (2001b) in defining PRTE without normalizing it by the net change in treatment status, that is, we could define PRTE as $E(Y^*) - E(Y)$. However, some normalization is required when taking limits. If we were to define PRTE in that manner and then normalize by $E(D^*) - E(D)$ when taking limits, we would obtain the same limits as would be obtained from analyzing the marginal version of PRTE defined by equation (3.1). Using the unnormalized version of PRTE, we can analyze alternative normalizations when taking limits. In footnote 5 below, we discuss one such normalization which results in an alternative version of MPRTE that is a rescaled version of the marginal version of PRTE defined by equation (3.1).

3.2. PRTE as a Function of Proposed Policy Changes

The PRTE depends on a policy change only through the distribution of P^* after the policy change. Given our assumptions, F_{P^*} is sufficient to summarize everything about the proposed policy change that is relevant for calculating the average effect of the policy change. We can thus define the PRTE function as a function mapping the proposed policy change (corresponding to a distribution of P^*) to the resulting per-person change in outcomes. Expressing the PRTE this way is important in the next step of our analysis that uses sequences of PRTEs to define a marginal version of the PRTE.

Let \mathcal{G} denote the space of all cumulative distribution functions for random variables that lie in the unit interval such that $\int t dG(t) \neq \int t dF_P(t)$, that is, all right-continuous, nondecreasing functions that satisfy $G(t) = 1$ for $t \geq 1$ and $G(t) = 0$ for $t < 0$, and such that the first moment of G does not equal the first moment of F_P . Any $G \in \mathcal{G}$ corresponds to a potential distribution of P^* and thus corresponds to a potential alternative policy, with $\text{PRTE}(G)$ denoting the corresponding policy effect. We define the PRTE function, $\text{PRTE}: \mathcal{G} \mapsto \mathbb{R}$, by

$$(3.3) \quad \text{PRTE}(G) = \int_0^1 \text{MTE}(u) \omega_{\text{PRTE}}(u; G) du,$$

where

$$(3.4) \quad \omega_{\text{PRTE}}(u, G) = \frac{F_P(u) - G(u)}{E_G(P) - E_{F_P}(P)}.$$

In many cases, the class of policy alternatives under consideration can be indexed by a scalar variable. Let $P_0 = P$ denote the baseline probability for $D = 1$. Let \mathbf{M} denote a subset of \mathbb{R} with $0 \in \mathbf{M}$. Let $\{P_\alpha : \alpha \in \mathbf{M}\}$ denote a class of alternative probabilities corresponding to alternative policy regimes with

associated cumulative distribution functions F_α . For example, one alternative policy could increase the probability of attending college by an amount α , so that $P_\alpha = P_0 + \alpha$ and $F_\alpha(t) = F_0(t - \alpha)$. An alternative policy could change each person's probability of attending college by the proportion $(1 + \alpha)$, so that $P_\alpha = (1 + \alpha)P_0$ and $F_\alpha(t) = F_0(\frac{t}{1+\alpha})$. The policy intervention might have an effect similar to a shift in one of the components of Z , say $Z^{[k]}$. In particular, suppose $Z_\alpha^{[k]} = Z^{[k]} + \alpha$ and $Z_\alpha^{[j]} = Z^{[j]}$ for $j \neq k$. For example, the k th element of Z might be college tuition, and the policy under consideration subsidizes college tuition by the fixed amount α . Suppose that the linear latent index Assumption B-1 holds. Then $P_\alpha(Z) = F_V(Z\gamma + \alpha\gamma^{[k]})$ and $F_\alpha(t) = F_{Z\gamma}(F_V^{-1}(t) - \alpha\gamma^{[k]})$. Notice that the first two examples have the form $P_\alpha = q_\alpha(P_0)$ for some function q_α , while the last example has the form $P_\alpha = F_V(Z^{[-k]}\gamma^{[-k]} + q_\alpha(Z^{[k]})\gamma^{[k]})$, where $Z^{[-k]}$ denotes the elements of Z not in $Z^{[k]}$ and $\gamma^{[-k]}$ is the corresponding coefficient vector. We will explore more general examples in the next section.

4. MARGINAL POLICY CHANGES

The PRTE is defined for a discrete change from a baseline policy to a fixed alternative. We now consider a marginal version of the PRTE parameter that corresponds to a marginal change from a baseline policy.⁴ It is expositorily convenient to think of the treatment as college attendance and the policy as a change in tuition. The marginal version of the PRTE depends on the nature of the perturbation that defines the marginal change. For example, a policy change that subsidizes tuition by a fixed amount has different effects than a policy change that subsidizes tuition by a fixed proportion. The limits of these two policies for infinitesimally small subsidies are different.

To define the marginal version of the PRTE, we could consider the limit of $\Delta^{\text{PRTE}}(G)$ as G gets close to F_P in some metric. We could define the marginal PRTE as a directional derivative. For ease of exposition, we do not work with this more general formulation, but instead work with a one-dimensional version of it. Thus we do not analyze general perturbations within the function space \mathcal{G} , but only one-dimensional curves within \mathcal{G} .

Consider a class of alternative policies indexed by α , $\{F_\alpha : \alpha \in \mathbf{M}\}$, where 0 is a limit point of \mathbf{M} and 0 represents the baseline policy so that $F_0 = F_P$. Consider the effect of a marginal change in α in a neighborhood of the current base policy of $\alpha = 0$. When the limit exists, we can define the MPRTE as

$$\text{MPRTE}(\{F_\alpha\}) = \lim_{\tau \rightarrow 0} \text{PRTE}(F_\tau),$$

where the MPRTE parameter depends on the class of alternative policies $\{F_\alpha : \alpha \in \mathbf{M}\}$, that is, on the choice of a particular curve within \mathcal{G} . Thus, the

⁴Ichimura and Taber (2002) present a discussion of local policy analysis in a model without the MTE structure using a framework developed by Hurwicz (1962).

MPRTE can be seen as a path derivative along the path $\{F_\alpha : \alpha \in \mathbf{M}\}$. We impose the following sufficient conditions for the limit to exist:

ASSUMPTION A-6: *For α in a neighborhood of 0, F_α has a density f_α with respect to Lebesgue measure, f_α is differentiable in α , $\sup_{t \in (0,1)} |\frac{\partial}{\partial \alpha} f_\alpha(t)| < \infty$, and $\int_0^1 t \frac{\partial}{\partial \alpha} f_\alpha(t) dt \neq 0$.*

Under Assumptions **A-1–A-6**, $\text{MPRTE}(\{F_\alpha\}) = \lim_{\tau \rightarrow 0} \text{PRTE}(F_\tau)$ exists and is given by

$$(4.1) \quad \text{MPRTE}(\{F_\alpha\}) = \int_0^1 \text{MTE}(u) \omega_{\text{MPRTE}}(u; \{F_\alpha\}) du,$$

where $\omega_{\text{MPRTE}}(u; \{F_\alpha\})$ is

$$(4.2) \quad \omega_{\text{MPRTE}}(u; \{F_\alpha\}) = -\frac{\frac{\partial}{\partial \alpha} F_0(u)}{\frac{\partial}{\partial \alpha} E_{F_0}(P)} = -\frac{\int_0^u \left(\frac{\partial}{\partial \alpha} f_0(p) \right) dp}{\int_0^1 p \left(\frac{\partial}{\partial \alpha} f_0(p) \right) dp},$$

where we write $\frac{\partial}{\partial \alpha} F_0(p)$ and $\frac{\partial}{\partial \alpha} f_0(p)$ as shorthand expressions for $\frac{\partial}{\partial \alpha} F_\alpha(p)|_{\alpha=0}$ and $\frac{\partial}{\partial \alpha} f_\alpha(p)|_{\alpha=0}$, respectively. An alternative way to express the form of the weights uses the property that $0 \leq P_\alpha \leq 1$ so that $E(P_\alpha) = \int_0^1 (1 - F_\alpha(t)) dt$ to obtain

$$(4.3) \quad \omega_{\text{MPRTE}}(u; \{F_\alpha\}) = \frac{\frac{\partial}{\partial \alpha} F_0(u)}{\int_0^1 \left(\frac{\partial}{\partial \alpha} F_0(t) \right) dt},$$

which makes it clear that the weights always integrate to unity.⁵

Just as the PRTE parameter depends on which particular policy counterfactual is being considered, the marginal PRTE parameter depends on which particular class of policy perturbations is being considered. Just as the effect of a fixed dollar amount tuition subsidy will result in a different PRTE parameter than a proportional tuition subsidy, the limit parameter for a marginal change in an additive tuition subsidy will be different from the limit parameter for a marginal change in a proportional tuition subsidy. Of particular interest to us are the following two cases for which we will define classes of functions. Let

⁵An alternative way to define MPRTE that does not require the condition $E(D) \neq E(D^*)$ normalizes by α instead of $E(D) - E(D^*)$: $\lim_{\alpha \rightarrow 0} \int_0^1 \text{MTE}(u)([F_\alpha(u) - F_0(u)]/\alpha) du$. With this definition the corresponding weights on MTE are given by $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = \frac{\partial}{\partial \alpha} F_0(u)$. This change in the normalization only affects the constant of integration for the weights and results in an alternative MPRTE that is a rescaled version of the MPRTE analyzed in this paper.

\mathcal{Q} denote the set of sequences of functions $\{q_\alpha(\cdot) : \alpha \in \mathbf{M}\}$ such that $q_\alpha(t)$ is strictly increasing in t for any $\alpha \in \mathbf{M}$; such that $q_0(\cdot)$ is the identity function; such that $q_\alpha(t)$ is differentiable in α with $\frac{\partial}{\partial \alpha} q_\alpha(t)$ bounded; such that r_α is a bounded function where r_α denotes the inverse of q_α ; and such that r_α is differentiable in α with $\frac{\partial}{\partial \alpha} r_\alpha(t)$ bounded. The following examples can be trivially modified to allow q_α to be monotonically decreasing.

EXAMPLE 1: Suppose that the alternative sequence of policies has the form $P_\alpha = q_\alpha(P_0)$ for some $\{q_\alpha\} \in \mathcal{Q}$. Then $F_\alpha(t) = F_P(r_\alpha(t))$ and

$$(4.4) \quad \omega_{\text{MPRTE}}(u; \{F_\alpha\}) = -\frac{f_P(u) \frac{\partial}{\partial \alpha} r_0(u)}{E_P\left(\frac{\partial}{\partial \alpha} q_0(P)\right)} = \frac{f_P(u) \frac{\partial}{\partial \alpha} r_0(u)}{\int_0^1 f_P(t) \frac{\partial}{\partial \alpha} r_0(t) dt}.$$

EXAMPLE 2: Suppose that the alternative sequence of policies shifts the k th component of Z , $Z^{[k]}$, to $q_\alpha(Z^{[k]})$ for some $\{q_\alpha\} \in \mathcal{Q}$. Suppose that Assumption B-1 holds. Then

$$F_\alpha(t) = \int F_{Z^{[k]}|\tilde{Z}\tilde{\gamma}}\left[r_\alpha\left(\frac{F_V^{-1}(t) - s}{\gamma^{[k]}}\right)\right] f_{\tilde{Z}\tilde{\gamma}}(s) ds$$

and

$$(4.5) \quad \begin{aligned} \omega_{\text{MPRTE}}(u; \{F_\alpha\}) &= \frac{E_{\tilde{Z}\tilde{\gamma}}\left[f_{Z^{[k]}|\tilde{Z}\tilde{\gamma}}\left[\frac{F_V^{-1}(u) - \tilde{Z}\tilde{\gamma}}{\gamma^{[k]}}\right] \frac{\partial}{\partial \alpha} r_0\left(\frac{F_V^{-1}(u) - \tilde{Z}\tilde{\gamma}}{\gamma^{[k]}}\right)\right]}{\int_0^1 E_{\tilde{Z}\tilde{\gamma}}\left[f_{Z^{[k]}|\tilde{Z}\tilde{\gamma}}\left[\frac{F_V^{-1}(t) - \tilde{Z}\tilde{\gamma}}{\gamma^{[k]}}\right] \frac{\partial}{\partial \alpha} r_0\left(\frac{F_V^{-1}(t) - \tilde{Z}\tilde{\gamma}}{\gamma^{[k]}}\right)\right] dt}. \end{aligned}$$

The expressions in equations (4.4) and (4.5) look different, but both can be represented as weighted densities

$$(4.6) \quad \omega_{\text{MPRTE}}(u; \{F_\alpha\}) = f_P(u)h(u)/E(h(P))$$

for some function $h(\cdot)$. For Example 1, this form is immediate, substituting $h(u) = \frac{\partial}{\partial \alpha} r_0(u)$. For Example 2, note that, for $\omega_{\text{MPRTE}}(u; \{F_\alpha\})$ given by equation (4.5),

$$\begin{aligned} \omega_{\text{MPRTE}}(u; \{F_\alpha\}) \neq 0 &\Rightarrow \int f_{Z^{[k]}|\tilde{Z}\tilde{\gamma}}\left[\frac{F_V^{-1}(u) - s}{\gamma^{[k]}}\right] f_{\tilde{Z}\tilde{\gamma}}(s) ds \neq 0 \\ &\Leftrightarrow f_{Z\gamma}(F_V^{-1}(u)) \neq 0 \\ &\Leftrightarrow f_P(u)f_V(F_V^{-1}(u)) \neq 0 \\ &\Rightarrow f_P(u) \neq 0, \end{aligned}$$

where we use the fact that $\Pr[P(Z) \leq u] = \Pr[Z\gamma \leq F_V^{-1}(u)]$ and the chain rule to obtain $f_P(u) = (f_{Z\gamma}(F_V^{-1}(u)))/(f_V(F_V^{-1}(u)))$. Thus, the weights for Example 2 are of the form $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = f_P(u)h(u)/C$ for some function h and some constant C . The weights integrate to 1, because $C = \int f_P(u)h(u) du = E(h(P))$. Thus the weights for both Example 1 and Example 2 are of the form (4.6). This expression plays an important role when we consider identification and estimation of the MPRTE.

Special cases include $q_\alpha(t) = t + \alpha$ and $q_\alpha(t) = (1 + \alpha)t$ for policy changes that act like constant shifts or proportional shifts either in P or in a component of Z . For Example 1, if $q_\alpha(t) = t + \alpha$, then $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = f_P(u)$, so that the MPRTE weights MTE according to the density of $P(Z)$. In contrast, if $q_\alpha(t) = (1 + \alpha)t$, then $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = uf_P(u)/E(P)$. Thus, for example, the limit form associated with increasing the probability of college attendance by a fixed amount and the limit form associated with increasing the probability of college attendance by a proportional amount produce different weights on MTE. The limit of the latter puts higher weight on MTE for higher u evaluation points, that is, puts higher weight on MTE for individuals whose unobservables make them less likely to go to college. For Example 2, setting $q_\alpha(t) = t + \alpha$ results in

$$\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = \frac{f_{Z\gamma}(F_V^{-1}(u))}{E(f_V(Z\gamma))} = \frac{f_P(u)f_V(F_V^{-1}(u))}{E(f_V(Z\gamma))},$$

which again will weight the MTE and thus weight people with different unobserved preferences for treatment differently from the way MTE is weighted in the other examples.

5. IDENTIFICATION

Equations (3.3)–(3.4) and (4.1)–(4.2) show that both the PRTE and the MPRTE treatment parameters can be expressed in the form

$$(5.1) \quad \text{treatment parameter}(j) = \int_0^1 \text{MTE}(u)\omega_j(u) du,$$

where $\omega_j(u)$ is the weighting function corresponding to treatment parameter j and where ω_j will depend on which policy change/marginal policy change is being considered. Heckman and Vytlacil (1999, 2005) show that, under our assumptions, the standard treatment parameters can also be expressed in this form and that the form can be used for identification. In particular, the weights are often easy to identify, in which case equation (5.1) implies identification of the parameter if we can identify the MTE at any evaluation point for which the weights are nonzero. Under our assumptions, Heckman and Vytlacil (1999,

2005) show that the MTE can be identified by the method of local instrumental variables (LIV)⁶ at any p in the support of $P(Z)$:

$$(5.2) \quad \frac{\partial}{\partial p} E(Y|P(Z) = p) = \text{MTE}(p).$$

Any parameter that can be represented as a weighted average of MTE can be identified by first estimating MTE over the appropriate support and then integrating the identified MTE function using the appropriate weights. To identify parameter j , not only does $P(Z)$ have to be a continuous random variable, but, in addition, the support of $P(Z)$ must contain all values of u such that $\omega_j(u) \neq 0$. For the standard treatment parameters, identification requires strong conditions on the support of the distribution of $P(Z)$. For example, for the average treatment effect, $E(Y_1 - Y_0)$, the weights are given by $\omega_{ATE}(u) = 1$ for $u \in [0, 1]$ so that the required support condition is that the support of $P(Z)$ is the full unit interval. It is possible to point-identify the parameters under only slightly weaker conditions than those required by this strategy, as shown in Heckman and Vytlacil (2001a, 2007).

Consider identification of PRTE. Suppose that we can identify F_{P^*} and thus the weights. To identify the parameter using the strategy just discussed, the support of $P(Z)$ must contain all values of u such that $F_P(u) - F_{P^*}(u) \neq 0$. Thus, if the support of P^* is not contained in the support of P , it is not possible to identify the PRTE.

For example, suppose that the largest estimated probability of attending college is strictly less than 1. For analyzing a tuition subsidy policy, it is possible that the largest probability of attending college under a tuition subsidy will be greater than the largest probability of attending college without a tuition subsidy, so the support condition for identifying the corresponding PRTE parameter is violated. More generally, in Examples 1 and 2 in Section 4, the PRTE parameters will not be identified unless the support of P is the full unit interval. Thus, the very strong support conditions required for identification of the standard treatment parameters are also required for identification of the PRTE parameter.

In contrast, the MPRTE parameter can generally be identified under weaker assumptions. The MPRTE weights are nonzero only if the density of $P(Z)$ is nonzero. Consider the classes of MPRTE parameters produced from Examples 1 and 2. For policy counterfactuals that act like transformations of $P(Z)$ as in Example 1, or act like transformations of Z as in Example 2, we have from

⁶LIV can be interpreted as the limit form of the Imbens and Angrist (1994) local average treatment effect (LATE) parameter (see Heckman and Vytlacil (1999)). The ideas of the marginal treatment effect and the limit form of LATE were first introduced in the context of a parametric normal generalized Roy model by Björklund and Moffitt (1987), and were analyzed more generally in Heckman (1997). Angrist, Graddy, and Imbens (2000) also defined and developed a limit form of LATE.

equation (4.6) that $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) \neq 0 \Rightarrow f_P(u) \neq 0$, so that the MPRTE parameters of Examples 1 and 2 can be identified without any additional support conditions even though the corresponding PRTE parameters cannot be identified for any α . In these examples, the marginal PRTE parameters place positive weight on $\text{MTE}(u)$ for values of u where the density of P is positive, that is, they only place positive weight on $\text{MTE}(u)$ for values of u where we can use local instrumental variables to point-identify $\text{MTE}(u)$. Thus, even though in each example the PRTE is not identified for any value of $\alpha \neq 0$ without large support assumptions, the marginal PRTE is identified using the assumption that P is a continuous random variable. The MPRTE parameter is thus fundamentally easier to identify than either the conventional treatment parameters or the PRTE parameters.⁷

6. ISSUES IN ESTIMATION

In addition to the support requirement, an additional difficulty in estimating the standard treatment parameters is that under our assumptions they are not \sqrt{N} -estimable. Suppose that P is known. Using equations (5.1) and (5.2), we obtain

$$(6.1) \quad \text{treatment parameter}(j) = \int \frac{\partial}{\partial p} E(Y|P(Z) = p) \omega_j(p) dp \\ = E(g'(P) q_j(P)),$$

where $g'(p) = \frac{\partial}{\partial p} E(Y|P(Z) = p)$ and $q_j(p) = \omega_j(p)/f_P(p)$, and we assume that $\omega_j(p) \neq 0 \Rightarrow f_P(p) \neq 0$ as is required for identification. A critical requirement for weighted average derivative estimators to be \sqrt{N} -consistent is that $q_j(p)f_P(p)$ vanish on the boundary of the support of P (see Newey and Stoker (1993)). Requiring that $q_j(p)f_P(p) = 0$ on the boundaries of the support of $P(Z)$ is equivalent to requiring that $\omega_j(p) = 0$ at the boundaries of the support of P . This requirement is not satisfied by conventional treatment parameters. For example, the average treatment effect imposes the requirement that $\omega_{\text{ATE}}(p) = 1 \forall p \in [0, 1]$, so that $\omega_{\text{ATE}}(p) \neq 0$ at the boundaries of the support and $E(g'(P) q_{\text{ATE}}(P))$ is not \sqrt{N} -estimable under our assumptions. A parallel analysis holds for the PRTE parameters. From equation (3.4), $\omega_{\text{PRTE}}(p; G) = 0$ at the boundaries of the support of P only if $F_P(t) - G(t) = 0$ at the boundaries of the support of P . This requires equality of the supremum of the supports of the two distributions, so that a necessary condition for \sqrt{N} -consistent estimation of the PRTE is that the policy counterfactual does not increase or decrease the largest value of the probability of participation.

⁷As is clear from our analysis, Assumption A-4 can be relaxed to only require that $E(|Y_1||V = v)$ and $E(|Y_0||V = v)$ are bounded for v contained in any arbitrarily small enlargement of the support of $F_V^{-1}(P(Z))$.

In contrast, the MPRTE parameters are \sqrt{N} -estimable under standard regularity conditions. Consider examples of the form of Example 1—policy changes that act like transformations of the P —or examples of the form of Example 2—policy changes that act like transformations of Z . From equation (4.6), $\omega_{\text{MPRTE}}(\cdot; \{F_\alpha\}) \neq 0$ only if $f_P(\cdot) \neq 0$, so we again have that the weights will equal zero on the boundaries of the support of $P(Z)$ if f_P goes to zero on the boundary of the support. In each of these cases, the marginal PRTE parameter is given by a weighted average derivative $E(g'(P)q(P))$ with weights such that $q_j(p)f_P(p)$ equals zero on the boundary of the support of P if $f_P(p) = 0$ on the boundary of the support of P . Thus, each of these parameters is \sqrt{N} -estimable under appropriate regularity conditions (see Newey and Stoker (1993)).

Our discussion thus far ignores three issues. First, in general, $P(Z)$ is not known but must be estimated. Second, the weights sometimes need to be estimated. Third, we have not analyzed how to deal with observed regressors X that enter the outcome equations for Y_0 and Y_1 . If X contains only discrete elements, then the estimation theory just presented is still valid conditional on X . If X contains continuous elements, then we may instead work with marginal PRTE parameters averaged over the distribution of X to obtain \sqrt{N} estimates. One example of an estimator of the weighted average derivative that applies to our problem is the sieve minimum distance estimator of Ai and Chen (2007). We next consider identification and estimation of average marginal treatment effects.

7. THE AVERAGE MARGINAL TREATMENT EFFECT

We now relate the MPRTE parameter to the AMTE, the average effect of treatment for the marginal person who is indifferent between participation and nonparticipation. More precisely, for a given choice of how to measure the distance between P and U , and thus for a given choice of how to measure how close an individual is to being indifferent between treatment or not, we define the AMTE as the average effect of treatment for those who are arbitrarily close to being indifferent between treatment or not. While the MPRTE depends on the direction of the marginal policy change, the AMTE parameter depends on the metric by which one measures how close individuals are to being indifferent. We show an equivalence between the MPRTE and AMTE parameters. Choosing a particular distance measure for the AMTE is equivalent to examining a particular policy direction for the MPRTE. The effect of a marginal policy change in a particular direction is equal to the average effect of treatment for those at the margin of indifference in the precise sense that a marginal policy change in that direction would change their treatment participation decision.

Consider the average effect of treatment for those who are close to being indifferent between treatment or not. For any metric $m(\cdot, \cdot)$, we have

$$E(Y_1 - Y_0 | m(P, U) \leq \varepsilon) = \int_0^1 \text{MTE}(u) \frac{\Pr[m(P, u) \leq \varepsilon]}{\Pr[m(P, U) \leq \varepsilon]} du.$$

Suppose the metric $m(\cdot, \cdot)$ is such that, for some strictly monotonic and differentiable function q , $m(P, U) = |q(P) - q(U)|$. Let r denote the inverse of q . Then, under the regularity conditions that allow us to interchange limits and integration, the average effect for those arbitrarily close to indifference between treatment or not is

$$\lim_{\varepsilon \rightarrow 0} E(Y_1 - Y_0 | m(P, U) \leq \varepsilon) = \int_0^1 \text{MTE}(u) \omega_{\text{AMTE}}(u) du,$$

where

$$(7.1) \quad \omega_{\text{AMTE}}(u) = \frac{f_P(u)r'(u)}{\int f_P(u)r'(u) du} = \frac{f_P(u)r'(u)}{E(r'(P))}.$$

The form of the weights depends on the choice of a metric. Choosing different ways to measure the distance between P and U , and thus different ways to measure how close individuals are to being indifferent between treatment or not, produces different weighting functions. For any choice of q , the weights will be positive, will integrate to unity, and will only be nonzero where the density of P is nonzero, but otherwise the weights can be shifted around arbitrarily by picking alternative metrics. One might think that there is a natural way to pick a particular metric. Taking $m(P, U) = |P - U|$ seems more natural than taking $m(P, U) = |q(P) - q(U)|$ for general $q(\cdot)$ including $F_V^{-1}(\cdot)$. However, under Assumption B-1, it is not any more natural to pick $m(P, U) = |P - U|$ than it is to pick $m(P, U) = |F_V^{-1}(P) - F_V^{-1}(U)| = |Z\gamma - V|$. Yet, the two choices of metrics give different limit results.

Ambiguity over which metric to use for the AMTE can be resolved by connecting the AMTE to the marginal PRTE. The connection is natural. Marginal policy changes only affect people who are indifferent between treatment or not. Thus there is a close connection between the two types of parameters. An arbitrary choice of a metric to define AMTE lacks an economic motivation. The choice of which marginal policy to study to define the AMTE is well motivated. Comparing equations (4.6) and (7.1), we see a duality between MPRTE parameters and alternative definitions of the AMTE. For example, note that AMTE taking q to be the identity function is exactly the MPRTE expression for policy alternatives of the form $P_\alpha = P + \alpha$. If B-1 holds, then AMTE taking $q(\cdot) = F_V^{-1}(\cdot)$ is exactly the MPRTE expression for policy alternatives of the form of changing one component of Z by α . Different policy alternatives define different AMTE parameters.

Consider an analysis of college attendance. Each marginal PRTE parameter defines a different set of “marginal people” who are indifferent between college participation or not and who would change their college participation in response to a marginal change in the policy. Notice that any inframarginal individual with $\mu(Z) > V$ will not be affected by any marginal policy change. Only individuals with $\mu(Z) = V$ will be affected by a marginal change in the policy. Each marginal PRTE parameter will correspond to a different valid definition of the average marginal treatment effect. The average effect of college attendance on those who are on the margin defined by an infinitesimal level shift in college tuition is different from the average effect of college attendance on those who are on the margin defined by an infinitesimal proportional shift in college tuition.

Since each average marginal treatment effect can be equated with a marginal PRTE parameter, we can use the analysis of Section 4 to define the AMTE so that it can be identified without large support assumptions. In addition, the parameter will be \sqrt{N} -estimable under appropriate regularity conditions. Thus, under our assumptions, the AMTE parameter is fundamentally easier to identify and to estimate than are the conventional treatment parameters.

8. APPLYING THE ANALYSIS TO DATA

Following Carneiro, Heckman, and Vytlacil (2009), we estimate the MTE for a sample of white males from the National Longitudinal Survey of Youth (NLSY).⁸ Using this sample, we also estimate the weights needed to construct the MPRTE parameters that correspond to policy perturbations in alternative directions. We group individuals in two groups: persons with a high school education or below who do not go to college ($D = 0$) and persons with some college or above ($D = 1$). The outcome variable is the log of the average of nonmissing values of the hourly wage between 1989 and 1993, which we interpret as an estimate of the log hourly wage in 1991. We estimate the selection probability $P(Z)$ using a logit. We assume that $Y_1 = X\beta_1 + U_1$, $Y_0 = X\beta_0 + U_0$, and that U_1 and U_0 are independent of X , so that the marginal treatment effect is given by $MTE(x, u) = x(\beta_1 - \beta_0) + E(U_1 - U_0 | U = u)$.⁹

⁸See the Supplemental Material Appendix.

⁹The components of Z are the Armed Forces Qualifying Test (AFQT) and its square, mother’s years of schooling and its square, permanent local earnings (average earnings between 1973 and 2000) in the county of residence at age 17 and its square, permanent local unemployment (average unemployment between 1973 and 2000) in state of residence at age 17 and its square, the presence of a four year college in the county of residence at age 17, average tuition in public four year colleges in the county of residence at age 17, log average wage in the county of residence at age 17, and unemployment rate in the state of residence at age 17. The latter four variables are excluded from the set of variables in the wage equations (X) for $D = 1$ and $D = 0$, which are years of experience and its square, log average wage in the county of residence in 1991, unemployment rate in the state of residence in 1991, AFQT and its square, mother’s years of schooling and its square, and the interaction of AFQT and its square with years of experience.

To compute the MTE, we estimate $E(Y|X, P) = X\beta_0 + PX(\beta_1 - \beta_0) + K(P)$ and we take its derivative with respect to P : $\text{MTE}(x, p) = \frac{\partial E(Y|X=x, P=p)}{\partial p}$. We estimate β_1 and β_0 by applying the partially linear regression method of Robinson (1988). $K(\cdot)$ is estimated using locally quadratic regression.¹⁰ We trim 2% of the observations from the extremes of the distribution of P , which means that we are only able to identify the MTE at evaluation points between 0.1 and 0.93. Thus, we can only identify parameters over this support. We present annualized returns, obtained by dividing the MTE (and the parameters it generates) by 4 (corresponding to 4 years of college).

Figure 1 plots three alternative MPRTE weights. We take two cases (which we label A and C) based on Example 1, where $q_\alpha^A(P) = P + \alpha$ and $q_\alpha^C(P) = (1 + \alpha)P$. A third case (which we label B) is taken from Example 2, where $q_\alpha^B(Z^{[k]}) = Z^{[k]} + \alpha$.¹¹ ω_{MPRTE}^A puts more weight than ω_{MPRTE}^C on higher values of U , and both sets of weights are zero in the region of U for which we cannot identify the MTE. The intuitive reason for this pattern is that a proportional change in P results in larger absolute changes for high levels of P than for low levels of P , while an additive change in P results in uniform changes in P across the whole distribution. The MPRTE for a marginal proportional change in P is estimated to be 0.1296, while the MPRTE for a marginal additive change in P is estimated to be 0.0880, since the former puts more weight on high values of U where the MTE is estimated to be smaller. ω_{MPRTE}^B puts more weight than either ω_{MPRTE}^A or ω_{MPRTE}^C in the center of the support of U , and the MPRTE is estimated to be 0.1274. Figure 1 also shows that corresponding to each MPRTE weight there is an AMTE weight, defined by different choices of metric for measuring the distance between P and U . For case A , $m(P, U) = |P - U|$. For case B , $m(P, U) = |F_V^{-1}(P) - F_V^{-1}(U)| = |Z\gamma - V|$. For case C , $m(P, U) = |\frac{P}{U} - 1|$.

square, permanent local earnings (average earnings between 1973 and 2000) in the county of residence at age 17 and its square, and permanent local unemployment (average unemployment between 1973 and 2000) in state of residence at age 17 and its square.

¹⁰We first run kernel regressions of each X on \hat{P} using a bandwidth of 0.05 and trimming 2% of the observations, and we compute the resulting residuals. We then estimate a linear regression of the outcome variable on these residuals and obtain estimates $\hat{\beta}_1$ and $\hat{\beta}_0$. To estimate $K(\cdot)$ and its first derivative, we run a locally quadratic regression of $Y - X\hat{\beta}_0 - \hat{P}X(\hat{\beta}_1 - \hat{\beta}_0)$ on \hat{P} , using a bandwidth of 0.322 (determined by cross validation) and trimming 2% of the observations. $\hat{K}'(\cdot)$ is constructed by taking the coefficient on the linear term of the locally quadratic regression.

¹¹In the empirical work, we impose an index sufficiency restriction when estimating $f_{P|X}$: $f_{P|X} = f_{P|X\delta}$. To estimate the scalar index $X\delta$, we use the fact that $f_{P|X} = f_{P|X\delta}$ implies that $E(P|X) = E(P|X\delta)$ to estimate δ using semiparametric least squares (Ichimura (1993)). We estimate $f_{P|X\delta}$ by running a locally linear regression of $(\frac{1}{h}) * K(\frac{\hat{P}-P}{h})$ on $X\delta$, where $K(\cdot)$ is a standard normal density and $h = 1.06 * [\widehat{\text{Var}}(P)]^{1/2} * n^{(-1/5)}$. We use a bandwidth equal to $1.06 * [\widehat{\text{Var}}(X\delta)]^{1/2} * n^{(-1/5)}$. The figure fixes $X\delta$ at $\bar{X}\delta$. The weights are rescaled so that both the MTE and the weights fit in the figure.

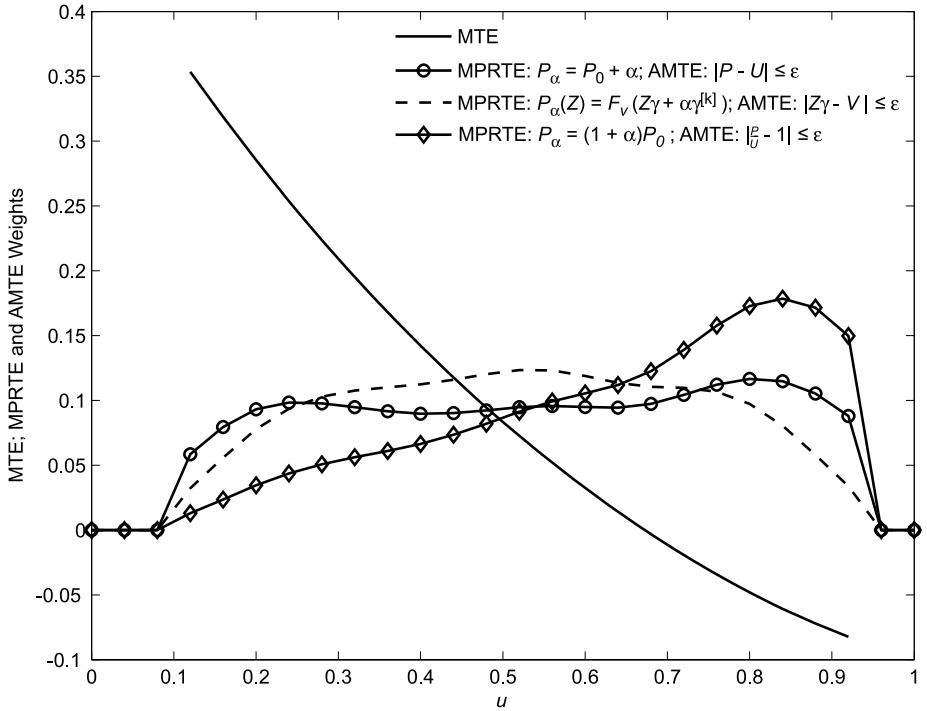


FIGURE 1.—Alternative definitions of the MPRTE and the AMTE. This figure plots the marginal treatment effect ($MTE = E(Y_1 - Y_0|X = x, U = u)$) and the marginal policy relevant treatment effect (MPRTE) weights for three types of policy shifts: $P_a^A = P + \alpha$, $P_a^B(Z) = F_V(Z\gamma + \alpha\gamma^{[k]})$, and $P_a^C = (1 + \alpha)P$, where P is the probability of receiving treatment conditional on the observed covariates (Z). The equivalent definitions of the AMTE correspond, respectively, to the following three metrics for measuring the distance between P and U ($m(P, U)$): $|P - U|$, $|Z\gamma - V|$, and $|\frac{P}{U} - 1|$. The MTE, the MPRTE, and the AMTE weights are evaluated at values of X such that $E(Y_1 - Y_0|X = x) = 0.13$. The average marginal policy effects are 0.1296 for the additive shift in P (A), 0.1274 for the additive shift in Z (B), and 0.0880 for the proportional shift in P (C). Estimated from NLSY data (see Carneiro, Heckman, and Vytlacil (2009), for details on estimating MTE).

9. SUMMARY AND CONCLUSIONS

This paper extends the analysis of Heckman and Vytlacil (1999, 2005, 2007) by using the MTE to identify the effect of a marginal policy change, and to identify the average effect of treatment on individuals who are indifferent between treatment or not. Conventional treatment parameters and the PRTE require large support conditions for identification and often are not \sqrt{N} -estimable. Under our assumptions, the marginal policy effect parameters and the average marginal treatment effects are generally identified without large support conditions and are \sqrt{N} -estimable. An analysis of the effect of

marginal changes in tuition policies on college attendance illustrates the empirical relevance of this analysis.

REFERENCES

- AI, C., AND X. CHEN (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models With Different Conditioning Variables," *Journal of Econometrics*, 141, 5–43. [388]
- ANGRIST, J., K. GRADDY, AND G. IMBENS (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models With an Application to the Demand for Fish," *Review of Economic Studies*, 67, 499–527. [386]
- BJÖRKLUND, A., AND R. MOFFITT (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection," *Review of Economics and Statistics*, 69, 42–49. [386]
- CARD, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160. [380]
- CARNEIRO, P., J. J. HECKMAN, AND E. J. VYTLACIL (2009): "Estimating Marginal Returns to Education," Unpublished Manuscript, University College London, Department of Economics. [390,392]
- HECKMAN, J. J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462; addendum published (1998), *Journal of Human Resources*, 33. [386]
- HECKMAN, J. J., AND E. J. VYTLACIL (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734. [385,386,392]
- (2001a): "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect," in *Econometric Evaluation of Labour Market Policies*, ed. by M. Lechner and F. Pfeiffer. New York: Center for European Economic Research, 1–15. [386]
- (2001b): "Policy-Relevant Treatment Effects," *American Economic Review*, 91, 107–111. [377,379–381]
- (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. [377–380,385,386,392]
- (2007): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, Vol. 6B, ed. by J. Heckman and E. Leamer. Amsterdam: Elsevier, 4875–5144. [379,380,386,392]
- HURWICZ, L. (1962): "On the Structural Form of Interdependent Systems," in *Logic, Methodology and Philosophy of Sciences*, ed. by E. Nagel, P. Suppes, and A. Tarski. Stanford, CA: Stanford University Press, 232–239. [382]
- ICHIMURA, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120. [391]
- ICHIMURA, H., AND C. TABER (2002): "Direct Estimation of Policy Impacts," Unpublished Working Paper, University College London, Department of Economics. [382]
- IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [379,386]
- NEWHEY, W., AND T. M. STOKER (1993): "Efficiency of Weighted Average Derivative Estimators and Index Models," *Econometrica*, 61, 1199–1223. [387,388]
- ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954. [391]
- STOCK, J. H. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association*, 84, 565–575. [380]

VYTLACIL, E. J. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. [379]

Dept. of Economics, University College London, Gower Street, London WC1E 6BT, U.K. and IFS and CeMMAP; p.carneiro@ucl.ac.uk,

*Dept. of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, U.S.A. and Geary Institute, University College Dublin; Cowles Foundation, Yale University; American Bar Foundation; jjh@uchicago.edu,
and*

Dept. of Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, U.S.A.; edward.vytlacil@yale.edu.

Manuscript received April, 2007; final revision received July, 2009.

COPULAS AND TEMPORAL DEPENDENCE

BY BRENDAN K. BEARE¹

An emerging literature in time series econometrics concerns the modeling of potentially nonlinear temporal dependence in stationary Markov chains using copula functions. We obtain sufficient conditions for a geometric rate of mixing in models of this kind. Geometric β -mixing is established under a rather strong sufficient condition that rules out asymmetry and tail dependence in the copula function. Geometric ρ -mixing is obtained under a weaker condition that permits both asymmetry and tail dependence. We verify one or both of these conditions for a range of parametric copula functions that are popular in applied work.

KEYWORDS: Copula, Markov chain, maximal correlation, mean square contingency, mixing, canonical correlation, tail dependence.

1. INTRODUCTION

IN RECENT YEARS, a number of authors have considered the possibility of modeling a univariate stationary discrete-time Markov chain by specifying (a) the invariant distribution F and (b) a bivariate copula C that characterizes the dependence between consecutive realizations. All finite dimensional distributions of the chain are uniquely determined by C and F . Some interesting patterns of temporal dependence can be generated using certain copula functions. A copula that exhibits tail dependence may generate a Markov chain which appears to become substantially more serially dependent as it draws toward the extremes of the state space. For instance, one could think of the phenomenon of unemployment hysteresis, in which unemployment appears to become “stuck” at relatively high levels, as corresponding to a copula that exhibits upper tail dependence. Asymmetric copulas can be used to model nonreversible temporal behavior, such as the tendency for many economic and financial time series to exhibit relatively long periods of fairly steady growth, interspersed with shorter periods of sharp decline.

Various procedures for estimating models of this kind have been proposed, ranging from fully parametric (Joe (1997, Chap. 8)) to semiparametric (Chen and Fan (2006), Chen, Wu, and Yi (2009)) and quasi-nonparametric (Gagliardini and Gouriéroux (2007)). Fentaw and Naik-Nimbalkar (2008) developed methods applicable to Markov chains generated by copula functions with time-varying parameters. Applications of these procedures have considered air quality measurements (Joe (1997, Chap. 8)), stochastic volatility in stock

¹Parts of this paper derive from my doctoral research at Yale University. I thank my advisor Peter Phillips and committee members Donald Andrews and Yuichi Kitamura for their support and advice. I also thank Xiaohong Chen, Rustam Ibragimov, Bent Nielsen, Andres Santos, and three anonymous referees for helpful comments. Financial support from the Cowles Foundation under a Carl Arvid Anderson Prize Fellowship is gratefully acknowledged.

returns (Ibragimov and Lentzas (2008)), and coffee prices in Ethiopia (Fentaw and Naik-Nimbalkar (2008)). Gagliardini and Gouriéroux (2008) discussed the application of their methods to intertrade durations in financial markets. In a panel data context, related methods have been applied by Bonhomme and Robin (2006) and Dearden, Fitzsimons, Goodman, and Kaplan (2008) to model earnings dynamics. Patton (2008) provided a brief review of a number of these papers, among others.

An important issue that has arisen in much of this literature is the following: How does the form of the copula C relate to the strength of temporal dependence in the Markov chain? More specifically, what conditions on C will ensure that weak dependence conditions sufficient for the application of invariance principles are satisfied? As suggested by Chen and Fan (2006), verification of the stability conditions of Meyn and Tweedie (1993) that guarantee geometric ergodicity provides one possible approach to demonstrating weak dependence properties for specific copula functions. Gagliardini and Gouriéroux (2008) used this approach to identify a condition under which proportional hazard copulas generate a geometric rate of β -mixing. Ibragimov and Lentzas (2008) claimed to provide numerical evidence that is suggestive of long memory in chains generated by Clayton copulas. Ibragimov (2009) proposed a class of Fourier copulas which generate m -dependent Markov chains under suitable conditions. m -dependence implies the satisfaction of mixing conditions at arbitrarily fast rates.

In this paper, we identify conditions on C that suffice for geometrically fast mixing rates. Geometric β -mixing, equivalent to geometric ergodicity for stationary Markov chains, is established under a rather strong condition that excludes copulas that exhibit tail dependence or asymmetry. Geometric ρ -mixing, which implies geometric α -mixing, is obtained under a much weaker condition. We verify this condition for various parametric copula functions that are popular in applied work. ρ -, β -, and α -mixing conditions may be used as the basis for a range of inequalities and limit theorems that are useful in demonstrating the asymptotic validity of statistical methods; see, for example, the monograph by Doukhan (1994) or the recent three volume series by Bradley (2007).

Contemporaneous work by Chen, Wu, and Yi (2009) relates closely to the results reported here. Specifically, Chen, Wu, and Yi established that the Clayton, Gumbel, and t -copulas generate Markov chains with a geometric rate of β -mixing. Since these three copula functions exhibit positive tail dependence, they are not covered by our results concerning geometric β -mixing. Nevertheless, we do provide results pertaining to the ρ -mixing properties of Markov chains generated by these copula functions. Since the β -mixing property does not imply ρ -mixing, or vice versa, our results are complementary to those of Chen, Wu, and Yi.

The structure of the paper is as follows. In Section 2, we introduce our notation and some basic definitions and results. In Section 3, we consider β -mixing

conditions, while in Section 4 we consider ρ -mixing conditions. Section 5 concludes. Mathematical proofs are collected in the Appendix in the Supplemental Material (Beare (2010)).

2. BASIC SETUP

We begin with the following rather basic definition of a bivariate copula function.

DEFINITION 2.1: A bivariate copula function is a bivariate probability distribution function on $[0, 1]^2$ for which the two univariate marginal distribution functions are uniform on $[0, 1]$.

Our concern in this paper is with bivariate dependence; when we refer to a copula function or copula, we mean a bivariate copula function. Suppose that X and Y are real-valued random variables with joint distribution function $F_{X,Y}$ and marginal distribution functions F_X and F_Y . We will say that X and Y admit the copula C if $C(F_X(x), F_Y(y)) = F_{X,Y}(x, y)$ for all $x, y \in \mathbb{R}$. A fundamental result concerning copulas is Sklar's theorem (1959), which is proved in Schweizer and Sklar (1974). A useful, more recent treatment of Sklar's theorem can be found in Nelsen (1999). Sklar's theorem ensures that for any random variables X and Y , there exists a copula C such that X and Y admit C . Moreover, C is uniquely defined on the product of the ranges of the marginal distribution functions of X and Y . Hence, C is unique if X and Y are continuous random variables. If X or Y is not continuous, C may nevertheless be uniquely defined by bilinear interpolation between uniquely defined values; see, for example, the proof of Lemma 2.3.5 in Nelsen (1999). Following Darsow, Nguyen, and Olsen (1992), we will refer to this unique copula as the copula of X and Y .

Let $\{Z_t : t \in \mathbb{Z}\}$ be a stationary sequence of real-valued random variables defined on a probability space (Ω, \mathcal{F}, P) and for $s, t \in \mathbb{Z} \cup \{-\infty, \infty\}$, let $\mathcal{F}_s^t \subseteq \mathcal{F}$ denote the σ -field generated by the random variables $\{Z_r : s \leq r \leq t\}$. We assume that $\{Z_t\}$ is a Markov chain. That is, for any finite collection of integers $\{t_1, \dots, t_n\}$ satisfying $t_1 < t_2 < \dots < t_n$ and any $z \in \mathbb{R}$, we have

$$E(1_{\{Z_{t_n} \leq z\}} | Z_{t_1}, \dots, Z_{t_{n-1}}) = E(1_{\{Z_{t_n} \leq z\}} | Z_{t_{n-1}})$$

almost surely. Let F denote the marginal distribution function of Z_0 and for $k \in \mathbb{N}$ let C_k be the copula of Z_0 and Z_k . We will often write C in place of C_1 . Theorem 3.2 of Darsow, Nguyen, and Olsen (1992) asserts that the copulas C_k satisfy

$$(2.1) \quad C_{k+1}(x, y) = \int_0^1 \frac{\partial C_k(x, z)}{\partial z} \cdot \frac{\partial C(z, y)}{\partial z} dz$$

for all $k \in \mathbb{N}$ and all $x, y \in [0, 1]^2$. The existence of partial derivatives in (2.1) for almost all $z \in [0, 1]$ is a well known property of copula functions; see, for example, Theorem 2.2.7 in Nelsen (1999).

Equation (2.1) constitutes a version of the Chapman–Kolmogorov equations for Markov chains, expressed in terms of copula functions. It implies that all bivariate copulas C_k can be expressed in terms of the copula C . In the next two sections we will identify conditions on C such that mixing coefficients determined by the copulas C_k decay to zero at a geometric rate as k increases.

3. SUFFICIENT CONDITIONS FOR GEOMETRIC β -MIXING

Sequences of β -mixing coefficients provide one way to characterize the serial persistence of time series. We shall employ the following definition.

DEFINITION 3.1: The β -mixing coefficients $\{\beta_k : k \in \mathbb{N}\}$ that correspond to the sequence of random variables $\{Z_t\}$ are given by

$$\beta_k = \frac{1}{2} \sup_{m \in \mathbb{Z}} \sup_{\{A_i\}, \{B_j\}} \sum_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i)P(B_j)|,$$

where the second supremum is taken over all finite partitions $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, B_J\}$ of Ω such that $A_i \in \mathcal{F}_{-\infty}^m$ for each i and $B_j \in \mathcal{F}_{m+k}^\infty$ for each j .

An equivalent form of Definition 3.1 was originally stated by Volkonskii and Rozanov (1959), although they attribute it to Kolmogorov. The condition $\beta_k \rightarrow 0$ as $k \rightarrow \infty$ is variously referred to as strong regularity, complete regularity, absolute regularity, or simply β -mixing. Several equivalent formulations of Definition 3.1 have been used in the literature, often involving conditional probabilities or total variation norms. Numerous results concerning these equivalencies may be found in Chapter 3 of Bradley (2007).

Before stating our first theorem, we require an additional definition.

DEFINITION 3.2: The maximal correlation ρ_C of the copula C is given by

$$(3.1) \quad \sup_{f,g} \left| \int_0^1 \int_0^1 f(x)g(y)C(dx, dy) \right|,$$

where the supremum is taken over all $f, g \in L_2[0, 1]$ such that $\int f = \int g = 0$ and $\int f^2 = \int g^2 = 1$.

The integral in (3.1) is defined in the usual Lebesgue–Stieltjes sense. Maximal correlation is an old concept; Rényi (1959) provided an early discussion of its properties as a measure of dependence between random variables.

The first result of this section is as follows.

THEOREM 3.1: Suppose that C is symmetric and absolutely continuous with square-integrable density c , and that $\rho_C < 1$. Then there exists $A < \infty$ and $\gamma > 0$ such that $\beta_k \leq Ae^{-\gamma k}$ for all k .

REMARK 3.1: Here and elsewhere, we say that a copula C is symmetric if $C(x, y) = C(y, x)$ for all $x, y \in [0, 1]$.

REMARK 3.2: Theorem 3.1 is proved in two main steps. First, one establishes the bound $\beta_k \leq \frac{1}{2} \|c_k - 1\|_2$, where c_k is the density of C_k . Next, a spectral decomposition of c_k is used to show $\|c_k - 1\|_2 \leq \rho_C^{k-1} \|c - 1\|_2$. The validity of this decomposition depends crucially on c being symmetric and square integrable.

REMARK 3.3: The quantity $\|c - 1\|_2^2$ is referred to as the mean square contingency of the joint distribution of Z_0 and Z_1 . Mean square contingency was defined formally by Lancaster (1958) without reference to copula functions, but its origins can be traced back to work by Pearson in the early twentieth century. Lancaster (1958), Rényi (1959), and Sarmanov (1958a, 1958b, 1961) used spectral or singular value decompositions to study the structure of bivariate distributions that exhibit finite mean square contingency. The work of Sarmanov (1961) is of special relevance here, as he was concerned in particular with the bivariate distributions associated with stationary Markov chains.

REMARK 3.4: The proof of Theorem 3.1 establishes geometric decay of $\|c_k - 1\|_2$, which is stronger than the statement of the theorem. Geometric decay of $\|c_k - 1\|_2$ implies geometric decay of a range of dependence measures between Z_0 and Z_k other than β -mixing coefficients. For instance, Proposition 1 of Ibragimov and Lentzas (2008) established $\|c_k - 1\|_2$ or $\|c_k - 1\|_2^2$ as an upper bound on the relative entropy, Hellinger distance, linear correlation, and Schweizer–Wolff (1981) distances between Z_0 and Z_k . These quantities must therefore decay geometrically as $k \rightarrow \infty$, under the assumptions of Theorem 3.1.

REMARK 3.5: For stationary Markov chains, a geometric rate of β -mixing holds if and only if the chain satisfies geometric ergodicity. See Definition 21.18 and Theorem 21.19 in Bradley (2007) for a definition of geometric ergodicity and statement of this result.

REMARK 3.6: The following result may prove useful in verifying the condition $\rho_C < 1$ that appears in the statement of Theorem 3.1.

THEOREM 3.2: Suppose that C is absolutely continuous with square-integrable density c . Then $\rho_C = 1$ if and only if there exist measurable sets $A, B \subset [0, 1]$ with measure strictly between 0 and 1 such that $c = 0$ almost everywhere on $(A \times B) \cup (A^c \times B^c)$. In particular, if $c > 0$ almost everywhere on $[0, 1]^2$, then $\rho_C < 1$.

Theorem 3.2 indicates that, for absolutely continuous copulas with square-integrable densities, the condition $\rho_C < 1$ is rather easy to satisfy. Copula functions used in applied work typically have a density that is positive almost everywhere.

REMARK 3.7: The assumption that C is symmetric implies that the Markov chain $\{Z_t\}$ is time reversible. There is substantial evidence that many economic and financial series exhibit irreversible behavior; see, for example, McCausland (2007). The proportional hazard copulas considered by Gagliardini and Gouriéroux (2008) are asymmetric in general. Nevertheless, many parametric families of copulas used in applied work do satisfy the symmetry condition. In particular, all Archimedean copulas are symmetric.

REMARK 3.8: Commonly used parametric copula functions that satisfy the conditions of Theorem 3.1 include the Farlie–Gumbel–Morgenstern, Frank, and Gaussian copulas. See Nelsen (1999) for definitions and origins. For the latter two copulas, this assumes that the copula parameter is in the interior of the parameter space, so that the copula does not degenerate to the Fréchet–Höffding upper or lower bound. Square integrability of the Farlie–Gumbel–Morgenstern and Frank copula densities follows from the fact that they are bounded. For the Gaussian copula density, square integrability follows from Mehler’s identity, which provides a mean square convergent expansion of the bivariate Gaussian density in terms of Hermite polynomials. That is,

$$(3.2) \quad f_\rho(x, y) = f(x)f(y) \sum_{k=0}^{\infty} \rho^k H_k(x)H_k(y),$$

where f_ρ is the standard bivariate Gaussian density with correlation coefficient $\rho \in (0, 1)$, f is the standard Gaussian density, and $\{H_k\}$ is a sequence of functions that is orthonormal with respect to the density f , with each H_k being a polynomial of order k . Equation (3.2) implies that the Gaussian copula density c_ρ satisfies

$$(3.3) \quad c_\rho(x, y) = \sum_{k=0}^{\infty} \rho^k \phi_k(x)\phi_k(y),$$

with $\{\phi_k\}$ being an orthonormal sequence of functions in $L_2[0, 1]$. It follows that $\|c_\rho\|_2 = 1/\sqrt{1 - \rho^2} < \infty$. Other bivariate distributions that admit expansions analogous to (3.2) include the bivariate gamma, Poisson, binomial, and hypergeometric distributions, and the compound correlated bivariate Poisson distribution. See Hamdan and Al-Bayyati (1971) for references and further discussion. The copula functions that correspond to these distributions satisfy the conditions of Theorem 3.1.

REMARK 3.9: Not all copulas of interest satisfy the conditions of Theorem 3.1. The Marshall–Olkin copula (see, e.g., Nelsen (1999, p. 46)) is not absolutely continuous nor is it symmetric in general. Furthermore, any copula exhibiting upper or lower tail dependence will not admit a square-integrable density. Following, for example, McNeil, Frey, and Embrechts (2005, p. 209), we define tail dependence as follows.

DEFINITION 3.3: The coefficient of lower tail dependence μ_L that corresponds to a copula C is given by

$$\mu_L = \lim_{x \rightarrow 0^+} \frac{C(x, x)}{x}$$

if the limit exists. The coefficient of upper tail dependence is given by

$$\mu_U = \lim_{x \rightarrow 1^-} \frac{1 - 2x + C(x, x)}{1 - x}$$

if the limit exists.

If μ_L exists and is positive, we say that C exhibits lower tail dependence, while if μ_U exists and is positive, we say that C exhibits upper tail dependence.

The following result states that tail dependence is ruled out by our requirement in Theorem 3.1 that C have square-integrable density c .

THEOREM 3.3: *Let C be an absolutely continuous copula with square-integrable density c . Then C exhibits neither upper nor lower tail dependence.*

Theorem 3.3 implies that several parametric classes of copulas used frequently in applied work have a density that is not square integrable. In particular, the Gumbel, Clayton, and t -copulas all exhibit upper or lower tail dependence; see, for example, Examples 5.3.1 and 5.3.3 in McNeil, Frey, and Embrechts (2005). The connection between tail dependence and the square integrability of c does not seem to have been noted in previous literature.

REMARK 3.10: Although Theorem 3.3 implies that the conditions of Theorem 3.1 are not satisfied by copulas that exhibit upper or lower tail dependence, we do not assert that such copulas necessarily generate stationary Markov chains for which the decay rate of β -mixing coefficients is subgeometric. Indeed, Chen, Wu, and Yi (2009) established that the Gumbel, Clayton, and t -copulas all generate a geometric rate of β -mixing. This result indicates that the conditions of Theorem 3.1 are sufficient but not necessary for the stated conclusion.

REMARK 3.11: Theorem 3.3 bears an additional implication that is not related to time series applications in particular. Gagliardini and Gouriéroux (2007) proposed a quasi-nonparametric approach to copula estimation that involves minimizing a weighted chi-squared distance between an unconstrained kernel estimate of c and a constrained estimate. The constrained estimate restricts c to be in a class of copula densities that is specified up to a one-dimensional functional parameter; that is, a vector-valued function of one variable. Archimedean copulas are an example of a class of copulas that may be parameterized in this way, with the additive generator of each copula being its functional parameter; see, for example, Nelsen (1999, p. 92). Another example is the class of proportional hazard copulas, as shown by Gagliardini and Gouriéroux (2008).

Suppose we have a sample of n pairs of observations, assumed for simplicity to have marginal distributions that are uniform on $[0, 1]$. Given a set of functional parameters Θ , the estimated functional parameter $\hat{\theta}_n$ minimizes the criterion function

$$M_n(\theta) = \int_0^1 \int_0^1 \frac{[c(x, y; \theta) - \hat{c}_n(x, y)]^2}{\hat{c}_n(x, y)} w_n(x, y) dx dy,$$

where $\hat{c}_n(x, y)$ is the unconstrained estimate of $c(x, y)$, $c(x, y; \theta)$ is the copula density corresponding to the functional parameter θ , and $w_n(x, y)$ is a smooth weighting function that converges pointwise to 1 as $n \rightarrow \infty$. Gagliardini and Gouriéroux (2007) did not discuss the choice of the weighting function w_n . Theorem 3.3 implies that if $c(x, y; \theta)$ exhibits tail dependence for values of θ in a neighborhood of the true value, then the choice of w_n may be rather important. Specifically, for each n , we will need to have $w_n(x, x) \rightarrow 0$ at some rate as $x \rightarrow 0$ and/or as $x \rightarrow 1$ in order for $M_n(\theta)$ to be finite in a neighborhood of the true value. This issue merits further investigation, but goes beyond the scope of this paper.

4. SUFFICIENT CONDITIONS FOR GEOMETRIC ρ -MIXING

β -mixing coefficients provide one way to characterize the serial persistence of $\{Z_t\}$. One may also characterize this persistence using ρ -mixing coefficients. We shall employ the following definition.

DEFINITION 4.1: The ρ -mixing coefficients $\{\rho_k : k \in \mathbb{N}\}$ that correspond to the sequence of random variables $\{Z_t\}$ are given by

$$\rho_k = \sup_{m \in \mathbb{Z}} \sup_{f, g} |\text{Corr}(f, g)|,$$

where the second supremum is taken over all square-integrable random variables f and g measurable with respect to $\mathcal{F}_{-\infty}^m$ and \mathcal{F}_{m+k}^∞ respectively, with pos-

itive and finite variance, and where $\text{Corr}(f, g)$ denotes the correlation between f and g .

Definition 4.1 was originally stated by Kolmogorov and Rozanov (1960). Rosenblatt (1971, Chap. 7) provided an important discussion of ρ -mixing conditions in the context of stationary Markov chains. ρ -mixing conditions appear to have been largely ignored in the econometrics literature. A recent exception is Chen, Hansen, and Carrasco (2008), who studied the ρ - and β -mixing properties of nonlinear diffusion processes.

A definition of mixing that has been much more heavily employed in econometrics is that of α -mixing.

DEFINITION 4.2: The α -mixing coefficients $\{\alpha_k : k \in \mathbb{N}\}$ that correspond to the sequence of random variables $\{Z_t\}$ are given by

$$\alpha_k = \sup_{m \in \mathbb{Z}} \sup_{A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+k}^\infty} |P(A \cap B) - P(A)P(B)|.$$

Definition 4.2 is commonly attributed to Rosenblatt (1956). However, Beare (2007) observed that this definition is in fact different from that proposed by Rosenblatt and appears to have been stated first by Volkonskii and Rozanov (1959), who referred to their condition as being “analogous” to that of Rosenblatt. The inequalities $2\alpha_k \leq \beta_k$ and $4\alpha_k \leq \rho_k$ (see, e.g., Proposition 3.11 in Bradley (2007)) ensure that α -mixing is a weaker dependence condition than both β - and ρ -mixing. Neither of the β - and ρ -mixing conditions is implied by the other: there exist processes for which $\beta_k \rightarrow 0$ but $\rho_k \not\rightarrow 0$ as $k \rightarrow \infty$ and vice versa. Central limit theorems are available for ρ -mixing processes that involve weaker assumptions on moments and mixing rates than do those available for β -mixing processes. See, for instance, Remark 10.11(8) and Theorems 10.7 and 11.4 in Bradley (2007). On the other hand, β -mixing conditions have been used to establish a variety of other useful results, such as, for instance, the central limit theorem for U -statistics in Arcones (1995).

The following result identifies a simple condition on C such that ρ_k decays at a geometric rate.

THEOREM 4.1: Suppose $\rho_C < 1$. Then there exist $A < \infty$ and $\gamma > 0$ such that $\rho_k \leq Ae^{-\gamma k}$ for all k .

REMARK 4.1: Theorem 4.1 is little more than a reformulation of Theorem 7.5(I)(a) of Bradley (2007) in terms of copula functions. Although well understood in the probability literature, the connection between mixing conditions and maximal correlation appears to have gone unmentioned in the recent literature in statistics and econometrics on copula-based time series.

REMARK 4.2: The following example of a stationary autoregressive process that is not α -mixing may be familiar to econometricians. It was studied in detail by Andrews (1984); see also Example 2.15 in Bradley (2007) and the references given there. Let $\{\varepsilon_t : t \in \mathbb{Z}\}$ be an independent sequence of random variables that each take the value 0 with probability 1/2 and the value 1/2 with probability 1/2, and for $t \in \mathbb{Z}$, let Z_t be the limit in mean square of the series $\sum_{k=0}^{\infty} 2^{-k} \varepsilon_{t-k}$. Andrews (1984) showed that the α -mixing coefficients that correspond to $\{Z_t\}$ satisfy $\alpha_k = 1/4$ for all k . Since $\{Z_t\}$ is a stationary Markov chain, it follows from Theorem 4.1 and the inequality $4\alpha_k \leq \rho_k$ that the copula C_A of Z_0 and Z_1 , which we will refer to as the Andrews copula, must have maximal correlation coefficient $\rho_{C_A} = 1$.

Let us consider the form of the Andrews copula in more detail. It is known that the marginal distribution of Z_0 and of Z_1 is uniform on $[0, 1]$ (Bradley (2007), Vol. 1, p. 58)). Consequently, C_A is simply the joint distribution function of Z_0 and Z_1 . It takes the form

$$(4.1) \quad C_A(x, y) = \min \left\{ x, y, \frac{1}{2}x - \frac{1}{2} + \max \left\{ y, \frac{1}{2} \right\} \right\}.$$

Equation (4.1) can be verified using elementary methods. Viewed as a probability distribution on $[0, 1]^2$, the Andrews copula is absolutely singular with respect to Lebesgue measure, assigning half its mass uniformly along the line $y = x/2$ and half its mass uniformly along the line $y = x/2 + 1/2$. Letting $f(x) = x$ and $g(y) = 2y1(y \leq 1/2) + (2y - 1)1(y > 1/2)$, it is straightforward to see that $f(Z_0) = g(Z_1)$ almost surely, confirming that the Andrews copula does indeed satisfy $\rho_{C_A} = 1$.

REMARK 4.3: From Theorem 3.2 and Remark 3.8, we know that the Farlie–Gumbel–Morgenstern, Frank, and Gaussian copulas satisfy $\rho_C < 1$, provided in the case of the latter two copulas that the copula parameters are in the interior of the respective parameter spaces. Consequently, stationary Markov chains generated using such copula functions satisfy ρ -mixing conditions at a geometric rate.

REMARK 4.4: The following result may prove useful in verifying the assumption $\rho_C < 1$ for specific copula functions.

THEOREM 4.2: *If the density of the absolutely continuous part of C is bounded away from zero on a set of measure 1, then $\rho_C < 1$.*

The proof of Theorem 4.2 uses only elementary methods. Compared to Theorem 3.2, Theorem 4.2 relaxes the requirement that C be absolutely continuous with square-integrable density, but rules out copulas whose densities become arbitrarily close to zero. The same trade-off can be found in Corollary 2.7 and Proposition 2.8 of Bryc (1996), which concern the solubility of a certain in-

verse problem involving conditional expectation operators. The proof of Theorem 4.2 resembles the proof of Lemma 3.4 in that paper.

The t -copula and Marshall–Olkin copula (with parameters in the interior of the respective parameter spaces) provide two examples of copula functions whose densities are bounded away from zero. Consequently, Theorems 4.1 and 4.2 jointly imply that stationary Markov chains constructed using t -copulas or Marshall–Olkin copulas will satisfy ρ - and α -mixing conditions at a geometric rate. Unfortunately, not all copula functions used in applied work satisfy the condition of Theorem 4.2. In particular, the Clayton and the Gumbel copula densities tend to zero at the off-diagonal corners $(0, 1)$ and $(1, 0)$.

REMARK 4.5: One might conjecture that Theorem 4.2 would remain true if one required that the density of the absolutely continuous component of C merely be positive on a set of measure 1, rather than bounded away from zero. In fact, this conjecture is false. We now provide an example of a pair of random variables X and Y whose unique copula function C satisfies $\rho_C = 1$ and is absolutely continuous with density $c > 0$ almost everywhere. This example was inspired by Figure 1 in Rényi (1959). Let X, Y have joint probability density function (p.d.f.) $h: [0, 1]^2 \rightarrow \mathbb{R}$ given by

$$h(x, y) = x^3 y^3 + A \cdot 1(\log(1+x) \leq y \leq e^x - 1),$$

where $A > 0$ is such that $\int_0^1 \int_0^1 h(x, y) dx dy = 1$. Note that $h(x, y) = h(y, x)$ for all $x, y \in [0, 1]$. Since the joint and marginal densities of X and Y are positive almost everywhere, X and Y admit a unique copula C that is absolutely continuous with density $c > 0$ almost everywhere (see, e.g., McNeil, Frey, and Embrechts (2005, p. 197)). As $z \rightarrow 0^+$, one can show using Taylor approximations that

$$\begin{aligned} P(X \leq z, Y \leq z) &= \frac{1}{16} z^8 + A(z^2 + 2z - 2(1+z)\log(1+z)) \\ &= \frac{1}{3} Az^3 + O(z^4) \end{aligned}$$

and that

$$P(X \leq z) = \frac{1}{16} z^4 + A(e^z - 1 - (1+z)\log(1+z)) = \frac{1}{3} Az^3 + O(z^4),$$

with the first equality in the second line holding for $z \in [0, \log 2]$. Therefore, the coefficient of lower tail dependence that corresponds to C satisfies

$$\begin{aligned} \mu_L &= \lim_{z \rightarrow 0^+} \frac{C(z, z)}{z} = \lim_{z \rightarrow 0^+} \frac{C(F(z), F(z))}{F(z)} \\ &= \lim_{z \rightarrow 0^+} \frac{P(X \leq z, Y \leq z)}{P(X \leq z)} = 1, \end{aligned}$$

where F is the common cumulative distribution function (c.d.f.) of X and Y . For $n \in \mathbb{N}$, define the function $f_n \in L_2[0, 1]$ by

$$f_n(x) = \frac{n}{\sqrt{n-1}} \left(1\left(x \leq \frac{1}{n} \right) - \frac{1}{n} \right).$$

Note that $\int f_n = 0$ and $\int f_n^2 = 1$. It is simple to show that

$$\begin{aligned} \int_0^1 \int_0^1 f_n(x) f_n(y) C(dx, dy) &= \frac{n^2}{n-1} C\left(\frac{1}{n}, \frac{1}{n}\right) - \frac{1}{n-1} \\ &= nC\left(\frac{1}{n}, \frac{1}{n}\right) + o\left(\frac{1}{n}\right). \end{aligned}$$

Letting $n \rightarrow \infty$, we obtain

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 f_n(x) f_n(y) C(dx, dy) = \lim_{z \rightarrow 0^+} \frac{C(z, z)}{z} = \mu_L = 1,$$

implying that $\rho_C = 1$.

REMARK 4.6: For copulas that do not satisfy the condition of Theorem 4.2 and for which ρ_C is unknown, one may nevertheless seek to verify the condition $\rho_C < 1$ numerically.

THEOREM 4.3: *Given a copula C , for $n \in \mathbb{N}$, let K_n be the $n \times n$ matrix with (i, j) th element given by*

$$\begin{aligned} K_n(i, j) &= C\left(\frac{i}{n}, \frac{j}{n}\right) - C\left(\frac{i}{n}, \frac{j-1}{n}\right) - C\left(\frac{i-1}{n}, \frac{j}{n}\right) \\ &\quad + C\left(\frac{i-1}{n}, \frac{j-1}{n}\right) - \frac{1}{n^2}. \end{aligned}$$

Let ϱ_n denote the maximum eigenvalue of K_n . Then $n\varrho_n \rightarrow \rho_C$ as $n \rightarrow \infty$.

Theorem 4.3 shows that ρ_C may be approximated arbitrarily well by computing the maximum eigenvalue of a matrix of sufficiently large dimension. No assumptions whatsoever are placed on the copula C . In Figure 4.1 we plot approximate values of ρ_C for the Clayton and Gumbel copulas. These copulas are given by

$$C_{\text{Clayton}}(x, y; \theta) = (x^{-\theta} + y^{-\theta} - 1)^{-1/\theta},$$

$$C_{\text{Gumbel}}(x, y; \theta) = \exp(-((-\ln x)^\theta + (-\ln y)^\theta)^{1/\theta}),$$

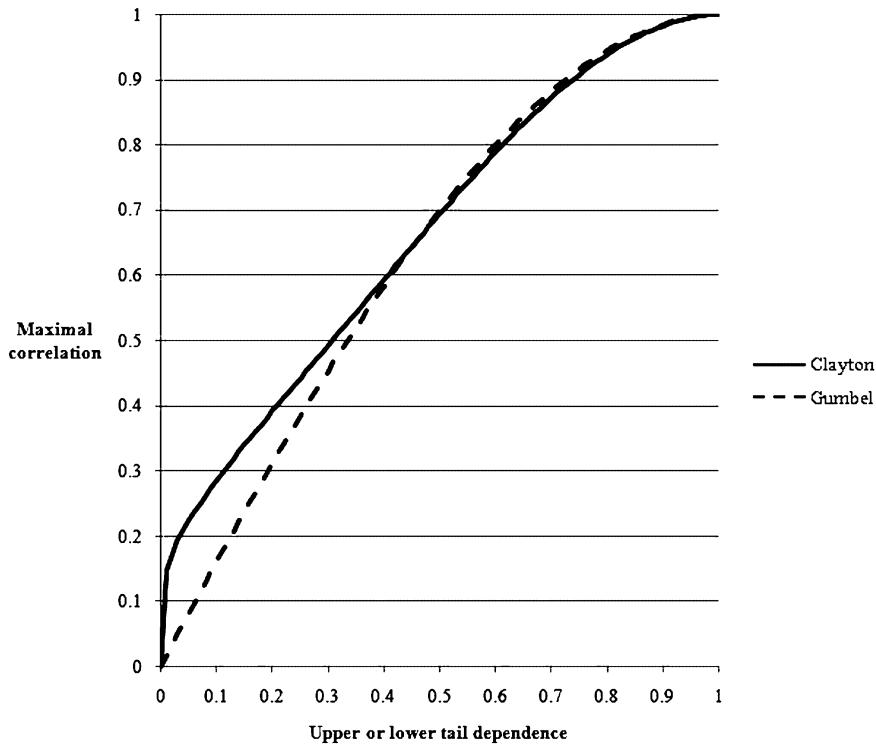


FIGURE 4.1.—Maximal correlation coefficients for Clayton and Gumbel copulas.

where we have $\theta \in (0, \infty)$ for the Clayton copula and $\theta \in (1, \infty)$ for the Gumbel copula. The horizontal axis measures lower tail dependence for the Clayton copula and upper tail dependence for the Gumbel copula. In the former case, we have $\mu_L = 2^{-1/\theta}$, while in the latter case, we have $\mu_U = 2 - 2^{1/\theta}$; see, for example, McNeil, Frey, and Embrechts (2005). At each increment of 0.01 on the horizontal axis, an approximate maximal correlation coefficient for each copula was calculated by computing the maximum eigenvalue of a 200×200 matrix. Maximal correlation values between these increments were obtained by linear interpolation. Computations were implemented using Ox version 5.10; see Doornik (2007).

Figure 4.1 confirms what basic intuition would suggest: the maximal correlation coefficient for both the Clayton and the Gumbel copulas increases smoothly from zero to one as the tail dependence coefficient increases from zero to one. Consequently, Theorem 4.1 implies that stationary Markov chains constructed using a Clayton or Gumbel copula with $\theta < \infty$ will satisfy ρ - and α -mixing conditions at a geometric rate. Note that the curves in Figure 4.1 that plot maximal correlation never fall below the 45 degree line. This is because the upper and lower tail dependence coefficients can be written as the limit of

correlations between indicator functions, and are, therefore, bounded by the maximal correlation coefficient.

We are unable to give precise bounds on the accuracy of the numerical approximations used to generate Figure 4.1. In the case of the Gaussian copula with linear correlation parameter ρ ranging between 0.01 and 0.99 in intervals of 0.01, we found that the approximate maximal correlation obtained using a 200×200 matrix differed from the true maximal correlation (known to be equal to ρ) by no more than 0.001.

Our conclusion that Clayton and Gumbel copulas generate a geometric rate of ρ -mixing is at odds with a claim made by Ibragimov and Lentzas (2008), who argued that Clayton copulas generate a polynomial decay rate of the L_1 Schweizer–Wolff distance $\kappa_k = \|C_k(x, y) - xy\|_1$. Since $|C_k(x, y) - xy| \leq \alpha_k \leq (1/4)\rho_k$, a geometric rate of ρ -mixing implies that κ_k must decay at least geometrically fast and not at a polynomial rate. Ibragimov and Lentzas arrived at their conclusion by approximating κ_k numerically using a discretization method and then regressing the log of the approximate κ_k on $\log k$ and a constant. They found that the coefficient of $\log k$ is significantly negative and argued that this implies a polynomial decay rate of κ_k . However, in view of the slowly varying nature of $\log k$ as $k \rightarrow \infty$, one may expect the regression estimates reported by Ibragimov and Lentzas to be largely determined by the regression fit at small values of k , yet our interest lies with the behavior of κ_k as $k \rightarrow \infty$. A further concern is that positive bias in the approximation of κ_k when κ_k is close to zero may cause the rate of decay of κ_k to appear slower than is actually the case.

5. CONCLUSION

In this paper we have identified conditions under which a copula function generates a stationary Markov chain that satisfies mixing conditions at a geometric rate. In particular, for nonextreme parameter values, we have demonstrated that the Farlie–Gumbel–Morgenstern, Frank, and Gaussian copulas generate geometric rates of β - and ρ -mixing, and that the Marshall–Olkin, Clayton, Gumbel, and t -copulas generate a geometric rate of ρ -mixing. The conditions under which we obtain geometric ρ -mixing are substantially weaker than those under which we obtain geometric β -mixing, as well as being easily verifiable. Our results complement those of Chen, Wu, and Yi (2009), who established a geometric rate of β -mixing for Clayton, Gumbel, and t -copulas. Taken together, the results here and in that paper establish that, for a wide class of copula functions, copula-based Markov models exhibit dependence properties typical of short memory time series.

REFERENCES

- ANDREWS, D. W. K. (1984): “Non Strong Mixing Autoregressive Processes,” *Journal of Applied Probability*, 21, 930–934. [404]

- ARCONES, M. A. (1995): "On the Central Limit Theorem for U -Statistics Under Absolute Regularity," *Statistics and Probability Letters*, 24, 245–249. [403]
- BEARE, B. K. (2007): "A New Mixing Condition," Discussion Paper 348, Department of Economics, University of Oxford. [403]
- (2010): "Supplement to 'Copulas and Temporal Dependence,'" *Econometrica Supplemental Material*, 79, http://www.econometricsociety.org/ecta/Supmat/8152_Proofs.pdf. [397]
- BONHOMME, S., AND J.-M. ROBIN (2006): "Modelling Individual Earnings Trajectories Using Copulas: France, 1990–2002," in *Structural Models of Wage and Employment Dynamics*. Contributions to Economic Analysis, Vol. 275, ed. by H. Bunzel, B. J. Christensen, G. R. Neumann, and J.-M. Robin. Amsterdam: Elsevier, Chapter 18. [396]
- BRADLEY, R. C. (2007): *Introduction to Strong Mixing Conditions*, Vols. 1–3. Heber City: Kendrick Press. [396,398,399,403,404]
- BRYC, W. (1996): "Conditional Moment Representations for Dependent Random Variables," *Electronic Journal of Probability*, 1, 1–14. [404]
- CHEN, X., AND Y. FAN (2006): "Estimation of Copula-Based Semiparametric Time Series Models," *Journal of Econometrics*, 130, 307–335. [395,396]
- CHEN, X., L. P. HANSEN, AND M. CARRASCO (2008): "Nonlinearity and Temporal Dependence," Discussion Paper 1652, Cowles Foundation. [403]
- CHEN, X., W. B. WU, AND Y. YI (2009): "Efficient Estimation of Copula-Based Semiparametric Markov Models," *The Annals of Statistics*, 37, 4214–4253. [395,396,401,408]
- DARSOW, W. F., B. NGUYEN, AND E. T. OLSEN (1992): "Copulas and Markov Processes," *Illinois Journal of Mathematics*, 36, 600–642. [397]
- DEARDEN, L., E. FITZSIMONS, A. GOODMAN, AND G. KAPLAN (2008): "Higher Education Funding Reforms in England: The Distributional Effects and the Shifting Balance of Costs," *Economic Journal*, 118, F100–F125. [396]
- DOORNIK, J. A. (2007): *Object-Oriented Matrix Programming Using Ox* (Third Ed.). London: Timberlake Consultants Press. [407]
- DOUKHAN, P. (1994): *Mixing: Properties and Examples*. New York: Springer-Verlag. [396]
- FENTAW, A., AND U. V. NAIK-NIMBALKAR (2008): "Dynamic Copula-Based Markov Time Series," *Communications in Statistics: Theory and Methods*, 37, 2447–2460. [395,396]
- GAGLIARDINI, P., AND C. GOURIÉROUX (2007): "An Efficient Nonparametric Estimator for Models With Nonlinear Dependence," *Journal of Econometrics*, 137, 189–229. [395,402]
- (2008): "Duration Time-Series Models With Proportional Hazard," *Journal of Time Series Analysis*, 29, 74–124. [396,400,402]
- HAMDAN, M. A., AND H. A. AL-BAYYATI (1971): "Canonical Expansion of the Compound Correlated Bivariate Poisson Distribution," *Journal of the American Statistical Association*, 66, 390–393. [400]
- IBRAGIMOV, R. (2009): "Copula-Based Characterizations for Higher Order Markov Processes," *Econometric Theory*, 25, 819–846. [396]
- IBRAGIMOV, R., AND G. LENTZAS (2008): "Copulas and Long Memory," Discussion Paper 2160, Harvard Institute of Economic Research. [396,399,408]
- JOE, H. (1997): *Multivariate Models and Dependence Concepts*. London: Chapman & Hall. [395]
- KOLMOGOROV, A. N., AND Y. A. ROZANOV (1960): "On the Strong Mixing Conditions for Stationary Gaussian Sequences," *Theory of Probability and Its Applications*, 5, 204–207. [403]
- LANCASTER, H. O. (1958): "The Structure of Bivariate Distributions," *Annals of Mathematical Statistics*, 29, 719–736. [399]
- MCCAUSLAND, W. J. (2007): "Time Reversibility of Stationary Regular Finite-State Markov Chains," *Journal of Econometrics*, 136, 303–318. [400]
- MCNEIL, A. J., R. FREY, AND P. EMBRECHTS (2005): *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton University Press. [401,405,407]
- MEYN, S. P., AND R. L. TWEEDIE (1993): *Markov Chains and Stochastic Stability*. London: Springer-Verlag. [396]
- NELSEN, R. B. (1999): *An Introduction to Copulas*. New York: Springer-Verlag. [397,398,400–402]

- PATTON, A. J. (2008): "Copula-Based Models for Financial Time Series," in *Handbook of Financial Time Series*, ed. by T. G. Anderson, R. A. Davis, J.-P. Kreiss, and T. Mikosch. New York: Springer-Verlag. [396]
- RÉNYI, A. (1959): "On Measures of Dependence," *Acta Mathematica Academiae Scientiarum Hungaricae*, 10, 441–451. [398,399,405]
- ROSENBLATT, M. (1956): "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Sciences of the United States of America*, 42, 43–47. [403]
- _____. (1971): *Markov Processes: Structure and Asymptotic Behavior*. Berlin: Springer-Verlag. [403]
- SARMANOV, O. V. (1958a): "Maximum Correlation Coefficient (Symmetric Case)," *Doklady Akademii Nauk SSSR*, 120, 715–718 (in Russian). English translation: *Selected Translations in Mathematical Statistics and Probability*, 4, 271–275 (1963). [399]
- _____. (1958b): "Maximum Correlation Coefficient (Nonsymmetric Case)," *Doklady Akademii Nauk SSSR*, 121, 52–55 (in Russian). [399]
- _____. (1961): "Investigation of Stationary Markov Processes by the Method of Eigenfunction Expansion," *Trudy Matematicheskogo Instituta Imeni V. A. Steklova*, 60, 238–261 (in Russian). English translation: *Selected Translations in Mathematical Statistics and Probability*, 4, 245–269 (1963). [399]
- SCHWEIZER, B., AND A. SKLAR (1974): "Operations on Distribution Functions Not Derivable From Operations on Random Variables," *Studia Mathematica*, 52, 43–52. [397]
- SCHWEIZER, B., AND E. F. WOLFF (1981): "On Nonparametric Measures of Dependence for Random Variables," *The Annals of Statistics*, 9, 879–885. [399]
- SKLAR, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231 (in French). [397]
- VOLKONSKII, V. A., AND Y. A. ROZANOV (1959): "Some Limit Theorems for Random Functions, Part I," *Theory of Probability and Its Applications*, 4, 178–197. [398,403]

Dept. of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, U.S.A.; bbeare@ucsd.edu.

Manuscript received September, 2008; final revision received June, 2009.

ANNOUNCEMENTS

2010 WORLD CONGRESS OF THE ECONOMETRIC SOCIETY

THE TENTH WORLD CONGRESS of the Econometric Society will be held in Shanghai from August 17th to August 21th, 2010. It is hosted by Shanghai Jiao Tong University in cooperation with Shanghai University of Finance and Economics, Fudan University, China Europe International Business School, and the Chinese Association of Quantitative Economics.

The congress is open to all economists, including those who are not now members of the Econometric Society. It is hoped that papers presented at the Congress will represent a broad spectrum of applied and theoretical economics and econometrics.

The Program Co-Chairs are:

Professor Daron Acemoglu, MIT Department of Economics, E52-380B,
50 Memorial Drive, Cambridge, MA 02142-1347, U.S.A.

Professor Manuel Arellano, CEMFI, Casado del Alisal 5, 28014 Madrid,
Spain.

Professor Eddie Dekel, Department of Economics, Northwestern University,
2003 Sheridan Rd., Evanston, IL 60208-2600, U.S.A., and Eitan Berglas
School of Economics, Tel Aviv University, Tel Aviv 69978, Israel.

Submissions will be open from November 1st, 2009 and will be accepted only in electronic form at www.eswc2010.com. The deadline for such submissions will be January 30th, 2010. There will be financial assistance for young scholars to be allocated once the decisions on submitted papers have been made. At least one co-author must be a member of the Society or must join prior to submission. This can be done electronically at www.econometricsociety.org.

The Chair of the Local Organizing Committee is:

Professor Lin Zhou, Department of Economics, Shanghai Jiao Tong
University, Shanghai 200052, China, and Department of Economics,
Arizona State University, Tempe, AZ 85287, U.S.A.

Detailed information on registration and housing will be sent by email to all members of the Econometric Society in due course and will be available at www.eswc2010.com.

FORTHCOMING PAPERS

THE FOLLOWING MANUSCRIPTS, in addition to those listed in previous issues, have been accepted for publication in forthcoming issues of *Econometrica*.

- BARTOLUCCI, FRANCESCO, AND VALENTINA NIGRO: "A Dynamic Model for Binary Panel Data With Unobserved Heterogeneity Admitting a Root- n Consistent Conditional Estimator."
- BERGEMANN, DIRK, AND JUUSO VÄLIMÄKI: "The Dynamic Pivot Mechanism."
- BUGNI, FEDERICO A.: "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set."
- CHERNOZHUKOV, VICTOR, IVÁN FERNÁNDEZ-VAL, AND ALFRED GALICHON: "Quantile and Probability Curves Without Crossing."
- CHESHER, ANDREW: "Instrumental Variable Models for Discrete Outcomes."
- CUNHA, FLAVIO, JAMES J. HECKMAN, AND SUSANNE M. SCHENNACH: "Estimating the Technology of Cognitive and Noncognitive Skill Formation."
- EECKHOUT, JAN, AND PHILIPP KIRCHER: "Sorting and Decentralized Price Competition."
- EINAV, LIRAN, AMY FINKELSTEIN, AND PAUL SCHRIMPFF: "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence From the U.K. Annuity Market."
- GONZALEZ, FRANCISCO M., AND SHOUYONG SHI: "An Equilibrium Theory of Learning, Search and Wages."
- KUERSTEINER, GUIDO, AND RYO OKUI: "Constructing Optimal Instruments by First Stage Prediction Averaging."
- MARTINS-DA-ROCHA, V. FILIPE, AND YIANNIS VAILAKIS: "Existence and Uniqueness of a Fixed-Point for Local Contractions."
- STRULOVICI, BRUNO: "Learning While Voting: Determinants of Collective Experimentation."
- WOODERS, JOHN: "Does Experience Teach? Professionals and Minimax Play in the Lab."

THE ECONOMETRIC SOCIETY ANNUAL REPORTS
REPORT OF THE SECRETARY

BARCELONA, SPAIN
AUGUST 22–23, 2009

1. MEMBERSHIP AND CIRCULATION

THIS REPORT STARTS by describing the evolution of the Society's membership and of the number of institutional subscribers. Information is provided on both a midyear and an end-of-year basis. The latest information available, as of June 30 of the current year and of selected previous years, is provided in the top panel of Table I. The bottom panel of Table I reports the final number of members and subscribers as of the end of 2008 and selected previous years. For any given year the figures in the bottom half of Table I are larger than in the top half, reflecting those memberships and subscriptions that are initiated between the middle and the end of that calendar year.

The membership of the Society seems to have stabilized after five years of very high growth. At the end of 2008 there were a total of 5,714 members, which represents an increase of 34.3 percent with respect to the 2003 figure. The increase has been particularly strong in ordinary members. In contrast, in the last three years there has been a decrease in the number of student members that has been more than compensated by the increase in ordinary members.

The number of institutional subscribers has continued its downward trend, reaching a total of 1,786 subscribers at the end of 2008, which represents a decrease of 3.0 percent with respect to the figure in 2007, and of 19.5 percent with respect to the one in 2003. However, the latest data indicates that this decline may not continue in 2009.

Table II displays the division between print and online and online only memberships and subscriptions. Since the choice between these two alternatives was offered in 2004, there has been a continued shift toward online only. This is especially significant for student members, 82.8 percent of whom chose this option as of June 2009, but the shift is also very significant for ordinary members, for whom the proportion of online only crossed the 50 percent threshold this year, and it is also noticeable in institutional subscriptions, for which the proportion of online only went up from 17.9 percent in June 2008 to 24.6 percent in June 2009.

Table III compares the Society's membership and the number of institutional subscribers with those of the American Economic Association. (For the membership category these figures include ordinary, student, free, and life members for both the ES and the AEA.) The increase in ES membership and the decline in AEA membership seem to have stabilized in the last year, so the ES/AEA ratio for members in 2008 is very close to the record 34.1 percent reached in

TABLE I
INSTITUTIONAL SUBSCRIBERS AND MEMBERS

Year	Members						Total Circulation
	Institutions	Ordinary	Student	Soft Currency	Free ^a	Life	
<i>1. Institutional subscribers and members at the middle of the year</i>							
1980	2,829	1,978	411	53	45	74	5,390
1985	2,428	2,316	536	28	55	71	5,434
1990	2,482	2,571	388	57	73	69	5,643
1995	2,469	2,624	603	46	77	66	5,885
2000	2,277	2,563	437	—	112	62	5,471
2001	2,222	2,456	363	—	71	62	5,174
2002	2,109	2,419	461	—	103	61	5,153
2003	1,971	2,839	633	—	117	60	5,620
2004	1,995	2,965	784	—	111	60	5,915
2005	1,832	3,996	1,094	—	106	57	7,085
2006	1,776	4,020	1,020	—	110	58	6,984
2007	1,786	4,393	916	—	97	58	7,250
2008	1,691	4,257	759	—	89	56	6,852
2009	1,686	4,268	744	—	81	56	6,835
<i>2. Institutional subscribers and members at the end of the year</i>							
1980	3,063	2,294	491	49	47	74	6,018
1985	2,646	2,589	704	53	61	70	6,123
1990	2,636	3,240	530	60	74	68	6,608
1995	2,569	3,072	805	43	96	66	6,651
2000	2,438	3,091	648	—	77	62	6,316
2001	2,314	3,094	680	—	87	61	6,233
2002	2,221	3,103	758	—	105	60	6,247
2003	2,218	3,360	836	—	112	60	6,586
2004	2,029	3,810	1,097	—	101	58	7,095
2005	1,949	4,282	1,222	—	110	58	7,621
2006	1,931	4,382	1,165	—	93	58	7,629
2007	1,842	4,691	1,019	—	86	56	7,694
2008	1,786	4,742	916	—	89	56	7,589

^aIncludes free libraries.

2007. At the same time, the long-run proportional decline in the number of institutional subscribers has been similar for both organizations.

The geographic distribution of members (including students) by countries and regions as of June 30 of the current year and of selected previous years is shown in Table IV. The format of this table was slightly changed in 2008, and it now shows individual data on countries with more than 10 members. Previously some countries were grouped together, so their individual membership data is not available. In comparison with the 2008 figures, the membership has significantly increased in Brazil and Japan, which is probably explained by the

TABLE II
INSTITUTIONAL SUBSCRIBERS AND MEMBERS BY TYPE OF SUBSCRIPTION (MIDYEAR)

	2008		2009	
	Total	Percent	Total	Percent
Institutions	1,691	100.0	1,686	100.0
Print + Online	1,388	82.1	1,271	75.4
Online only	303	17.9	415	24.6
Ordinary members	4,257	100.0	4,268	4,268
Print + Online	2,233	52.5	1,902	44.6
Online only	2,024	47.5	2,366	55.4
Student members	759	100.0	744	100.0
Print + Online	191	25.2	128	17.2
Online only	568	74.8	616	82.8

TABLE III
INSTITUTIONAL SUBSCRIBERS AND MEMBERS ECONOMETRIC SOCIETY AND AMERICAN
ECONOMIC ASSOCIATION (END OF YEAR)

Year	Institutions			Members		
	ES	AEA	ES/AEA (%)	ES	AEA	ES/AEA (%)
1975	3,207	7,223	44.4	2,627	19,564	13.4
1980	3,063	7,094	43.2	2,955	19,401	15.2
1985	2,646	5,852	45.2	3,416	20,606	16.0
1990	2,636	5,785	45.6	3,972	21,578	18.4
1995	2,569	5,384	47.7	4,082	21,565	18.9
2000	2,438	4,780	50.8	3,878	19,668	19.7
2001	2,314	4,838	47.8	3,919	18,761	20.9
2002	2,221	4,712	47.1	4,026	18,698	21.5
2003	2,218	4,482	49.5	4,368	19,172	22.8
2004	2,029	4,328	46.9	5,066	18,908	26.8
2005	1,949	4,234	46.0	5,672	18,067	31.4
2006	1,931	3,945	48.9	5,698	17,811	32.0
2007	1,842	3,910	47.1	5,852	17,143	34.1
2008	1,786	3,726	47.9	5,803	17,096	33.9

organization of the 2008 Latin American Meeting in Rio de Janeiro and the 2009 Far East and South Asian Meeting in Tokyo.

Table V shows the percentage distribution of members (including students) by regions as of June 30 of the current year and of selected previous years. The share of North America in total membership fell for the first time in 2009 below that of Europe and Other Areas, and it is now at 40.3 percent.

TABLE IV
GEOGRAPHIC DISTRIBUTION OF MEMBERS^a (MIDYEAR)

Region and Country	1980	1985	1990	1995	2000	2005	2008	2009
<i>Australasia</i>	57	60	95	98	90	162	201	209
Australia	52	57	84	88	78	137	167	182
New Zealand	5	3	11	10	12	25	34	27
<i>Europe and Other Areas</i>	625	716	803	1,031	992	2,092	2,106	2,067
Austria	15	21	25	27	24	49	44	38
Belgium	23	21	30	31	32	61	41	44
Czech Republic	—	—	—	—	—	—	17	11
Denmark	19	22	27	38	22	47	43	42
Finland	19	26	17	15	13	27	48	43
France ^b	53	36	56	81	73	188	206	186
Germany	92	106	112	135	153	354	390	399
Greece ^c	12	12	6	14	15	18	22	28
Hungary	34	30	30	5	5	13	19	16
Ireland	4	5	5	6	6	15	18	18
Israel	0	16	25	32	37	56	38	36
Italy ^d	16	43	48	57	59	126	167	158
Netherlands	75	68	90	103	86	130	151	148
Norway	24	26	23	29	21	52	51	40
Poland	4	6	20	27	27	22	18	13
Portugal	5	5	11	11	19	32	33	38
Russia ^e	5	2	4	4	5	11	13	9
Spain	34	43	36	88	81	171	184	204
Sweden	27	31	25	45	42	72	57	47
Switzerland	26	27	25	34	25	79	91	90
Turkey	0	1	3	8	9	21	12	21
United Kingdom	135	145	162	210	207	509	398	385
Other Europe	3	6	10	17	19	23	29	33
Other Asia	0	4	2	5	7	6	5	6
Other Africa	0	14	11	9	5	10	11	14
<i>Far East</i>	105	134	144	228	189	315	391	459
China	—	—	—	—	—	—	25	28
Hong Kong ^f	—	—	—	—	—	—	27	28
Japan	83	114	101	143	130	203	248	316
Korea	—	—	—	—	—	—	47	45
Taiwan	—	—	—	—	—	—	43	41
Other Far East	22	20	43	85	59	112	1	1
<i>North America</i>	1,645	2,059	2,150	1,989	1,498	2,409	2,187	2,058
Canada	159	192	194	200	127	208	226	227
United States	1,486	1,867	1,956	1,789	1,371	2,201	1,961	1,831

(Continues)

TABLE IV—*Continued*

Region and Country	1980	1985	1990	1995	2000	2005	2008	2009
<i>Latin America</i>	53	39	30	87	105	180	162	233
Argentina	—	—	—	—	—	—	21	20
Brazil	—	—	—	—	—	—	69	101
Chile	—	—	—	—	—	—	21	35
Colombia	—	—	—	—	—	—	13	15
Mexico	10	5	1	16	15	33	25	40
Other Latin America	43	34	29	71	90	147	13	22
<i>South and South East Asia</i>	27	49	42	49	31	105	74	76
India	6	30	18	10	14	22	22	21
Singapore	—	—	—	—	—	—	36	41
Other South and South East Asia ^f	21	19	24	39	17	83	16	14
Total	2,512	3,057	3,264	3,482	2,905	5,263	5,121	5,102

^aOnly countries with more than 10 members in 2008 are listed individually. Until 2005 some countries were grouped together, so their individual membership data is not available.

^bUntil 2005 the data for France includes Luxembourg.

^cUntil 2005 the data for Greece includes Cyprus.

^dUntil 2005 the data for Italy includes Malta.

^eUntil 2005 the data for Russia corresponds to the Commonwealth of Independent States or the USSR.

^fUntil 2005 Hong Kong was included in South and South East Asia.

TABLE V
PERCENTAGE DISTRIBUTION OF MEMBERS (MIDYEAR)

	1980	1985	1990	1995	2000	2005	2008	2009
Australasia	2.3	2.0	2.9	2.8	3.1	3.1	3.9	4.1
Europe and Other Areas	24.9	23.4	24.6	29.6	34.1	39.7	41.1	40.5
Far East	4.2	4.4	4.4	6.5	6.5	6.0	7.6	9.0
North America	65.5	67.4	65.9	57.1	51.6	45.8	42.7	40.3
Latin America	2.1	1.3	0.9	2.5	3.6	3.4	3.2	4.6
South and Southeast Asia	1.1	1.6	1.3	1.4	1.1	2.0	1.4	1.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

2. FELLOWS

Table VI displays the geographic distribution of Fellows as of June 30, 2009. As noted in previous reports, this distribution is very skewed, with 69.5 percent of the Fellows based in North America and 26.2 percent in Europe and Other Areas. It should also be noted that the ratio of Fellows to members fell below 1% in 2009 in Latin America.

Table VII provides information on the nomination and election of Fellows. Since 2006, the election has been conducted with an electronic ballot system.

TABLE VI
GEOGRAPHIC DISTRIBUTION OF FELLOWS, 2008

<i>Australasia</i>	6	Switzerland	3
Australia	6	Turkey	1
		United Kingdom	51
<i>Europe and Other Areas</i>	163		
Austria	2	<i>Far East</i>	16
Belgium	9	Japan	15
Denmark	2	Korea	1
Finland	2		
France	31	<i>North America</i>	433
Germany	9	Canada	10
Hungary	6	United States	423
Israel	21		
Italy	4	<i>Latin America</i>	2
Netherlands	6	Brazil	2
Norway	1		
Poland	2	<i>South and Southeast Asia</i>	3
Russia	4	India	3
Spain	6		
Sweden	3	Total (as of June 30, 2009)	623

TABLE VII
FELLOWS' VOTING STATISTICS

Year	Total Fellows	Inactive	Eligible to Vote	Returned Ballots	Percent Returning Ballots		Number of Nominees	Number Elected	Percent Elected to Nominee	Late Ballots Returned but Not Counted
					Returning	Ballots				
1975	197	26	171	100	58.5		63	21	33.3	n.a.
1980	299	49	251	150	59.8		73	18	24.7	n.a.
1985	354	57	301	164	54.4		60	13	21.7	17
1990	422	47	375	209	55.7		44	23	52.3	5
1995	499	119	380	225	59.2		52	15	28.8	2
2000	546	147	399	217	54.4		59	14	23.7	10
2001	564	170	394	245	62.2		55	10	18.2	0
2002	577	189	388	236	60.8		45	17	37.8	2
2003	590	200	390	217	55.6		53	20	37.7	10
2004	582	145	437	239	54.7		51	15	29.4	8
2005	604	140	464	211	45.5		50	14	28.0	16
2006	601	154	447	325	72.7		55	5	9.1	—
2007	599	166	433	305	70.4		50	16	32.0	—
2008	610	163	447	310	69.4		61	15	24.6	—

This has led to a very significant increase in the participation rate, which averaged 70.8 percent in the last three elections compared to an average of 55.8 percent in the previous five elections and a historical minimum of 45.5 percent in 2005. The number of nominees in 2008 was 61 and the number of new Fellows elected was 15. This outcome is very similar to the one in 2007, when 16 new Fellows were elected. In this respect, it is important to note the change in the electronic ballot agreed by the Executive Committee in 2006 and implemented in 2007, which added the possibility of selecting by a single click all the candidates nominated by the Nominating Committee. In fact, of the 15 new Fellows elected in 2008, 11 had been nominated by the Committee (which also nominated 7 other candidates that were not elected). Comparing the 2007 with the 2008 election, the number of votes needed to be elected (30 percent of the number of ballots submitted) went up from 90 to 92, while the average number of votes per ballot went up from 12.8 to 14.2.

3. NEW ECONOMETRIC SOCIETY JOURNALS

In 2005, following a report written by Lars Peter Hansen, Matthew Jackson, David Levine (chair), and Robert Porter, the Executive Committee started discussing the possibility of launching new open access journals. The report stated three reasons for creating new journals: “The growing supply of high quality manuscripts, problems with existing journals, and enabling the Society to experiment with alternative forms of distribution.” Subsequent discussions, which included an online forum, led to a second report, written in 2007, by Richard Blundell, Lars Peter Hansen, and Torsten Persson. This report proposed launching two journals, “one with a base in economic theory and its applications, and another with a base in quantitative methods and applications broadly defined.”

In 2007, the Executive Committee agreed (i) to initiate the process of starting these two journals, (ii) to discuss with the Editors of *Theoretical Economics*, an open access journal launched in 2006, the possibility of becoming an Econometric Society journal, and (iii) to appoint a committee with Manuel Arellano, Orazio Attanasio, Steven Durlauf, Robert Porter, Jean-Marc Robin (chair), and Thomas Sargent to further look into the purpose and scope of the quantitative economics journal (all these reports are available at the Society’s website).

Following this report, the Executive Committee decided in 2008 to prepare a formal proposal to be considered by the Council, in its capacity as the highest decision-making body of the Society. It was the unanimous view of the Executive Committee that the new journals would complement the Society’s efforts to broaden its impact and promote innovative research in theoretical and quantitative economics, in a manner consistent with Ragnar Frisch’s original vision of the Society.

The proposal was approved by the Council in January 2009, with 27 members voting in favor and none against. The Society's Constitution states that "The Fellows represent the highest authority of the Society. The Council shall consult them on any vital question that affects the policy of the Society, obtaining their decision by mail vote." In accordance with this provision, a ballot was sent to the Fellows. 277 active Fellows (66.4% of the total) cast their vote, of whom 240 (86.6%) were in favor of the proposal, 30 (10.8%) were against, and 7 (2.5%) abstained.

Starting in 2010, the Econometric Society will publish two new journals, called *Quantitative Economics* and *Theoretical Economics*. They will be open access journals, so electronic access will be free for both members and non-members of the Society. Each journal will initially publish three issues per year. In addition to the electronic publication, they will also be printed. Each journal will have its own Editorial Board, with an Editor, several Co-Editors, and a set of Associate Editors. The rules for appointment of the members of the Editorial Board will be identical to those used for *Econometrica*. It is expected that each journal will develop its own identity, separate of *Econometrica*. While the new journals will have very high quality requirements, it is anticipated that they may experiment more with publishing papers in new research areas.

Quantitative Economics (*QE*) will cover quantitative economics in a broad sense, including econometric theory, computational methods, and empirical applications based on structural or reduced-form estimation, computation or simulation. Applications will be welcome in all fields, such as e.g., finance, macroeconomics, labor economics, industrial organization, and development economics, as well as quantitative research on economic dynamics. The journal will emphasize empirical orientation. The first Editor of *QE* will be Orazio Attanasio, with Steven Durlauf, Victor Rios-Rull, and Elie Tamer serving as Co-Editors.

Theoretical Economics (*TE*) will publish both pure and applied theoretical research in economics. All fields of economic theory will be covered. In particular, the journal will publish theoretical research in microeconomics, macroeconomics, behavioral economics, industrial organization, finance, labor economics, public economics, political economy, urban economics, development economics, and growth. Papers may contain empirical and experimental results, but need to have a substantial and innovative theoretical component to be published. An agreement between the Econometric Society and the Society for Economic Theory, which launched *TE* in 2006, about the terms of adoption of the journal was signed in February 2009. The first Editor of *TE* will be Martin Osborne, the Managing Editor of the existing journal, with Jeffrey Ely, Edward Green, Narayana Kocherlakota, Barton Lipman, and Debraj Ray serving as Co-Editors.

4. REGIONAL MEETINGS AND WORLD CONGRESS

In 2009, all six regions of the Society are organizing meetings, according to the following timetable:

North American Winter Meeting, San Francisco, California, January 3–5, 2009

North American Summer Meeting, Boston, Massachusetts, June 4–7, 2009

Australasian Meeting, Canberra, Australia, July 7–10, 2009

Far East and South Asian Meeting, Tokyo, Japan, August 3–5, 2009

European Summer Meeting, Barcelona, Spain, August 23–27, 2009

Latin American Meeting, Buenos Aires, Argentina, October 1–3, 2009

European Winter Meeting, Budapest, Hungary, November 12–13, 2009

The North American Winter Meetings have traditionally taken place within the meetings of the Allied Social Sciences Association (ASSA). Since 2003, the European Summer Meeting has run in parallel with the Annual Congress of the European Economic Association, since 2006, the Latin Amerian Meeting has run in parallel with the Annual Meeting of the Latin American and Caribbean Economic Association (LACEA), and since 2008 there has been a joint Far East and South Asian Meeting.

The 2010 World Congress will take place in Shanghai, China, August 17–21. The Congress will take place in the Shanghai International Convention Center, and it will be organized by the Shanghai Jiao Tong University in cooperation with the Shanghai University of Finance and Economics, Fudan University, the China Europe International Business School, and the Chinese Association of Quantitative Economics. The Local Arrangements Chair is Lin Zhou, and the Program Chairs are Daron Acemoglu, Manuel Arellano, and Eddie Dekel. The Plenary Lectures will be given by Orazio Attanasio (Walras-Bowley), Drew Fudenberg (Fisher-Schultz), Elhanan Helpman (Frisch Memorial), Whitney Newey (Shanghai), and John Moore (Presidential Address).

5. A FINAL NOTE

To conclude, I would like to thank the members of the Executive Committee, and in particular Torsten Persson, for their help and support during 2008. I am also very grateful to Claire Sashi, the Society's General Manager in charge of the office at New York University, for her excellent work during this year.

RAFAEL REPULLO

THE ECONOMETRIC SOCIETY REPORTS
REPORT OF THE TREASURER

BARCELONA, SPAIN
AUGUST 22–23, 2009

1. 2008 ACCOUNTS

THE 2008 ACCOUNTS of the Econometric Society show a deficit of \$315,638 (Table III, Line G). The deficit is significantly larger than the estimate of \$250,000 at this time last year. This is due to the heavy losses in the Society's investment portfolio during the second half of 2008, which were partly compensated by a higher than expected operational surplus. In fact, membership and subscriptions plus other revenues were \$259,493 higher than expected (Table II, Lines A and B), while total expenses were \$85,156 lower than expected (Table III, Line F).

The net worth of the Society on 12/31/2008 was down to \$1,109,581 (Table I, Line C). Consequently, the ratio of net worth to total expenses on 12/31/2008 was 113 percent, a figure which is below the target range between 120 and 160 percent agreed by the Executive Committee in August 2007.

Table I shows the balance sheets of the Society for the years 2004–2008, distinguishing between unrestricted assets and liabilities, whose difference gives the Society's net worth, and five restricted accounts: The World Congress Fund, which is a purely bookkeeping entry that serves to smooth the expenses every five years on travel grants to the World Congress, the Jacob Marschak Fund, devoted to support the Marschak lectures at regional meetings outside Europe and North America, and the Far Eastern, Latin American, and European Funds, which are held in custody for the convenience of the corresponding Regional Standing Committees. Tables IV and V show the movements in the World Congress Fund and the other restricted accounts for the years 2004–2008.

Table II shows the actual revenues for 2007, the estimated and actual revenues for 2008, and the estimated revenues for 2009 and 2010. The losses in the investment portfolio during 2008 were \$560,286 (Line C). The situation is likely to be very different in 2009, due to the rise in stock markets since the beginning of the year, with investment income estimated to be of the order of \$200,000. Thus total revenues are expected to increase to \$1,316,000 (Line D), which implies a 97 percent increase relative to the figure in 2008. The budget for 2010 incorporates a significant increase in membership and subscription revenues, due to the decisions on rates agreed by the Executive Committee.

Table III shows the actual expenses for 2007, the estimated and actual expenses for 2008, and the estimated expenses for 2009 and 2010. Publishing expenses in 2008 have been significantly lower than the estimate at this time last year, while all other expenses have been in line with the estimate. Total

TABLE I
ECONOMETRIC SOCIETY BALANCE SHEETS, 2004–2008

	12/31/04 \$	12/31/05 \$	12/31/06 \$	12/31/07 \$	12/31/08 \$
A. Unrestricted Assets	1,771,179	1,707,036	1,801,710	2,203,312	1,896,510
1. Short Term Assets	453,857	123,106	63,854	117,574	96,245
2. Investments	855,848	1,114,981	1,548,878	1,753,807	1,428,604
3. Accounts Receivable	436,678	450,198	167,360	245,755	266,921
4. Back Issue Inventory	7,285	7,067	1,884	7,913	26,587
5. Furniture and Equipment	10,448	5,791	2,459	3,161	1,545
6. Other Assets	7,063	5,893	17,275	75,102	76,608
B. Unrestricted Liabilities	882,989	755,569	554,504	778,093	786,929
1. Accounts Payable	19,470	68,293	37,861	99,103	117,257
2. Deferred Revenue	563,519	607,276	356,643	438,990	349,672
3. World Congress Fund	300,000	80,000	160,000	240,000	320,000
C. Unrestricted Fund Balance	888,190	951,467	1,247,206	1,425,219	1,109,581
D. World Congress Fund Balance	300,000	80,000	160,000	240,000	320,000
E. Jacob Marschak Fund Balance	27,876	29,011	26,560	24,926	21,649
F. Far Eastern Fund Balance	61,710	63,576	66,624	70,016	68,047
G. Latin American Fund Balance	14,489	22,046	23,103	21,941	22,577
H. European Fund Balance	—	—	—	64,903	38,011

TABLE II
ECONOMETRIC SOCIETY REVENUES, 2007–2010

	Actual 2007 \$	Estimate 2008 \$	Actual 2008 \$	Estimate 2009 \$	Budget 2010 \$
A. Membership and Subscriptions	966,345	920,000	1,138,648	1,050,000	1,250,000
B. Other Revenues	48,651	50,000	90,845	66,000	60,000
1. Back Issues	15,660	15,000	53,508	30,000	30,000
2. Reprints	831	1,000	3,165	3,000	3,000
3. Advertising	2,518	7,000	5,549	5,000	5,000
4. List Rentals	1,366	2,000	1,846	2,000	2,000
5. Permissions	12,085	10,000	5,954	6,000	6,000
6. North American Meetings	16,191	15,000	20,823	20,000	14,000
C. Investment Income	167,229	(150,000)	(560,286)	200,000	120,000
1. Interest and Dividends	64,998	50,000	52,982	40,000	40,000
2. Capital Gains (Losses)	102,231	(200,000)	(613,268)	160,000	80,000
D. Total Revenues	1,182,225	820,000	669,206	1,316,000	1,430,000

TABLE III
ECONOMETRIC SOCIETY EXPENSES, 2007–2010

	Actual 2007	Estimate 2008	Actual 2008	Estimate 2009	Budget 2010
	\$	\$	\$	\$	\$
A. Publishing	372,450	360,000	270,567	290,000	370,000
1. Composition	71,226	60,000	49,323	55,000	80,000
2. Printing	67,907	70,000	61,347	70,000	105,000
3. Inventory (net)	(6,029)	0	7,913	0	0
4. Circulation	102,556	100,000	98,625	100,000	100,000
5. Postage	136,790	130,000	53,359	65,000	85,000
B. Editorial	278,828	339,000	333,745	408,000	484,000
1. Editors	173,875	232,000	231,540	299,000	355,000
2. Editorial Assistants	92,637	92,000	89,940	95,000	105,000
3. Software	3,000	3,000	6,000	6,000	8,000
4. Meetings	9,316	12,000	6,265	8,000	16,000
C. Administrative	206,964	195,000	200,650	203,000	204,000
1. Salaries and Honoraria	131,598	136,000	140,110	140,000	140,000
2. Administrative Support	10,000	10,000	10,000	10,000	10,000
3. Accounting and Auditing	38,040	34,000	33,870	34,000	32,000
4. Office	4,467	6,000	6,328	6,000	6,000
5. Website	23,494	8,000	10,086	12,000	15,000
6. IRS	(635)	1,000	256	1,000	1,000
D. Executive Committee	54,479	52,000	51,770	52,000	52,000
E. Meetings	91,489	124,000	128,111	128,000	115,000
1. World Congress	80,000	80,000	80,500	80,000	115,000
2. Regional Meetings	11,489	44,000	47,611	48,000	0
F. Total Expenses	1,004,210	1,070,000	984,844	1,081,000	1,225,000
G. Surplus	178,015	(250,000)	(315,638)	235,000	205,000
H. Unrestricted Fund Balance	1,425,219	1,175,219	1,109,581	1,344,581	1,549,581
I. Ratio of Unrestricted Fund Balance to Total Expenses	1.42	1.10	1.13	1.24	1.26

expenses for 2009 and 2010 are expected to increase by 9.8 and 13.3 percent, respectively, due to the launch of *Quantitative Economics* and *Theoretical Economics*, the two new Econometric Society journals. This together with the expected behaviour of total revenues implies an estimated surplus of \$235,000 for 2009 and of \$205,000 for 2010, after allocating \$80,000 and \$115,000, respectively, to the World Congress Fund. Thus the ratio of net worth to total expenses is expected to go up to 124 percent in 2009 and to 126 percent in 2010, within the target range agreed by the Executive Committee in August 2007.

The 2008 financial statements have been compiled by David Ciciyavili, 42 Vista Drive, Morganville, NJ 07751, and will be audited by Rothstein, Kass & Company, 1350 Avenue of the Americas, New York, NY 10019.

TABLE IV
WORLD CONGRESS FUND, 2004–2008

	2004 \$	2005 \$	2006 \$	2007 \$	2008 \$
A. Income	60,000	180,000	80,000	80,000	80,000
1. Transfer from General Fund	60,000	180,000	80,000	80,000	80,000
B. Expenses	0	400,000	0	0	0
1. Travel Grants	0	326,385	0	0	0
2. Transfer to General Fund	0	73,615	0	0	0
C. Fund Balance	300,000	80,000	160,000	240,000	320,000

TABLE V
RESTRICTED ACCOUNTS, 2004–2008

	2004 \$	2005 \$	2006 \$	2007 \$	2008 \$
A. Jacob Marschak Fund					
1. Investment Income	387	1,135	1,393	1,393	723
2. Expenses	0	0	3,844	3,029	4,000
3. Fund Balance	27,877	29,012	26,561	24,926	21,649
B. Far Eastern Fund					
1. Investment Income	674	1,866	3,048	3,392	2,031
2. Expenses	0	0	0	0	4,000
3. Fund Balance	61,710	63,576	66,624	70,016	68,047
C. Latin American Fund					
1. Investment Income	197	558	1,057	1,162	636
2. Expenses (net)	10,000	(7,000)	0	2,324	0
3. Fund Balance	14,489	22,046	23,103	21,941	22,577
D. European Fund					
1. Transfer from European Region	—	—	—	62,612	0
2. Investment Income	—	—	—	2,291	−26,892
3. Fund Balance	—	—	—	64,903	38,011

2. MEMBERSHIP AND INSTITUTIONAL SUBSCRIPTION RATES

The Executive Committee decided in August 2008 to form a Committee with Tim Besley, Eddie Dekel (chair), Roger Myerson, and Rafael Repullo to think about alternative pricing models for the Society. The Executive Committee agreed by e-mail in May 2009 to a proposal of the Pricing Committee on institutional subscription rates for 2010. The proposal consists of moving from a two-tier to a three-tier scheme based on the World Bank classification of countries. In particular, there will be a high income tier (the same as the current one), and the concessionary tier will be divided into a middle income and a low

income tier. The latter comprises those economies classified as low income by the World Bank plus the International Development Association (IDA) countries, while the former comprises the countries that are neither high nor low income according to this definition.

Income classifications are set by the World Bank each year on July 1. In the latest classification high (low) income economies are those with 2008 gross national income per capita (calculated using the World Bank Atlas method) higher (lower) than \$11,906 (\$975). IDA countries are those that had a per capita income in 2008 of less than \$1,135 and lack the financial ability to borrow from the International Bank for Reconstruction and Development (IBRD).

The adjusted institutional subscription rates for 2010, together with those for 2009, are the following:

	<u>2009</u>	<u>2010</u>
High income		
Print + Online	\$550	\$650
Online only	\$500	\$500
Middle income		
Print + Online	\$50	\$175
Online only	Free	\$125
Low income		
Print + Online	\$50	\$60
Online only	Free	\$10

Print + Online subscribers receive hard copies of the three Econometric Society journals (*Econometrica*, *Quantitative Economics* and *Theoretical Economics*) for the corresponding year, and have free online access to volumes of *Econometrica* back to 1999 (the other two journals are open access). Online only subscribers do not get the hard copies of the journals. Since 2006, institutional subscribers to *Econometrica* have perpetual online access to the volumes to which they subscribed.

The Pricing Committee proposed the following adjustment in individual membership rates for 2010 (agreed by the Executive Committee):

	<u>2009</u>	<u>2010</u>
Ordinary member (High income)		
Print + Online	\$60	\$90
Online only	\$25	\$50
Ordinary member (Middle and low income)		
Print + Online	\$45	\$50
Online only	\$10	\$10
Student member		
Print + Online	\$45	\$50
Online only	\$10	\$10

Members that choose the Print + Online option receive hard copies of the three Econometric Society journals (*Econometrica*, *Quantitative Economics* and *Theoretical Economics*) for the corresponding year, and have free online access to volumes of *Econometrica* back to 1933.

In addition, and in order to encourage 3-year memberships, the Executive Committee agreed a 20 percent discount on 3-year rates.

3. INVESTMENTS

The Society's Investments Committee consists of the Executive Vice-President and two Fellows appointed by the Executive Committee for a term of three years that can be renewed once. The Executive Committee decided in August 2008 to reappoint John Campbell and Hyun Shin for a second term starting on January 1, 2009.

In the light of market developments, the Investments Committee kept during 2008 a portfolio allocation that underweighted equities by about 5 percentage points relative to the reference asset allocation of 20 percent cash, 10 percent bonds, and 70 percent equities (of which 45 percent correspond to US equities, 45 percent to international equities, and 10 percent to emerging market equities). Partial rebalancing toward the reference asset allocation was done in August and October 2008. The return of the unrestricted portfolio in the year ending December 31, 2008 was -27.33 percent, as compared to the return of the S&P 500 stock market index of -38.49 percent.

During the first seven months of 2009, the Committee has allowed the portfolio allocation to move back to the reference allocation. On 7/31/2009, the breakdown by type of asset was 18.8 percent cash, 9.2 percent bonds, 31.8 percent US equities, 32.4 percent international equities, and 7.8 percent emerging

TABLE VI
ECONOMETRIC SOCIETY INVESTMENT PORTFOLIO

Name of Fund	Market Value 7/31/2008		Market Value 12/31/2008		Market Value 7/31/2009	
	\$	%	\$	%	\$	%
<i>Unrestricted Investment Portfolio</i>						
Fidelity Money Market	587,956	30.2	292,694	20.5	344,279	18.8
Spartan Interm. Treasury Bond	133,328	6.9	150,190	10.5	167,700	9.2
Spartan 500 Index	563,438	29.0	441,616	30.9	581,480	31.8
Spartan International Index	560,915	28.8	458,134	32.1	592,009	32.4
Fidelity Emerging Markets	100,418	5.2	85,969	6.0	141,942	7.8
<i>Restricted Investment Portfolio</i>						
Fidelity Money Market	119,000	68.0	116,273	75.4	116,901	72.7
Spartan International Index	55,979	32.0	38,011	24.6	43,979	27.3
<i>Total Investment Portfolio</i>	2,121,034		1,582,888		1,988,290	

markets equities (Table VI, Column 3). The return of the unrestricted portfolio in the first seven months of 2009 was 10.88 percent, as compared to the return of the S&P 500 stock market index of 9.33 percent. The return of the unrestricted portfolio in the year ending July 31, 2009 was -11.59 percent, as compared to the return of the S&P 500 of -22.08 percent.

RAFAEL REPULLO

THE ECONOMETRIC SOCIETY ANNUAL REPORTS
REPORT OF THE EDITORS 2008–2009

THE THREE TABLES BELOW provide summary statistics on the editorial process in the form presented in previous editors' reports.

Table I indicates that we received 672 new submissions this year. This number represents a small drop from last year. However, submissions have been rising very fast in recent years, and this number is close to the average of the last four years. The number of revisions received (188) is the highest ever and the number of accepted papers (59) is one of the highest numbers in recent years.

Table III gives data on the time to first decision for decisions made in this reporting year, with 47% of papers decided within three months and 88% decided within six months. This is an improvement on recent years and the positive trend is illustrated by the fact that the corresponding numbers for decisions in the first half of 2009 were 49% within three months and 94% within six months. Revisions were also faster with 55% within three months and 88% within six months. Although not reported in the tables, we can report that papers published during 2008–2009 spent an average of a year in the hands of the journal (adding up all “rounds”) and nine months in the hands of the authors (carrying out revisions).

This year saw the announcement that the Econometric Society will launch a new journal, *Quantitative Economics*, and incorporate the journal *Theoretical Economics*. We believe that the three journals will complement each other in pursuing the Econometric Society's mission of promoting research across all areas of economics, using and developing appropriate theoretical and empirical tools to address important economic questions. The three journals will be independent of each other in terms of editorial policy and decisions.

TABLE I
STATUS OF MANUSCRIPTS

	03/04	04/05	05/06	06/07	07/08	08/09
In process at beginning of year	218	156	158	165	236	216
New papers received	589	617	615	691	744	672
Revisions received	122	130	161	127	146	188
Papers accepted	61	50	57	45	57	59
Papers conditionally accepted				16	32	29
Papers returned for revision	138	153	190	95	156	157
Papers rejected or active withdrawals	574	542	520	591	656	590
[Of these rejected without full refereeing]	[194]	[199]	[146]	[163]	[154]	[123]
Papers in process at end of year	156	158	165	236	216	241

TABLE II
DISTRIBUTION OF NEW PAPERS AMONG CO-EDITORS

	03/04	04/05	05/06	06/07	07/08	08/09
Current Editors						
Acemoglu					84	70
Berry				71	70	75
Morris					170	128
Newey		113	105	116	107	89
Pesendorfer					0	116
Samuelson			110	115	102	105
Uhlig				90	91	88
Guest	12	7	12	3	4	0
Previous Editors						
Dekel	193	192	184	169	4	0
Levine	118	121	129	127	110	1
Horowitz	93				1	
Meghir	56	83	75		0	0
Postlewaite	117	101			1	0
Total	589	617	615	691	744	672

Sharing referee reports between journals may give rise to significant costs and benefits. *Econometrica's* referees do provide remarkably valuable and detailed referee reports, and we believe there is a social value in allowing these reports to be shared. As an interim measure, we have decided to experiment with sharing editorial material with the other Econometric Society journals.

TABLE III
TIME TO DECISION

	Decisions on New Submissions			Decisions on Revisions			Decisions on All Papers		
	Number	Percent-	Cumulative	Number	Percent-	Cumulative	Number	Percent-	Cumulative
	age	%	age	age	%	%	age	%	%
In ≤ 1 months	141	21%	21%	57	34%	34%	198	24%	24%
In 2 months	50	7%	29%	12	7%	41%	62	7%	31%
In 3 months	124	19%	47%	24	14%	55%	148	18%	49%
In 4 months	123	18%	66%	18	11%	66%	141	17%	66%
In 5 months	80	12%	78%	18	11%	77%	98	12%	77%
In 6 months	72	11%	88%	19	11%	88%	91	11%	88%
In 7 months	42	6%	95%	10	6%	94%	52	6%	95%
In 8 months	13	2%	97%	4	2%	96%	17	2%	97%
In >8 months	22	3%	100%	6	4%	100%	28	3%	100%
Total	667			168			835		

Specifically, if both authors and the other journal request it, we will (1) forward (anonymous) referee reports and decision letters to the other journal's editor; (2) ask referees if they would like their names and cover letters to be shared with the other journal's editor. We have started implementing this interim policy and will keep it under review in the coming years.

This year sees significant turnover on the editorial board. Whitney Newey served a term as Co-Editor from 2004 to 2008 and agreed to extend his term for an extra year through 2009. Stephen has greatly valued the judgment and expertise he has brought to handling the large and important econometrics portfolio. We are delighted that Jim Stock has agreed to take his place. Steve Berry is also standing down. Stephen is very grateful for his success in getting first rate empirical work published in *Econometrica* and improving the processing time on empirical papers. We are delighted that Jean-Marc Robin has agreed to take his place.

The Associate Editors of *Econometrica* have always played a special role at *Econometrica* with their consistently high quality refereeing and advice. We try to balance a desire to have turnover and add new talents as AEs with the remarkable sustained input that we get from some long serving AEs. This year Don Andrews (Yale University) steps down after twenty one years and Larry Epstein (Boston University) steps down after eighteen years. Generations of Co-Editors have benefited from their wisdom. We also thank Michele Boldrin (Washington University at St. Louis), Yuichi Kitamura (Yale University), Eric Renault (UNC at Chapel Hill), and Chris Shannon (University of California at Berkeley) who will not be continuing on the board. We are very grateful for all they have done for the journal. We are delighted that Xiaohong Chen (Yale University), Mikhail Golosov (Yale University), Michael Jansson (University of California, Berkeley), Felix Kubler (University of Zurich), Nicola Persico (New York University), Ben Polak (Yale University), and Ed Vytlacil (Yale University) will be joining us. We are very grateful also to those who have agreed to extend their service for another term: Oliver Linton (London School of Economics), Bart Lipman (Boston University), Thierry Magnac (Toulouse School of Economics), David Martimort (Toulouse School of Economics), and Lee Ohanian (University of California-Los Angeles (UCLA)).

Our referees also maintain a tradition of writing referee reports to a remarkably high standard. We offer them our sincere gratitude for their willingness to invest their time in offering us their insightful views on the submissions we receive. Following this report we list those who advised us this year; we apologize to anyone whom we have mistakenly omitted.

Mary Beth Bellando continues to coordinate the editorial process of *Econometrica* from the editorial office at Princeton University. She also provides invaluable support to the Editor and Co-Editors in managing the review process. Princeton University provides us with facilities and backup services for the editorial office; we are grateful in particular to Matthew Parker (for technical support), Barbara Radvany and Laura Sciarotta and to the Economics Department Chair, Chris Paxson. We benefit from the help of the Co-Editors'

assistants: Emily Gallagher, Sharline Samuelson and Lauren Fahey. John Rust and Sarbartha Bandyopadhyay of Editorial Express® continue to assist us by developing and maintaining the software we use for running the journal. The Managing Editor Geri Mattson and her staff at Mattson Publishing Services supervise an efficient publication process. We appreciate the assistance of Keira McKee and Elisabetta O'Connell at Blackwells-Wiley with the journal web site. Vytas Statulevicius and his staff at VTEX continue their superb work typesetting the journal. The Econometric Society in the form of its General Manager, Claire Sashi, and its Executive Vice-President, Rafael Repullo, oversee the production process and the management of our editorial process. We thank them for their efficiency in doing this as well as their input and advice on running the journal.

STEPHEN MORRIS
DARON ACEMOGLU
STEVE BERRY
WHITNEY NEWHEY
WOLFGANG PESENDORFER
LARRY SAMUELSON
HARALD UHLIG

THE ECONOMETRIC SOCIETY ANNUAL REPORTS
ECONOMETRICA REFEREES 2008–2009

A

Abdulkadiroglu, A.	Alos-Ferrer, C.	Arellano, M.
Abreu, D.	Alvarez, F.	Armstrong, M.
Abrevaya, J.	Amador, M.	Artemov, G.
Acharya, V.	Ambrus, A.	Aruoba, S.
Ackerberg, D.	Andersen, T.	Asheim, G.
Admati, A.	Anderson, R.	Ashworth, S.
Aghion, P.	Anderson, S.	Asker, J.
Aguirregabiria, V.	Angrist, J.	Atakan, A.
Ahn, D.	Antinolfi, G.	Atkeson, A.
Albuquerque, R.	Antras, P.	Attanasio, O.
Alesina, A.	Aoyagi, M.	Auer, P.
Allen, F.	Apesteguia, J.	Austen-Smith, D.
Almeida, H.	Aradillas-Lopez, A.	
Al-Najjar, N.	Arcidiacono, P.	

B

Bachmann, R.	Beker, P.	Bond, S.
Back, K.	Benartzi, S.	Bonhomme, S.
Baeurle, G.	Benhabib, J.	Bontemps, C.
Bagnoli, M.	Benitez-Silva, H.	Borell, C.
Bagwell, K.	Benkard, L.	Borgers, T.
Bajari, P.	Berk, J.	Borovicka, J.
Balder, E.	Berliant, M.	Bossaerts, P.
Baldazzi, P.	Bernheim, B.	Bossert, W.
Baliga, S.	Bester, A.	Boyarchenko, N.
Balkenborg, D.	Bester, H.	Brandenburger, A.
Ballester, M.	Bhaskar, V.	Brandts, J.
Bandi, F.	Bhattacharya, S.	Breitung, J.
Banerjee, A.	Biais, B.	Bresnahan, T.
Barberis, N.	Binsbergen, J.	Bridges, D.
Barndorff-Nielsen, O.	Bisin, A.	Browning, M.
Bartelsman, E.	Bleakley, H.	Brugemann, B.
Basak, S.	Bloch, F.	Brunnermeier, M.
Basso, A.	Bloom, N.	Buchinsky, M.
Baucells, M.	Blume, A.	Buckle, R.
Baum-Snow, N.	Blume, L.	Burnside, C.
Beare, B.	Board, O.	Burriel, P.
Beck, T.	Bolton, G.	
Becker, R.	Bommier, A.	

C

- | | | |
|---------------|-----------------|--------------------|
| Cabralles, A. | Chamley, C. | Collado, M. |
| Cagetti, M. | Chaney, T. | Collard-Wexler, A. |
| Calzolari, G. | Charness, G. | Compte, O. |
| Camera, G. | Chateauneuf, A. | Conlon, C. |
| Camerer, C. | Chatterjee, S. | Conlon, J. |
| Canay, I. | Chen, S. | Corbae, D. |
| Caner, M. | Chen, X. | Corchon, L. |
| Canova, F. | Cheridito, P. | Cordoba, J. |
| Cantillon, E. | Chernov, M. | Cornand, C. |
| Cao, D. | Chesher, A. | Costa, A. |
| Carlsson, H. | Chiappori, P. | Costa-Gomes, M. |
| Carneiro, P. | Choo, E. | Costain, J. |
| Carranza, J. | Christensen, J. | Costinot, A. |
| Carrasco, M. | Chugh, S. | Cox, J. |
| Carro, J. | Chung, K. | Crawford, G. |
| Casari, M. | Citanna, A. | Cremer, J. |
| Caselli, F. | Coate, S. | Cressman, R. |
| Chalak, K. | Cochrane, J. | Cripps, M. |
| Chambers, C. | Cole, H. | Cunha, F. |

D

- | | | |
|----------------|------------------|----------------|
| Dal Bo, P. | Dessein, W. | Draganska, M. |
| Danielsson, J. | Dette, H. | Dranove, D. |
| Das, S. | Devalande, A. | Drost, F. |
| Davidson, C. | Devereux, P. | Dube, J. |
| Davidson, J. | Dewachter, H. | Dubra, J. |
| Davis, D. | Di Tillio, A. | Duffee, G. |
| Davydenko, S. | Dickens, W. | Dufwenberg, M. |
| Dayanik, S. | Diebold, F. | Duggan, J. |
| de Clippel, G. | Dillenberger, D. | Dulleck, U. |
| Deb, J. | DiNardo, J. | Dupuy, A. |
| Del Negro, M. | Dittmar, R. | Durlauf, S. |
| Dell, M. | Dixon, H. | Dutta, B. |
| DellaVigna, S. | Donald, S. | Dutu, R. |
| DeMarzo, P. | Donaldson, J. | Dybvig, P. |

E

- | | | |
|----------------|---------------|---------------|
| Easley, D. | Egorov, G. | Eliaz, K. |
| Eberly, J. | Ehlers, L. | Ellickson, P. |
| Echenique, F. | Einav, L. | Ellingsen, T. |
| Economides, N. | Eisenberg, A. | Elliott, G. |
| Ederer, F. | Ejarque, J. | Emons, W. |
| Eeckhout, J. | Ekmekci, M. | Engel, E. |

Engelmann, D.
Engstrom, E.
Epple, D.
Epstein, D.
Epstein, L.

Eraslan, H.
Erickson, G.
Escanciano, J.
Eso, P.
Esponda, I.

Evans, G.
Evans, R.
Eyster, E.

F

Fafchamps, M.
Faingold, E.
Fair, R.
Falk, A.
Fan, J.
Fan, Y.
Farhi, E.
Feenstra, R.
Fehr, E.
Fehr-Duda, H.
Feinberg, Y.
Feldhutter, P.

Fernandez-Villaverde, J.
Ferreira, D.
Ferrie, J.
Fershtman, C.
Feyrer, J.
Finan, F.
Finkelstein, A.
Firpo, S.
Fisher, J.
Fleurbaey, M.
Foellmer, H.
Foellmi, R.

Forges, F.
Foster, A.
Foucault, T.
Fox, J.
Frankel, D.
Frazer, G.
Frechette, G.
Friedenberg, A.
Friedman, D.
Fuchs, W.
Fudenberg, D.
Fulop, A.

G

Gabaix, X.
Gaechter, S.
Gale, D.
Galenianos, M.
Galeotti, A.
Galor, O.
Gancia, G.
Gans, J.
Garcia, R.
Garicano, L.
Garlappi, L.
Garleanu, N.
Gautier, E.
Gautier, P.
Genicot, G.
Gerardi, D.

Ghirardato, P.
Ghysels, E.
Giannone, D.
Gilboa, I.
Gilleskie, D.
Gjerstad, S.
Glasserman, P.
Glazer, J.
Glosten, L.
Gneezy, U.
Goeree, M.
Goldstein, I.
Goldstein, R.
Gollier, C.
Golosov, M.
Gorton, G.

Gossner, O.
Gottardi, P.
Gottschalk, P.
Gottschling, A.
Gowrisankaran, G.
Goyal, S.
Graham, J.
Grant, S.
Grauer, R.
Green, E.
Greenstein, S.
Greenwood, J.
Greiner, B.
Gromb, D.
Guerrieri, V.
Guggenberger, P.

H

Hafalir, I.
Hagedorn, M.
Haldrup, N.
Halevy, Y.
Hall, A.

Hall, R.
Haltiwanger, J.
Hamilton, J.
Hansen, B.
Hansen, C.

Hansen, L.
Hansen, P.
Harding, M.
Harris, C.
Harrison, G.

Harstad, B.	Heinemann, F.	Holmstrom, B.
Hasegawa, H.	Hellwig, C.	Hommes, C.
Hastings, J.	Hellwig, M.	Hong, H.
Hatfield, J.	Hendel, I.	Honkapohja, S.
Hausman, J.	Henderson, V.	Honore, B.
Hautsch, N.	Hermalin, B.	Hopkins, E.
Hawkins, W.	Heyma, A.	Hörner, J.
Hayashi, T.	Heyman, J.	Horowitz, J.
Hazan, M.	Hidalgo, J.	Hossain, T.
He, Z.	Hillier, G.	House, C.
Healy, P.	Hirano, K.	Huang, J.
Heathcote, J.	Hlawitschka, W.	Huck, S.
Heaton, J.	Hoderlein, S.	Hurkens, S.
Heifetz, A.	Hofbauer, J.	Hurst, E.

I

Ibragimov, R.	Imbens, G.	Iskrev, N.
Ichimura, H.	Ioannides, Y.	
Ichino, A.	Iribarri, N.	

J

Jacquet, N.	Jensen, M.	Johansen, S.
Jakiela, P.	Jeong, H.	Jones, C.
Janssen, M.	Jermann, U.	Ju, N.
Jansson, M.	Jewitt, I.	Judd, K.
Jayachandran, S.	Jia, P.	
Jenish, N.	Jofre, A.	

K

Kaboski, J.	Kartik, N.	Klein, R.
Kagel, J.	Kasa, K.	Kleinberg, R.
Kajii, A.	Katz, M.	Klenow, P.
Kamada, Y.	Keane, M.	Klibanoff, P.
Kamihigashi, T.	Keller, G.	Knight, B.
Kandori, M.	Kelsey, D.	Kocherlakota, N.
Kaneko, M.	Kennan, J.	Koenker, R.
Kaniel, R.	Khan, A.	Koessler, F.
Kanwar, S.	Khan, S.	Kogan, L.
Kapan, T.	Kilian, L.	Koijen, R.
Kapetanios, G.	Kimball, M.	Kojima, F.
Kapicka, M.	Kinnan, C.	Kollmann, R.
Kaplanski, G.	Kircher, P.	Kolyuzhnov, D.
Kariv, S.	Kirchsteiger, G.	Koopman, S.
Karlan, D.	Kiyotaki, N.	Kopylov, I.

Kortum, S.
 Koszegi, B.
 Kramer, L.
 Kranton, R.
 Krasnokutskaya, E.

Krebs, T.
 Krishna, V.
 Kristensen, D.
 Krueger, D.
 Krysiak, F.

Kuersteiner, G.
 Kuhn, K.
 Kumhof, M.
 Kuruscu, B.
 Kuzmics, C.

L

Lagos, R.
 Lagunoff, R.
 Laibson, D.
 Laitner, J.
 Landon-Lane, J.
 Landsberger, M.
 Lauermann, S.
 Lavy, V.
 Le Breton, M.
 Leahy, J.
 Lee, D.
 Lee, J.
 Lee, R.
 Lee, S.
 Leeper, E.

Lehmann, E.
 Levchenko, A.
 Levin, D.
 Levin, J.
 Levinsohn, J.
 Levy, G.
 Lewbel, A.
 Lewellen, J.
 Lewis, G.
 Li, T.
 Li, Y.
 LiCalzi, M.
 Linde, J.
 Lippi, M.
 Lise, J.

List, J.
 Liu, H.
 Liu, Q.
 Livshits, I.
 Lleras-Muney, A.
 Lochner, L.
 Lockwood, B.
 Lopomo, G.
 Lorenzoni, G.
 Lovo, S.
 Luetkepohl, H.
 Lustig, J.
 Luttmer, E.

M

Maccheroni, F.
 MacDonald, G.
 MacGee, J.
 Machina, M.
 Mackowiak, B.
 MacLeod, W.
 Makarov, I.
 Manea, M.
 Manelli, A.
 Manovskii, I.
 Manski, C.
 Mansur, E.
 Manzini, P.
 Marinacci, M.
 Mariotti, M.
 Mariotti, T.
 Mark, N.
 Martin, I.
 Martinelli, C.
 Masatlioglu, Y.
 Matouschek, N.

Matsushima, H.
 Matsuyama, K.
 Maug, E.
 Mayraz, G.
 McAdams, D.
 McCaffrey, D.
 McElroy, M.
 McLean, R.
 McLennan, A.
 McMillan, R.
 Meddahi, N.
 Meghir, C.
 Meirowitz, A.
 Melin, L.
 Melitz, M.
 Mendoza, E.
 Merz, M.
 Messner, M.
 Meyer-ter-Vehn, M.
 Midrigan, V.
 Miguel, E.

Mikusheva, A.
 Milgrom, P.
 Miller, D.
 Miller, R.
 Minehart, D.
 Minelli, E.
 Miravete, E.
 Moen, E.
 Moench, E.
 Moldovanu, B.
 Molinari, F.
 Moon, H.
 Moore, D.
 Morelli, M.
 Moreno, D.
 Moretti, E.
 Morgan, J.
 Moscarini, G.
 Mroz, T.
 Mueller, U.
 Mukerji, S.

Munshi, K.
Muthoo, A.

Myerson, R.
Mykland, P.

Mylovanov, T.

N

Nagypal, E.
Nakajima, D.
Nau, R.
Navarro, S.
Neary, J.
Neeman, Z.
Nehring, K.

Neilson, W.
Nevo, A.
Ng, S.
Nino-Mora, J.
Nishiyama, Y.
Nocke, V.
Noldeke, G.

Noor, J.
Norman, P.
Nowak, A.
Nunn, N.
Nyarko, Y.

O

Obara, I.
O'Donoghue, T.
Okui, R.
Olken, B.
Onatski, A.
O'Neill, B.
Opp, C.

Oreopoulos, P.
Ortoleva, P.
Osborne, M.
Ostrovsky, M.
Otrok, C.
Otsu, T.
Ottaviani, M.

Ottaviano, G.
Ou-Yang, H.
Ozaki, H.
Ozdagli, A.
Ozdenoren, E.

P

Padro i Miquel, G.
Pakes, A.
Palacios-Huerta, I.
Palmero, C.
Palomino, F.
Panageas, S.
Parreira, S.
Pathak, P.
Pauzner, A.
Pavan, A.
Pavan, R.
Pearce, D.
Peche, S.
Pedersen, L.
Pennebaker, J.
Penta, A.
Pepper, J.
Perri, F.

Perrigne, I.
Perron, B.
Perry, M.
Persico, N.
Persson, L.
Pesendorfer, M.
Peski, M.
Petajisto, A.
Peters, H.
Peters, M.
Petrie, R.
Petrin, A.
Philippon, T.
Pinkse, J.
Piqueira, N.
Pissarides, C.
Pistaferri, L.
Plantin, G.

Plott, C.
Podczeck, K.
Podolskij, M.
Polak, B.
Polborn, M.
Polemarchakis, H.
Porter, J.
Portier, F.
Postel-Vinay, F.
Pouzo, D.
Powell, J.
Powell, R.
Prelec, D.
Proto, E.
Puppe, C.
Pycia, M.

Q

Qu, Z.
Quah, J.

Quesada, L.

Quinzii, M.

R

Rabin, M.
Rady, S.
Ragusa, G.
Rahi, R.
Raith, M.
Rajan, U.
Ramey, V.
Rangel, A.
Ravn, M.
Ray, D.
Razin, R.
Redding, S.
Reichelstein, S.
Reiley, D.
Reis, R.

Reshef, A.
Ridder, G.
Riedel, F.
Riella, G.
Rigobon, R.
Riordan, M.
Ritzberger, K.
Roberts, M.
Robins, J.
Robinson, J.
Robotti, C.
Rochet, J.
Rodriguez-Mora, S.
Rogers, B.
Rogers, L.

Romano, J.
Rosenbaum, M.
Rosenberg, D.
Rosenzweig, M.
Ross, S.
Rossi-Hansberg, E.
Rostek, M.
Rothschild, C.
Routledge, B.
Rozen, K.
Rubio-Ramirez, J.
Rudanko, L.
Rust, J.
Ryan, S.
Rysman, M.

S

Sadowski, P.
Sadzik, T.
Sahin, A.
Said, M.
Sakata, S.
Sakovics, J.
Salanie, B.
Salant, Y.
Salmon, T.
Sandholm, W.
Sandroni, A.
Sannikov, Y.
Santos, A.
Santos, M.
Sappington, D.
Sargent, T.
Sarin, R.
Sarno, L.
Sarver, T.
Satterthwaite, M.
Sbordone, A.
Scarsini, M.

Scheuer, F.
Schipper, B.
Schlag, K.
Schlee, E.
Schlesinger, H.
Schmidt-Dengler, P.
Schmitt-Grohe, S.
Schmitz, P.
Scholl, A.
Schorfheide, F.
Schott, P.
Schotter, A.
Schummer, J.
Scotchmer, S.
Scott Morton, F.
Segal, I.
Segev, E.
Seim, K.
Seitz, S.
Sekiguchi, T.
Sen, A.
Sentana, E.

Seo, K.
Serrano, R.
Severinov, S.
Seymour, R.
Shaikh, A.
Shang, D.
Shanken, J.
Shcherbakov, A.
Shearer, B.
Sher, I.
Shi, S.
Shi, X.
Shmaya, E.
Shum, M.
Sieg, H.
Silverman, D.
Simsek, A.
Singleton, K.
Siow, A.
Skiadas, C.
Skreta, V.
Sleet, C.

Smith, J.	Spiliopoulos, L.	Strulovici, B.
Smith, L.	Sprumont, Y.	Strzalecki, T.
Smorodinsky, R.	Squintani, F.	Sun, Y.
Sobel, J.	Stacchetti, E.	Sunder, S.
Solan, E.	Stachurski, J.	Sung, J.
Sonin, K.	Stahl, D.	Suri, T.
Sonmez, T.	Starmer, C.	Svensson, L.
Sorensen, A.	Steiner, J.	Swanson, N.
Sorensen, P.	Stinchcombe, M.	Sweeting, A.
Sorger, G.	Stoker, T.	Szeidl, A.
Souleles, N.	Stovall, J.	Szentes, B.
Spiegel, Y.	Strausz, R.	
Spiegler, R.	Strelbulaev, I.	

T

Taber, C.	Tercieux, O.	Tornell, A.
Takahashi, S.	Terviö, M.	Train, K.
Tallon, J.	Thesmar, D.	Trebbi, F.
Tan, W.	Thomas, J.	Trefler, D.
Tartari, M.	Thomson, W.	Trenkler, C.
Taub, B.	Tian, G.	Tripathi, G.
Tchistyi, A.	Ticchi, D.	Troger, T.
Teague, V.	Tille, C.	Tsui, K.
Telmer, C.	Timmins, C.	Tsywinski, A.
Telyukova, I.	Tirole, J.	Turner, S.
Temzelides, T.	Todorov, V.	Tybout, J.
Teneketzis, D.	Tomala, T.	Tyson, C.

U

Unver, U.	Uribe, M.	Urzuá, S.
-----------	-----------	-----------

V

Valimaki, J.	Veldkamp, L.	Villeneuve, B.
Van Biesebroeck, J.	Veracierto, M.	Vincent, D.
Van Nieuwerburgh, S.	Verdelhan, A.	Vindigni, A.
Van Wesep, E.	Veronesi, P.	Vives, X.
Van Zandt, T.	Vesterlund, L.	Vogel, J.
Vartiainen, H.	Viceira, L.	Vogelsang, T.
Vayanos, D.	Vickers, J.	Vogt, B.
Vecer, J.	Vieille, N.	Volij, O.
Vega-Redondo, F.	Villas-Boas, S.	Vuong, Q.

W

- Wachter, J.
Wacziarg, J.
Wakker, P.
Waldfogel, J.
Waldman, M.
Walker, J.
Wang, R.
Wang, T.
Wang, Z.
Watson, J.
Watson, M.
Weber, R.
Weibull, J.
Weil, D.
- Weill, P.
Weinstein, J.
Weintraub, G.
Weitzman, M.
Wen, Q.
Wendner, R.
Weretka, M.
Werner, J.
West, K.
Weyl, E.
Weymark, J.
White, H.
White, L.
White, M.
- Whited, T.
Wiederholt, M.
Williams, N.
Williamson, S.
Wilson, A.
Windmeijer, F.
Winter, E.
Wiseman, T.
Wolfers, J.
Wooders, J.
Wouters, R.
Woutersen, T.
Wu, L.
Wurgler, J.

X

- Xu, S.
Xu, Y.

Y

- Yannelis, N.
Yao, Q.
Yao, T.
Yared, P.
Yariv, L.
- Yaron, A.
Yasuda, Y.
Yeaple, S.
Yeltekin, S.
Yilankaya, O.
- Yildirim, H.
Yildiz, M.
Yildiz, N.
Yilmaz, B.
Young, A.

Z

- Zaffaroni, P.
Zame, W.
Zapatero, F.
Zariphopoulou, T.
- Zaslavsky, A.
Zeiler, K.
Zhang, J.
Zhang, L.
- Zhou, C.
Zhou, G.
Zurita, F.

THE ECONOMETRIC SOCIETY ANNUAL REPORTS
REPORT OF THE EDITORS OF THE MONOGRAPH SERIES

THE GOAL OF THE SERIES is to promote the publication of high-quality research works in the fields of *Economic Theory*, *Econometrics*, and *Quantitative Economics* more generally. Publications may range from more or less extensive accounts of the state of the art in a field to which the authors have made significant contributions to shorter monographs representing important advances on more specific issues. In addition to the usual promotion by the Publisher (Cambridge University Press) in their advertising and displays at conferences, it also arranges for members of the Econometric Society to receive monographs at a special discount.

The publishing arrangement with Cambridge University Press specifies that the reviewing process and the decision to publish a monograph in the series rests solely in the hands of the Editors appointed by the Society, in the same way as for papers submitted to *Econometrica*. Our experience shows that this procedure generates quite valuable services to the authors. Referee reports are usually very professional, and contain detailed and specific suggestions on how to improve the manuscript. Such services, which are not normally offered by private publishing companies, are among the features that distinguish the Monograph Series of the Society from others.

The complete list of publications in the series follows; the original publication dates of the hardcover (HC) and paperback (PB) versions are given.

All 45 monographs are now available for electronic purchase, and available online to Econometric Society members free of charge.

1. W. Hildenbrand (ed.), *Advances in Economic Theory*, *HC:2/83, *PB:8/85.
2. W. Hildenbrand (ed.), *Advances in Econometrics*, *HC:2/83, *PB:8/85.
3. G. S. Maddala, *Limited Dependent and Qualitative Variables in Econometrics*, HC:3/83, PB:6/86.
4. G. Debreu, *Mathematical Economics*, HC:7/83, PB:10/86.
5. J.-M. Grandmont, *Money and Value*, *HC:11/83, *PB:9/85.
6. F. M. Fisher, *Disequilibrium Foundations of Equilibrium Economics*, HC:11/83, PB:3/89.
7. B. Peleg, *Game Theoretic Analysis of Voting in Committees*, *HC:7/84.
8. R. J. Bowden and D. A. Turkington, *Instrumental Variables*, *HC:1/85, *PB:1/90. Second edition in process.
9. A. Mas-Colell, *The Theory of General Economic Equilibrium: A Differentiable Approach*, HC:8/85, PB:1/90.
10. J. Heckman and B. Singer, *Longitudinal Analysis of Labor Market Data*, *HC:10/85.
11. C. Hsiao, *Analysis of Panel Data*, *HC:7/86, *PB:11/89.
12. T. Bewley (ed.), *Advances in Economic Theory: Fifth World Congress*, *HC:8/87, *PB:7/89.
13. T. Bewley (ed.), *Advances in Econometrics: Fifth World Congress, Vol. I*, HC:11/87, PB:4/94.
14. T. Bewley (ed.), *Advances in Econometrics: Fifth World Congress, Vol. II*, HC:11/87, PB:4/94.

15. H. Moulin, *Axioms of Cooperative Decision-Making*, HC:11/88, PB:7/91.
16. L. G. Godfrey, *Misspecification Tests in Econometrics*, HC:2/89, PB:7/91.
17. T. Lancaster, *The Econometric Analysis of Transition Data*, HC:9/90, PB:6/92.
18. A. Roth and M. Sotomayor, *Two-Sided Matching*, HC:9/90, PB:6/92.
19. W. Härdle, *Applied Nonparametric Regression Analysis*, HC:10/90, PB:1/92.
20. J.-J. Laffont (ed.), *Advances in Economic Theory: Sixth World Congress, Vol. I*, HC:12/92, PB:2/95.
21. J.-J. Laffont (ed.), *Advances in Economic Theory: Sixth World Congress, Vol. II*, HC:12/92, PB:2/95.
22. H. White, *Inference, Estimation and Specification Analysis*, HC:9/94, PB:6/96.
23. C. Sims (ed.), *Advances in Econometrics: Sixth World Congress, Vol. I*, HC:3/94, PB:3/96.
24. C. Sims (ed.), *Advances in Econometrics: Sixth World Congress, Vol. II*, HC:10/94, PB:3/96.
25. R. Guesnerie, *A Contribution to the Pure Theory of Taxation*, HC:9/95, PB:9/98.
26. D. M. Kreps and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications (Seventh World Congress)*, Vol. I, HC:2/97, PB:2/97.
27. D. M. Kreps and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications (Seventh World Congress)*, Vol. II, HC:2/97, PB:2/97.
28. D. M. Kreps and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications (Seventh World Congress)*, Vol. III, HC:2/97, PB:2/97.
29. D. Jacobs, E. Kalai, and M. Kamien (eds.), *Frontiers of Research in Economic Theory—The Nancy L. Schwartz Memorial Lectures*, HC & PB:11/98.
30. A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count—Data*, HC & PB:9/98.
31. S. Strøm (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, HC & PB:2/99.
32. E. Ghysels, N. Swanson, and M. Watson (eds.), *Essays in Econometrics—Collected Papers of Clive W. J. Granger*, Vol. I, HC & PB:7/01.
33. E. Ghysels, N. Swanson, and M. Watson (eds.), *Essays in Econometrics—Collected Papers of Clive W. J. Granger*, Vol. II, HC & PB:7/01.
34. M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications (Eighth World Congress)*, Vol. I, HC:2/03, PB:2/03.
35. M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications (Eighth World Congress)*, Vol. II, HC:2/03, PB:2/03.
36. M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications (Eighth World Congress)*, Vol. III, HC:2/03, PB:2/03.
37. C. Hsiao, *Analysis of Panel Data: Second Edition*, HC & PB:2/03.
38. R. Koenker, *Quantile Regression*, HC & PB:5/05.
39. C. Blackorby, W. Bossert, and D. Donaldson, *Population Issues in Social Choice Theory, Welfare Economics and Ethics*, HC & PB:11/05.
40. J. Roemer, *Democracy, Education, and Equality*, HC & PB:1/06.
41. R. Blundell, W. Newey, and T. Persson, *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress, Vol. I, HC & PB:8/06.

42. R. Blundell, W. Newey, and T. Persson, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Vol. II*, HC & PB:8/06.
43. R. Blundell, W. Newey, and T. Persson, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Vol. III*, HC & PB:3/07.
44. F. Vega-Redondo, *Complex Networks*, HC & PB:3/07.
45. I. Gilboa, *Theory of Decision Under Uncertainty*, HC & PB:3/09.

The editors welcome submissions of high-quality manuscripts, as well as inquiries from prospective authors at an early stage of planning their monographs. Note that the series now includes shorter more focussed manuscripts on the order of 100 to 150 pages, as well as the traditional longer series of 200 to 300 or more pages. Information on submissions can be found on the society webpage.

Andrew Chesher completes his tenure as Co-Editor this year. Rosa Matzkin (UCLA) takes over as the Co-Editor responsible for econometrics. George Mailath is the Co-Editor responsible for the economic theory side of the series.

ANDREW CHESHER
GEORGE MAILATH

SUBMISSION OF MANUSCRIPTS TO THE ECONOMETRIC SOCIETY
MONOGRAPH SERIES

FOR MONOGRAPHS IN ECONOMIC THEORY, a PDF of the manuscript should be emailed to Professor George J. Mailath at *gmailath@econ.upenn.edu*. Only electronic submissions will be accepted. In exceptional cases for those who are unable to submit electronic files in PDF format, three copies of the original manuscript can be sent to:

Professor George J. Mailath
Department of Economics
3718 Locust Walk
University of Pennsylvania
Philadelphia, PA 19104-6297, U.S.A.

For monographs in theoretical and applied econometrics, a PDF of the manuscript should be emailed to Professor Rosa Matzkin at *matzkin@econ.ucla.edu*. Only electronic submissions will be accepted. In exceptional cases for those who are unable to submit electronic files in PDF format, three copies of the original manuscript can be sent to:

Professor Rosa Matzkin
Department of Economics
University of California, Los Angeles
Bunche Hall 8349
Los Angeles, CA 90095

They must be accompanied by a letter of submission and be written in English. Authors submitting a manuscript are expected not to have submitted it elsewhere. It is the authors' responsibility to inform the Editors about these matters. There is no submission charge.

The Editors will also consider proposals consisting of a detailed table of contents and one or more sample chapters, and can offer a preliminary contingent decision subject to the receipt of a satisfactory complete manuscript.

All submitted manuscripts should be double spaced on paper of standard size, 8.5 by 11 inches or European A, and should have margins of at least one inch on all sides. The figures should be publication quality. The manuscript should be prepared in the same manner as papers submitted to *Econometrica*.

Manuscripts may be rejected, returned for specified revisions, or accepted. Once a monograph has been accepted, the author will sign a contract with the Econometric Society on the one hand, and with the Publisher, Cambridge University Press, on the other. Currently, monographs usually appear no more than twelve months from the date of final acceptance.

SUBMISSION OF MANUSCRIPTS TO ECONOMETRICA

1. Members of the Econometric Society may submit papers to *Econometrica* electronically in pdf format according to the guidelines at the Society's website:

<http://www.econometricsociety.org/submissions.asp>

Only electronic submissions will be accepted. In exceptional cases for those who are unable to submit electronic files in pdf format, one copy of a paper prepared according to the guidelines at the website above can be submitted, with a cover letter, by mail addressed to Professor Stephen Morris, Dept. of Economics, Princeton University, Fisher Hall, Prospect Avenue, Princeton, NJ 08544-1021, U.S.A.

2. There is no charge for submission to *Econometrica*, but only members of the Econometric Society may submit papers for consideration. In the case of coauthored manuscripts, at least one author must be a member of the Econometric Society. Nonmembers wishing to submit a paper may join the Society immediately via Blackwell Publishing's website. Note that *Econometrica* rejects a substantial number of submissions without consulting outside referees.

3. It is a condition of publication in *Econometrica* that copyright of any published article be transferred to the Econometric Society. Submission of a paper will be taken to imply that the author agrees that copyright of the material will be transferred to the Econometric Society if and when the article is accepted for publication, and that the contents of the paper represent original and unpublished work that has not been submitted for publication elsewhere. If the author has submitted related work elsewhere, or if he does so during the term in which *Econometrica* is considering the manuscript, then it is the author's responsibility to provide *Econometrica* with details. There is no page fee; nor is any payment made to the authors.

4. *Econometrica* has the policy that all empirical and experimental results as well as simulation experiments must be replicable. For this purpose the Journal editors require that all authors submit datasets, programs, and information on experiments that are needed for replication and some limited sensitivity analysis. (Authors of experimental papers can consult the posted list of what is required.) This material for replication will be made available through the *Econometrica* supplementary material website. The format is described in the posted information for authors.

Submitting this material indicates that you license users to download, copy, and modify it; when doing so such users must acknowledge all authors as the original creators and *Econometrica* as the original publishers. If you have compelling reason we may post restrictions regarding such usage.

At the same time the Journal understands that there may be some practical difficulties, such as in the case of proprietary datasets with limited access as well as public use datasets that require consent forms to be signed before use. In these cases the editors require that detailed data description and the programs used to generate the estimation datasets are deposited, as well as information of the source of the data so that researchers who do obtain access may be able to replicate the results. This exemption is offered on the understanding that the authors made reasonable effort to obtain permission to make available the final data used in estimation, but were not granted permission. We also understand that in some particularly complicated cases the estimation programs may have value in themselves and the authors may not make them public. This, together with any other difficulties relating to depositing data or restricting usage should be stated clearly when the paper is first submitted for review. In each case it will be at the editors' discretion whether the paper can be reviewed.

5. Papers may be rejected, returned for specified revision, or accepted. Approximately 10% of submitted papers are eventually accepted. Currently, a paper will appear approximately six months from the date of acceptance. In 2002, 90% of new submissions were reviewed in six months or less.

6. Submitted manuscripts should be formatted for paper of standard size with margins of at least 1.25 inches on all sides, 1.5 or double spaced with text in 12 point font (i.e., under about 2,000 characters, 380 words, or 30 lines per page). Material should be organized to maximize readability; for instance footnotes, figures, etc., should not be placed at the end of the manuscript. We strongly encourage authors to submit manuscripts that are under 45 pages (17,000 words) including everything (except appendices containing extensive and detailed data and experimental instructions).

While we understand some papers must be longer, if the main body of a manuscript (excluding appendices) is more than the aforementioned length, it will typically be rejected without review.

7. Additional information that may be of use to authors is contained in the “Manual for *Econometrica* Authors, Revised” written by Drew Fudenberg and Dorothy Hodges, and published in the July, 1997 issue of *Econometrica*. It explains editorial policy regarding style and standards of craftsmanship. One change from the procedures discussed in this document is that authors are not immediately told which coeditor is handling a manuscript. The manual also describes how footnotes, diagrams, tables, etc. need to be formatted once papers are accepted. It is not necessary to follow the formatting guidelines when first submitting a paper. Initial submissions need only be 1.5 or double-spaced and clearly organized.

8. Papers should be accompanied by an abstract of no more than 150 words that is full enough to convey the main results of the paper. On the same sheet as the abstract should appear the title of the paper, the name(s) and full address(es) of the author(s), and a list of keywords.

9. If you plan to submit a comment on an article which has appeared in *Econometrica*, we recommend corresponding with the author, but require this only if the comment indicates an error in the original paper. When you submit your comment, please include any correspondence with the author. Regarding comments pointing out errors, if an author does not respond to you after a reasonable amount of time, then indicate this when submitting. Authors will be invited to submit for consideration a reply to any accepted comment.

10. Manuscripts on experimental economics should adhere to the “Guidelines for Manuscripts on Experimental Economics” written by Thomas Palfrey and Robert Porter, and published in the July, 1991 issue of *Econometrica*.

Typeset at VTEX, Akademijos Str. 4, 08412 Vilnius, Lithuania. Printed at The Sheridan Press, 450 Fame Avenue, Hanover, PA 17331, USA.

Copyright © 2010 by The Econometric Society (ISSN 0012-9682). Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation, including the name of the author. Copyrights for components of this work owned by others than the Econometric Society must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works, requires prior specific permission and/or a fee. Posting of an article on the author's own website is allowed subject to the inclusion of a copyright statement; the text of this statement can be downloaded from the copyright page on the website www.econometricsociety.org/permis.asp. Any other permission requests or questions should be addressed to Claire Sashi, General Manager, The Econometric Society, Dept. of Economics, New York University, 19 West 4th Street, New York, NY 10012, USA. Email: permissions@econometricsociety.org.

Econometrica (ISSN 0012-9682) is published bi-monthly by the Econometric Society, Department of Economics, New York University, 19 West 4th Street, New York, NY 10012. Mailing agent: Sheridan Press, 450 Fame Avenue, Hanover, PA 17331. Periodicals postage paid at New York, NY and additional mailing offices.

U.S. POSTMASTER: Send all address changes to *Econometrica*, Blackwell Publishing Inc., Journals Dept., 350 Main St., Malden, MA 02148, USA.

THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic
Theory in its Relation to Statistics and Mathematics*
Founded December 29, 1930

Website: www.econometricsociety.org

Membership

Joining the Econometric Society, and paying by credit card the corresponding membership rate, can be done online at www.econometricsociety.org. Memberships are accepted on a calendar year basis, but the Society welcomes new members at any time of the year, and in the case of print subscriptions will promptly send all issues published earlier in the same calendar year.

Membership Benefits

- Possibility to submit papers to *Econometrica*, *Quantitative Economics*, and *Theoretical Economics*
- Possibility to submit papers to Econometric Society Regional Meetings and World Congresses
- Full text online access to all published issues of *Econometrica* (*Quantitative Economics* and *Theoretical Economics* are open access)
- Full text online access to papers forthcoming in *Econometrica* (*Quantitative Economics* and *Theoretical Economics* are open access)
- Free online access to Econometric Society Monographs, including the volumes of World Congress invited lectures
- Possibility to apply for travel grants for Econometric Society World Congresses
- 40% discount on all Econometric Society Monographs
- 20% discount on all John Wiley & Sons publications
- For print subscribers, hard copies of *Econometrica*, *Quantitative Economics*, and *Theoretical Economics* for the corresponding calendar year

Membership Rates

Membership rates depend on the type of member (ordinary or student), the class of subscription (print and online or online only) and the country classification (high income or middle and low income). The rates for 2010 are the following:

		High Income	Other Countries
Ordinary Members			
Print and Online	1 year (2010)	\$90 / €65 / £55	\$50
Online only	1 year (2010)	\$50 / €35 / £30	\$10
Print and Online	3 years (2010–2012)	\$216 / €156 / £132	\$120
Online only	3 years (2010–2012)	\$120 / €84 / £72	\$24
Student Members			
Print and Online	1 year (2010)	\$50 / €35 / £30	\$50
Online only	1 year (2010)	\$10 / €7 / £6	\$10

Euro rates are for members in Euro area countries only. Sterling rates are for members in the UK only. All other members pay the US dollar rate. Countries classified as high income by the World Bank are: Andorra, Antigua and Barbuda, Aruba, Australia, Austria, The Bahamas, Bahrain, Barbados, Belgium, Bermuda, Brunei, Canada, Cayman Islands, Channel Islands, Croatia, Cyprus, Czech Republic, Denmark, Equatorial Guinea, Estonia, Faeroe Islands, Finland, France, French Polynesia, Germany, Greece, Greenland, Guam, Hong Kong (China), Hungary, Iceland, Ireland, Isle of Man, Israel, Italy, Japan, Rep. of Korea, Kuwait, Liechtenstein, Luxembourg, Macao (China), Malta, Monaco, Netherlands, Netherlands Antilles, New Caledonia, New Zealand, Northern Mariana Islands, Norway, Oman, Portugal, Puerto Rico, Qatar, San Marino, Saudi Arabia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Taiwan (China), Trinidad and Tobago, United Arab Emirates, United Kingdom, United States, Virgin Islands (US).

Institutional Subscriptions

Information on *Econometrica* subscription rates for libraries and other institutions is available at www.econometricsociety.org. Subscription rates depend on the class of subscription (print and online or online only) and the country classification (high income, middle income, or low income).

Back Issues and Claims

For back issues and claims contact Wiley Blackwell at cs-websites@wiley.com.

THE ECONOMETRIC SOCIETY

*An International Society for the Advancement of Economic
Theory in its Relation to Statistics and Mathematics
Founded December 29, 1930*

Website: www.econometricsociety.org

Administrative Office: Department of Economics, New York University,
19 West 4th Street, New York, NY 10012, USA; Tel. 212-9983820; Fax 212-9954487
General Manager: Claire Sashi (sashi@econometricsociety.org)

2010 OFFICERS

JOHN MOORE, University of Edinburgh and London School of Economics, PRESIDENT
BENGT HOLMSTRÖM, Massachusetts Institute of Technology, FIRST VICE-PRESIDENT
JEAN-CHARLES ROCHE, Toulouse School of Economics, SECOND VICE-PRESIDENT
ROGER B. MYERSON, University of Chicago, PAST PRESIDENT
RAFAEL REPULLO, CEMFI, EXECUTIVE VICE-PRESIDENT

2010 COUNCIL

DARON ACEMOGLU, Massachusetts Institute of Technology	CESAR MARTINELLI, ITAM
MANUEL ARELLANO, CEMFI	ANDREW MCLENNAN, University of Queensland
SUSAN ATHEY, Harvard University	ANDREU MAS-COLELL, Universitat Pompeu Fabra and Barcelona GSE
ORAZIO ATTANASIO, University College London	AKIHIKO MATSUI, University of Tokyo
DAVID CARD, University of California, Berkeley	HITOSHI MATSUSHIMA, University of Tokyo
JACQUES CRÉMER, Toulouse School of Economics	MARGARET MEYER, University of Oxford
(*)EDDIE DEKEL, Tel Aviv University and Northwestern University	PAUL R. MILGROM, Stanford University
MATHIAS DEWATRIPONT, Free University of Brussels	STEPHEN MORRIS, Princeton University
DARRELL DUFFIE, Stanford University	JUAN PABLO NICOLINI, Universidad Torcuato di Tella
GLENN ELLISON, Massachusetts Institute of Technology	CHRISTOPHER A. PISSARIDES, London School of Economics
HIDEHIKO ICHIMURA, University of Tokyo	(*)ROBERT PORTER, Northwestern University
(*)MATTHEW O. JACKSON, Stanford University	JEAN-MARC ROBIN, Université de Paris I and University College London
MICHAEL P. KEANE, University of Technology Sydney	LARRY SAMUELSON, Yale University
LAWRENCE J. LAU, Chinese University of Hong Kong	ARUNAVA SEN, Indian Statistical Institute
	JÖRGEN W. WEIBULL, Stockholm School of Economics

The Executive Committee consists of the Officers, the Editors of *Econometrica* (Stephen Morris), *Quantitative Economics* (Orazio Attanasio), and *Theoretical Economics* (Martin J. Osborne), and the starred (*) members of the Council.

REGIONAL STANDING COMMITTEES

Australasia: Trevor S. Breusch, Australian National University, CHAIR; Maxwell L. King, Monash University, SECRETARY.

Europe and Other Areas: John Moore, University of Edinburgh and London School of Economics, CHAIR; Helmut Bester, Free University Berlin, SECRETARY; Enrique Sentana, CEMFI, TREASURER.

Far East: Hidehiko Ichimura, University of Tokyo, CHAIR.

Latin America: Pablo Andres Neumeyer, Universidad Torcuato di Tella, CHAIR; Juan Dubra, University of Montevideo, SECRETARY.

North America: Bengt Holmström, Massachusetts Institute of Technology, CHAIR; Claire Sashi, New York University, SECRETARY.

South and Southeast Asia: Arunava Sen, Indian Statistical Institute, CHAIR.