Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods

Raj Chetty

Department of Economics, Harvard University, Cambridge, Massachusetts 02138;
email: chetty@fas.harvard.edu

Key Words

program evaluation, treatment effects, public economics, taxation, social insurance

Abstract

The debate between structural and reduced-form approaches has generated substantial controversy in applied economics. This article reviews a recent literature in public economics that combines the advantages of reduced-form strategies—transparent and credible identification—with an important advantage of structural models—the ability to make predictions about counterfactual outcomes and welfare. This literature has developed formulas for the welfare consequences of various policies that are functions of reduced-form elasticities rather than structural primitives. I present a general framework that shows how many policy questions can be answered by estimating a small set of sufficient statistics using program-evaluation methods. I use this framework to synthesize the modern literature on taxation, social insurance, and behavioral welfare economics. Finally, I discuss problems in macroeconomics, labor, development, and industrial organization that could be tackled using the sufficient statistic approach.

1. INTRODUCTION

There are two competing paradigms for policy evaluation and welfare analysis in economics: the structural approach and the reduced-form approach (also known as the program-evaluation or treatment-effect approach). The division between structural and reduced-form methods has split the economics profession into two camps whose research programs have evolved almost independently despite focusing on similar questions. The structural approach specifies complete models of economic behavior and estimates or calibrates the primitives of such models. Armed with the fully estimated model, these studies then simulate the effects of changes in policies and the economic environment on behavior and welfare. This powerful methodology has been applied to an array of topics, ranging from the design of tax and transfer policies in public finance to the sources of inequality in labor economics and antitrust policy in industrial organization.

Critics of the structural approach argue that it is difficult to identify all primitive parameters in an empirically compelling manner because of selection effects, simultaneity bias, and omitted variables. These researchers instead advocate reduced-form strategies that estimate statistical relationships, placing priority on the identification of causal effects using research designs that exploit quasi-experimental variation.[1] Reduced-form studies have identified a variety of important empirical regularities, especially in labor, public, and development economics. Advocates of the structural paradigm criticize the reduced-form approach for estimating statistics that are not policy-invariant parameters of economic models and therefore have limited relevance for policy and welfare analysis (Rosenzweig & Wolpin 2000, Heckman & Vytlacil 2005, Deaton 2009).[2]

This article argues that a set of papers in public economics written over the past decade (see **Table 1**) provide a middle ground between the two methods. These papers develop sufficient-statistic formulas that combine the advantages of reduced-form empirics, transparent and credible identification, with an important advantage of structural models, the ability to make precise statements about welfare. The central concept of the sufficient-statistic approach (illustrated in **Figure 1**) is to derive formulas for the welfare consequences of policies that are functions of high-level elasticities rather than deep primitives. Even though there are multiple combinations of primitives that are consistent with the inputs to the formulas, all such combinations have the same welfare implications.[3] For example, Feldstein (1999) showed that the marginal welfare gain from raising the income-tax rate can be expressed purely as a function of the elasticity of taxable income even though taxable income may be a complex function of choices such as hours, training, and effort. Saez (2001) demonstrated that labor-supply elasticity estimates can be used to make inferences about the optimal progressive income-tax schedule in Mirrlees's (1971) model. Chetty (2008) showed that the welfare gains from social insurance can be expressed purely in terms of the liquidity and moral hazard effects of the program in a broad class of dynamic, stochastic models. Each of these papers answers a policy question

---

[1]The term reduced form is a misnomer: The relationships that are estimated are generally not reduced forms of economic models. I use the term reduced form here simply for consistency with standard terminology used to describe design-based studies that identify treatment effects.

[2]See section 1 of Rosenzweig & Wolpin (2000) and table V of Heckman & Vytlacil (2005) for a more detailed comparison of the structural and treatment-effect approaches.

[3]The term sufficient statistic is borrowed from the statistics literature: Conditional on the statistics that appear in the formula, other statistics that can be calculated from the same sample provide no additional information about the welfare consequences of the policy.

Table 1 Recent examples of structural, reduced-form, and sufficient-statistic studies

	Structural	Reduced form	Sufficient statistic
Taxation	Hoynes 1996	Eissa & Liebman 1996	Diamond 1998
	Keane & Moffitt 1998	Blundell et al. 1998	Feldstein 1999
	Blundell et al. 2000	Goolsbee 2000	Saez 2001
	Golosov & Tsyvinksi 2007	Meyer & Rosenbaum 2001	Goulder & Williams 2003
	Weinzierl 2008	Blau & Khan 2007	Chetty 2009
Social insurance	Rust & Phelan 1997	Anderson & Meyer 1997	Gruber 1997
	Golosov & Tsyvinski 2006	Gruber & Wise 1999	Chetty 2006a
	Blundell et al. 2008	Autor & Duggan 2003	Shimer & Werning 2007
	Einav et al. 2008b	Lalive et al. 2006	Chetty 2008
	Lentz 2009	Finkelstein 2007	Einav et al. 2008a
Behavioral models	Angeletos el at. 2001	Genesove & Mayer 2001	Bernheim & Rangel 2008
	İmrohoroğlu et al. 2003	Madrian & Shea 2001	Chetty et al. 2009
	Liebman & Zeckhauser 2004	Shapiro 2005	
	DellaVigna & Paserman 2005	Ashraf et al. 2006	
	Amador et al. 2006	Chetty & Saez 2009	

Notes: Categories used to classify papers are defined as follows: structural, estimate or calibrate primitives to make predictions about welfare; reduced form, estimate high-level behavioral elasticities qualitatively relevant for policy analysis, but do not provide quantitative welfare results; sufficient statistic, make predictions about welfare without estimating or specifying primitives. This list includes only selected examples that relate to the topics discussed in the text and omits many important contributions in each category.

using program-evaluation estimates, providing economic meaning to what might otherwise be viewed as atheoretical statistical estimates.

The goal of this review is to elucidate the key concepts of the sufficient-statistic methodology and encourage its use as a bridge between structural and reduced-form methods. I first provide a general framework for the derivation of sufficient-statistic formulas for welfare analysis. This framework shows how envelope conditions from optimization can be used to reduce the set of parameters that need to be identified. I then illustrate the approach by reviewing several recent papers on tax policy, social insurance, and behavioral public finance.

The idea that it is adequate to estimate certain sufficient statistics rather than the full primitive structure to answer certain questions is not new; it was well understood by the pioneers of structural estimation in the Cowles Commission (Marschak 1953, Koopmans 1953). Indeed, Heckman & Vytlacil (2007) labeled this idea "Marschak's maxim," arguing that "for many decisions (policy problems), only combinations of explicit economic parameters are required—no single economic parameter need be identified." In early microeconometric work, structural methods were preferred because the parameters of the simple
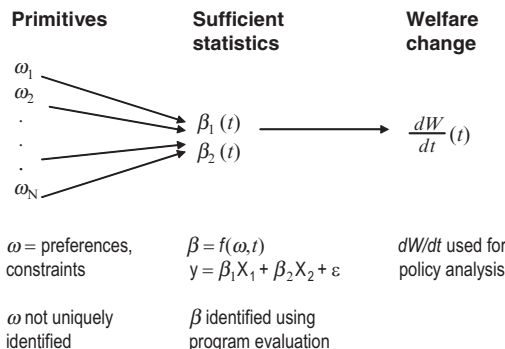
| **Primitives** | **Sufficient statistics** | **Welfare change** |

ω_1
ω_2
.
.
.
ω_N

$\beta_1(t)$
$\beta_2(t)$

$\frac{dW}{dt}(t)$

ω = preferences, constraints

$\beta = f(\omega, t)$
$y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

dW/dt used for policy analysis

ω not uniquely identified

β identified using program evaluation

Figure 1

The sufficient-statistic approach. Here we consider a policy instrument t that affects social welfare $W(t)$. The structural approach maps the primitives (ω) directly to the effects of the policy on welfare $\left(\frac{dW}{dt}\right)$. The sufficient-statistic approach leaves ω unidentified and instead identifies a smaller set of high-level parameters (β) using program-evaluation methods, e.g., via a regression of an outcome y on exogenous variables X. The β vector is sufficient for welfare analysis in that any vector ω consistent with β implies the same value of $\frac{dW}{dt}$. Identifying β does not identify ω because there are multiple ω vectors consistent with a single β vector.

models being studied in principle could be easily identified.[4] In the 1980s, it became clear that identifying primitives was difficult once one permitted dynamics, heterogeneity, and selection effects. Concerns about the identification of parameters in these richer models led a large group of empirical researchers to abandon structural methods in favor of more transparent quasi-experimental research designs (e.g., Lalonde 1986, Card 1990, Angrist & Krueger 1991).[5] A large library of treatment-effect estimates was developed in the 1980s and 1990s. The recent sufficient-statistic literature maps these treatment-effect estimates into statements about welfare in structural models that incorporate dynamics, heterogeneity, and general equilibrium effects.

The structural and sufficient-statistic approaches to welfare analysis should be viewed as complements rather than substitutes because each approach has certain advantages. The sufficient-statistic method has three benefits. First, it is simpler to implement empirically because less data and variation are needed to identify marginal treatment effects (MTEs) than to fully identify a structural model. Indeed, because of the estimation challenges, structural primitives are often calibrated to match reduced-form moments rather than formally estimated using microdata (Dawkins et al. 2001). The sufficient-statistic approach obviates the need to fully calibrate the structural model. This is especially beneficial in models with heterogeneity and discrete choice, in which the set of primitives is very large but the set of MTEs needed for welfare evaluation remains small. By estimating the relevant MTEs as a function of the policy instrument, one can integrate the formula for the marginal welfare gain between any two observed values to evaluate policy changes.

---

[4]An exception is the work of Harberger (1964), who advocated the use of elasticities as sufficient statistics for tax policy analysis in equilibrium models. As I discuss in Section 3, Harberger's work can be viewed as a predecessor to the modern sufficient-statistic literature.

[5]Imbens & Wooldridge (2008) reviewed program-evaluation methods, and Imbens (2009) discussed the advantages of such methods from a statistical perspective.

Second, the identification of structural models often requires strong assumptions given available data and variation. As it is unnecessary to identify all primitives, sufficient-statistic formulas can be implemented under weaker assumptions using design-based empirical methods. The results are therefore more transparent and empirically credible. Third, the sufficient-statistic approach can be applied even when one is uncertain about the positive model that generates observed behavior, such as in recent studies in the behavioral economics literature that document deviations from perfect rationality. In such cases, welfare analysis based on a structural model may be impossible, whereas the more agnostic sufficient-statistic approach permits some progress. For instance, Chetty et al. (2009) derived formulas for the deadweight cost of taxation in terms of price and tax elasticities in a model in which agents make arbitrary optimization errors with respect to taxes.

The parsimony of the sufficient-statistic approach naturally comes with costs. First, and most important, a new sufficient-statistic formula must be derived for each question. For example, Gruber (1997) calculated the optimal level of unemployment benefits using a sufficient-statistic formula. If one were interested in calculating the optimal duration of unemployment benefits, one would have to derive a new formula and estimate a new set of elasticities. In contrast, if one had estimated the structural primitives of the job-search model used to derive these formulas, different policy simulations could be conducted with ease. Moreover, for some questions, it may be difficult to derive a sufficient-statistic formula, and a structural approach may be the only feasible option.

A second potential weakness of sufficient-statistic formulas is that they are more easily misapplied than structural methods. This is because one can draw policy conclusions from a sufficient-statistic formula without assessing the validity of the model upon which it is based. For example, Gorodnichenko et al. (2009) showed that the assumptions about the costs of evasion underlying Feldstein's (1999) formula for the excess burden of income taxation are inconsistent with the data. In contrast, because structural methods require full estimation of the model, policy conclusions can be drawn only from models that fit the data.

A common argument in favor of the structural approach is that it has advantages in out-of-sample predictions. One can also make out-of-sample predictions using the sufficient-statistic approach by estimating MTEs as a function of the policy instrument and making a statistical extrapolation. For example, by estimating the elasticity of labor supply with respect to the tax rate at various tax rates, one can extrapolate to the welfare consequences of tax regimes that have not yet been observed. In principle, structural methods do not require such ad hoc extrapolations, as the primitive structure is by definition policy invariant. However, in practice, structural models often rely on extrapolations based on functional form assumptions (such as constant-elasticity utilities). Hence, the real advantage of structural models in out-of-sample predictions comes from the precision of the extrapolation. Statistical extrapolations may be less reliable than extrapolations guided by an economic model that imposes restrictions on how behavior changes with policies.[6]

The structural and sufficient-statistic methods can be combined to address the shortcomings of each strategy. For instance, a structural model can be calibrated to match the sufficient statistics that matter for local welfare analysis to improve its empirical credibility. Conversely, when making out-of-sample predictions using a sufficient-statistic formula,

---

[6]See Lumsdaine et al. (1992) and Keane & Wolpin (1997) for comparisons of reduced-form statistical extrapolations and model-based structural extrapolations. These studies found that structural predictions are more accurate, but statistical extrapolations that include the key variables suggested by the economic model come quite close.

one can use a structural model to guide the choice of functional forms used to extrapolate the key elasticities. By combining the two methods in this manner, researchers can pick a point in the interior of the continuum between reduced-form and structural estimation, without being pinned to one end point or the other. In addition, sufficient-statistic formulas provide theoretical guidance for reduced-form empirical work by identifying the parameters of greatest interest for a given question.

The review is organized as follows. The next section discusses Harberger's (1964) formula for the deadweight cost of taxation, a precursor to the modern literature on sufficient statistics. In Section 3, I develop a general framework that provides a six-step rubric for deriving sufficient-statistic formulas. Sections 4 through 6 review applications of the sufficient-statistic method to income taxation, social insurance, and behavioral (non-rational) models. These three sections provide a synthesis of the modern public-finance literature, showing how a dozen seemingly unrelated papers are essentially variants on the theme of finding sufficient statistics. The review concludes in Section 7 with a discussion of potential applications to fields beyond public finance.

2. A PRECEDENT: MEASURING DEADWEIGHT LOSS

Harberger (1964) popularized the measurement of the excess burden of a commodity tax using a simple elasticity-based formula. This result can be viewed as a precedent to the modern literature on sufficient statistics and provides a starting point from which to build intuition about the applications discussed below.

Let us consider a static general equilibrium model in which an individual is endowed with Z units of the numeraire (y), whose price is normalized to 1. Firms convert the numeraire good y (which can be interpreted as labor) into J other consumption goods, $x = (x_1, \ldots, x_J)$. Producing x_j units of good j requires an input of $c_j(x_j)$ units of y, where c_j is a weakly convex function. We let $c(x) = \sum_{j=1}^{J} c_j(x_j)$ denote the total cost of producing a vector x. Production is perfectly competitive. The government levies a unit tax t on good 1. We let $p = (p_1, \ldots, p_J)$ be the vector of pretax prices for the produced goods, which are determined endogenously in market equilibrium.

To simplify the exposition, we ignore income effects by assuming that utility is quasi-linear in y. The consumer takes the price vector as given and solves

$$\max_{x,y} u(x_1, \ldots, x_J) + y \qquad\qquad (1)$$
$$\text{s.t. } p \cdot x + t x_1 + y = Z,$$

where $u(x)$ is strictly quasiconcave. The representative firm takes prices as given and solves

$$\max_{x} p \cdot x - c(x). \qquad\qquad (2)$$

These two problems define maps from the price vector p to the demand and supply of the J goods, $x^D(p)$ and $x^S(p)$. The model is closed by the market-clearing condition $x^D(p) = x^S(p)$. We let $p(t)$ denote the market-clearing price vector as a function of the tax rate t.

Let us suppose the policy maker wants to measure the efficiency cost of the tax t. The efficiency (or deadweight) cost of a tax increase equals the loss in surplus from the transactions that fail to occur because of the tax. In calculating the efficiency cost, the conceptual experiment is to measure the net loss in welfare from raising the tax rate and returning

the tax revenue to the taxpayer through a lump-sum rebate.[7] With quasi-linear utility, the consumer will always choose to allocate the lump-sum rebate to consumption of the numeraire good y. Social welfare can therefore be written as the sum of the consumer's utility (which is a money metric given quasilinearity), producer profits, and tax revenue:

$$W(t) = \{\max_x u(x) + Z - tx_1 - p(t)\cdot x\} + \{\max_x p(t)\cdot x - c(x)\} + tx_1$$
$$= \{\max_x u(x) + Z - tx_1 - c(x)\} + tx_1, \tag{3}$$

where the second equation effectively recasts the decentralized equilibrium as a planner's allocation problem. The term in curly brackets measures private surplus; the tx_1 term measures tax revenue. The individual treats tax revenue as fixed when choosing x, failing to internalize the effects of his behavior on the lump-sum transfer he ultimately receives. This assumption, which is standard in efficiency-cost calculations, captures the intuition that in an economy populated by a large number of individuals, any one individual has a negligible impact on the government revenue and therefore treats it as fixed.

There are two approaches to estimating the effect of an increase in the tax on social welfare $\left(\frac{dW}{dt}\right)$. The first is to estimate (or calibrate) a J good demand and supply system to recover the utility function $u(x)$ and cost function $c(x)$. Once u and c are known, one can directly compute $W(t)$, recognizing that the tax t will affect equilibrium prices and quantities in all J markets. Preferences can be recovered using the parametric demand systems proposed, for instance, by Stone (1954) or Deaton & Muellbauer (1980). Alternatively, one can fit a demand system to the data and then integrate to obtain the expenditure function, as in Hausman (1981) or Hausman & Newey (1995). The econometric challenge in implementing any of these structural methods is simultaneity: Identifying the slope of the supply and demand curves requires $2J$ instruments.

Harberger (1964) proposed an alternative, simpler solution to the problem.[8] Differentiating Equation 3 and exploiting the first-order conditions from consumer and firm maximization yield

$$\frac{dW(t)}{dt} = -x_1 + x_1 + t\frac{dx_1}{dt} = t\frac{dx_1(t)}{dt}. \tag{4}$$

This formula shows that the effect of the tax on equilibrium quantity in the taxed market $\left(\frac{dx_1(t)}{dt}\right)$ is a sufficient statistic for analyzing the efficiency costs of tax changes. By estimating $\frac{dx_1}{dt}(t)$ for different values of t, one can calculate the welfare consequences of any tax change that lies within the observed support of t by integrating Equation 4: $\Delta W = W(t_2) - W(t_1) = \int_{t_1}^{t_2} t\frac{dx_1}{dt}(t)dt$. The key point is that the full system of supply and demand curves does not have to be identified to calculate ΔW.

Harberger's formula rests upon two conceptual insights that form the basis for modern sufficient-statistic applications. First, the behavioral responses $\left(\frac{dx}{dt}\right)$ in the curly brackets of Equation 3 can be ignored when calculating $\frac{dW}{dt}$ because of envelope conditions from consumer and firm optimization. Intuitively, social welfare has already been optimized by

---

[7]Formally, excess burden is defined using equivalent or compensating variation measures (e.g., see Auerbach 1985). The social-welfare calculation here is equivalent to these measures because we are considering a specification without income effects. With income effects, the Harberger formula discussed below applies with the Hicksian elasticity in place of the Marshallian elasticity.

[8]Hines (1999) colorfully recounts the intellectual history of the deadweight loss triangle.

individuals and firms (subject to constraints imposed by the government). Although the tax induces changes in behavior, these behavioral responses cannot have a first-order effect on private welfare; if they did, consumers or firms would not be optimizing. Second, the changes in equilibrium prices $\left(\frac{dp}{dt}\right)$ can also be ignored when calculating $\frac{dW}{dt}$. This is because prices cancel out of the expression for social welfare in Equation 3. Changes in prices simply redistribute income from producers to consumers without changing aggregate surplus.

These two observations imply that the loss in social surplus from the tax is determined purely by the difference between the agent's willingness to pay for good x_1 and the cost of producing good x_1. The difference can be measured by the area between the supply and demand curves and the initial and post-tax quantities (i.e., the Harberger triangle), which is proportional to $\frac{dx_1}{dt}$. The total derivative $\frac{dx_1}{dt} = \frac{\partial x_1}{\partial p_1}\frac{\partial p_1}{\partial t} + \frac{\partial x_1}{\partial p_2}\frac{\partial p_2}{\partial t} + \ldots + \frac{\partial x_1}{\partial p_J}\frac{\partial p_J}{\partial t}$ measures the effect of an increase in the tax rate on x_1, allowing all prices and equilibrium demands to change endogenously. The simplicity of Harberger's approach stems from estimating $\frac{dx_1}{dt}$ directly rather than estimating its various components, which is effectively what the structural approach requires.

The tradeoffs between the sufficient-statistic and structural approaches are apparent in the debate that followed Harberger's work. One limitation of Equation 4 is that it does not permit pre-existing distortions in the other markets; otherwise the spillover effects would have first-order effects on welfare. This limitation can be addressed by an extension of the formula that includes cross-price elasticities, as shown in Harberger's original analysis. Under empirically plausible approximations about the structure of the distortions, the formula can be expressed purely in terms of own-price elasticities (Goulder & Williams 2003). Hence, a different sufficient-statistic formula is required to handle cases with pre-existing distortions.[9]

A second limitation of Equation 4 is that it cannot be used directly to evaluate counterfactual policy changes, such as the imposition of a large new tax on good x_1. This limitation can be addressed by estimating $\frac{dx_1}{dt}(t)$ for various values of t and making functional-form assumptions to extrapolate out of sample. In practice, the Harberger formula is typically implemented under a linear or log-linear approximation to demand (e.g., $\frac{dx_1}{dt}$ constant) because data limitations preclude the estimation of higher-order properties of the demand curve. Simulations of calibrated models suggest that implementations of Harberger's formula using linear approximations are generally quite accurate (Shoven 1976, Ballard et al. 1985). Thus, despite its limitations, the simple Harberger triangle formula has become central to applied welfare analysis and has inspired a vast literature estimating tax elasticities.

The benefits of Harberger's approach are especially evident in modern structural models that permit heterogeneity across individuals and discrete choice. I now incorporate these features into the preceding analysis, following Small & Rosen's (1981) analysis for the discrete-choice case.

---

[9]A related concern is that one may inadvertently ignore some pre-existing distortions and apply an inaccurate version of the Harberger formula. Indeed, Goulder & Williams (2003) argued that previous applications of the simple formula in Equation 4 to assess the deadweight costs of commodity taxation are biased by an order of magnitude because they fail to account for interactions with the labor-income tax. This mistake would not have been made in a properly specified structural model.

Extension 1: Heterogeneity. Now suppose the economy has N individuals with heterogeneous preferences. Let x^i denote individual i's vector of demands and $x = \sum_{i=1}^{N} x^i$ denote aggregate demand. Individual i is endowed with Z^i units of the numeraire and has utility

$$u^i(x^i) + y. \tag{5}$$

Under a utilitarian criterion, social welfare is given by

$$W(t) = \left\{ \sum_{i=1}^{N} \max_{x^i} [u^i(x^i) + Z^i - tx_1^i] - c(x) \right\} + t \sum_{i=1}^{N} x_1^i. \tag{6}$$

The structural approach requires identifying the demand functions and utilities for all i agents. The sufficient-statistic approach simplifies the identification problem substantially here. Because there is an envelope condition for x^i for every agent, we can ignore all behavioral responses within the curly brackets when differentiating Equation 6 to obtain

$$\frac{dW(t)}{dt} = -\sum_{i=1}^{N} x_1^i + \sum_{i=1}^{N} x_1^i + t \frac{d\sum_{i=1}^{N} x_1^i}{dt} = t \frac{dx_1(t)}{dt}. \tag{7}$$

The effect of a tax increase on aggregate demand $\left(\frac{dx_1}{dt} \right)$ is a sufficient statistic for the marginal excess burden of that tax; there is no need to characterize the underlying heterogeneity in the population to implement Equation 7. Intuitively, even though each individual has a different demand elasticity, what matters for government revenue and aggregate welfare is the total change in behavior induced by the tax.[10]

   An important caveat is that with heterogeneity, $\frac{dx_1}{dt}$ may vary considerably with t because the individuals at the margin will differ with the tax rate. Hence, it is especially important to distinguish average and marginal treatment effects for welfare analysis by estimating $\frac{dx_1}{dt}(t)$ as a function of t in this case.

Extension 2: Discrete Choice. Now suppose individuals can only choose one of the J products $\{1, \ldots, J\}$. These products might represent models of cars, modes of transportation, or neighborhoods. Each product is characterized by a vector of K attributes $x_j = (x_{1j}, \ldots, x_{Kj})$ observed by the econometrician and an unobservable attribute ζ_j. If agent i chooses product j, his utility is

$$u_{ij} = v_{ij} + \varepsilon_{ij}$$
$$\text{with } v_{ij} = Z^i - p_j + \zeta_j + \phi^i(x_j),$$

where ε_{ij} is a random unobserved taste shock. Let P_{ij} denote the probability that individual i chooses option j, $P_j = \sum_i P_{ij}$ denote total (expected) demand for product j, and $P = (P_1, \ldots, P_J)$ denote the vector of aggregate product demands. Product j is produced by competitive firms using $c_j(P_j)$ units of the numeraire good y, and $c(P) = \sum_j c_j(P_j)$. This model differs from the one above in two respects: (*a*) Utility over the consumption goods is

---

[10]Of course, to analyze a policy that has heterogeneous impacts across groups (such as a progressive income tax), one needs group-specific elasticity estimates to calculate $\frac{dW}{dt}$. The key point, however, is that the only heterogeneity that matters is at the level of the policy impact; any additional heterogeneity within groups can be ignored. For instance, heterogeneous labor-supply responses within an income group need to not be characterized when analyzing optimal progressive income taxation.

replaced by utility over the product attributes $\phi^i(x_j) + \zeta_j + \varepsilon_{ij}$, and (b) the attributes can only be consumed in discrete bundles.

To build intuition, first consider a case in which individuals make binary decisions about whether to buy a single good x_1 ($J = 1$) and the price of x_1 is fixed at p_1 (constant marginal cost of supply). If an individual does not buy x_1, he spends his wealth on the numeraire and obtains utility $u_i = Z^i$. If he buys x_1, his utility is $u_i = Z^i - p_1 + \zeta_1 + \phi^i(x_1) + \varepsilon_{i1}$. Let $V^i = \zeta_1 + \phi^i(x_1) + \varepsilon_{i1}$ denote individual i's gross valuation of x_1, $F(V^i)$ denote the smooth distribution of valuations in the economy, and suppose there is a unit mass of agents. Let EZ denote the average level of wealth in the economy.

To calculate $\frac{dW}{dt}$ in this environment, we note that individuals with V^i above a cutoff $\bar{V}$ will buy x_1. The model is therefore isomorphic to one in which a representative agent chooses $\bar{V}$ to maximize total private surplus. Social surplus can be written as the sum of private surplus and tax revenue:

$$W(t) = \left\{ EZ + \max_{\bar{V}} \int_{\bar{V}}^{\infty} [V^i - (p_1 + t)]dF(V^i) \right\} + t \int_{\bar{V}}^{\infty} dF(V^i). \tag{8}$$

Differentiating this expression and using the envelope condition for $\bar{V}$, we see that the aggregate demand response remains a sufficient statistic for the marginal welfare gain:

$$\frac{dW}{dt} = -[1 - F(\bar{V})] + [1 - F(\bar{V})] + t\frac{d\int_{\bar{V}}^{\infty} dF(V^i)}{dt} = t\frac{dx_1}{dt}.$$

Intuitively, even though the individual demand functions are discontinuous, the model can be recast as the smooth choice of the demand threshold $\bar{V}$ by a representative agent. This permits application of the standard envelope condition to $\bar{V}$ in the social welfare function. At the microeconomic level, this envelope condition reflects the fact that the marginal individual who stops buying good x_1 when t is raised loses no utility, as he was indifferent about buying good x_1 to begin with.

The same logic applies in the J good case. To simplify exposition and link the results to standard multinomial logit discrete choice models, assume that ε_{ij} has a type 1 extreme value distribution.[11] Then it is well-known (e.g., see Train 2003) that the probability that a utility-maximizing individual i chooses product j is

$$P_{ij} = \frac{\exp(v_{ij})}{\sum_j \exp(v_{ij})} \tag{9}$$

and that agent i's expected utility from a vector of prices $p = (p_1, \ldots, p_J)$ is

$$S_i(p_1, \ldots, p_J) = E \, \max(u_{i1}, \ldots, u_{iJ}) = \log\left(\sum_j \exp v_{ij}\right).$$

Aggregating over the $i = 1, \ldots, N$ consumers, (expected) consumer surplus is

$$S = \sum_i \log\left(\sum_j \exp(v_{ij})\right).$$

---

[11]The results that follow do not rely on the assumption that the ε_{ij} errors have an extreme value distribution (Small & Rosen 1981). The distributional assumption simplifies the algebra by yielding a closed-form solution for total surplus, but the envelope conditions used to derive Equation 11 hold with any distribution.

Because utility is quasilinear, we can add producer profits to this expression to obtain social welfare:

$$W = \sum_i \log\left(\sum_j \exp(v_{ij})\right) + p \cdot P - c(P). \tag{10}$$

The classical approach to policy analysis in these models is to estimate the distributions of ϕ_i and ζ_j and simulate total surplus before and after a policy change (e.g., see Train 2003, p. 60). Identification of such models is challenging, especially if the econometrician does not observe all product attributes, as ζ_j will be correlated with p_j in equilibrium (Berry 1994, Berry et al. 1995).

Sufficient-statistic approaches offer a means of policy analysis that does not require identifying ϕ_i and ζ_j. For example, let us suppose the government levies a tax t on good 1, raising its price to $p_1 + t$. The government returns the proceeds to agents through a lump-sum transfer T so that y_i becomes $y_i + T$. As above, agents do not internalize the effects of their behavior on the size of the transfer T. Using the envelope condition for profit maximization,

$$\frac{dW(t)}{dt} = \sum_i \left[-\frac{\exp(v_{i1})}{\sum_j \exp(v_{ij})} - \sum_j \frac{dp_j}{dt} \frac{\exp(v_{ij})}{\sum_j \exp(v_{ij})} \right]$$
$$+ \sum_j \frac{dp_j}{dt} P_j + P_1 + t \frac{dP_1}{dt}$$
$$= t \frac{dP_1(t)}{dt}, \tag{11}$$

where the second equality follows from Equation 9. Identification of the welfare loss from taxation of good 1 requires estimation only of the effect of the tax on the aggregate market share $\left(\frac{dP_1}{dt}\right)$, as in the standard Harberger formula.

Now let us suppose that an ad-valorem tax τ is levied on all the products except the numeraire good, raising the price of product j to $(1 + \tau)p_j$. Again, tax revenue is returned to agents through a lump-sum grant. Following a similar derivation,

$$\frac{dW(\tau)}{d\tau} = \tau \sum_j p_j \frac{dP_j(\tau)}{d\tau} = \tau \frac{dE_P(\tau)}{d\tau},$$

where $E_p = \sum_j p_j P_j$ denotes total pretax expenditure in the market for the taxed good. The efficiency cost of a tax on all products depends on the aggregate expenditure elasticity for the taxed market; it does not require estimation of the substitution patterns within that market. A similar derivation can be used to show that the efficiency cost of a tax on a single characteristic, such as gas mileage, can be calculated simply by estimating the effect of the tax on the equilibrium quantity of that characteristic (e.g., gasoline consumption). Hence, many policy questions of interest can be answered simply by estimating reduced-form aggregate demand responses even in discrete-choice models.

The modern sufficient-statistic literature builds on Harberger's idea of identifying only the aspects of the model relevant for the question at hand. Before describing specific applications of this approach, I present a general framework that nests the papers in this literature and provides a recipe for developing such formulas.

3. GENERAL FRAMEWORK

Abstractly, many government policies amount to levying a tax t to finance a transfer $T(t)$. In the context of redistributive taxation, the transfer is to another agent; in the context of social insurance, it is to another state; and in the context of the excess-burden calculations above, the transfer can be thought of as being used to finance a public good. I now present a six-step rubric for calculating the welfare gain from raising the tax rate t [and the accompanying transfer $T(t)$] using sufficient statistics.

To simplify exposition, I formally present the rubric in a static model with a single agent. The same sequence of steps can be applied to obtain formulas for multiagent problems with heterogeneous preferences and discrete choice if $U(\cdot)$ is viewed as a (smooth) social-welfare function aggregating the utilities of all the agents, as in Equations 6 and 10. Similarly, dynamics can be incorporated by integrating the utility function over multiple periods.

3.1. Specify the Structure of the Model

Let $x = (x_1, \ldots, x_J)$ denote the vector of choices for the representative agent in the private sector. A unit tax t is levied on choice x_1 and the transfer $T(t)$ is paid in units of x_J. Let $\{G_1(x,t,T), \ldots, G_M(x,t,T)\}$ denote the $M < J$ constraints faced by the agent, which include budget constraints, restrictions on insurance or borrowing, hours constraints, and so on. The agent takes t and T as given and makes her choices by solving

$$\max\ U(x) \text{ s.t. } G_1(x,t,T) = 0, \ldots, G_M(x,t,T) = 0. \tag{12}$$

The solution to Equation 12 gives social welfare as a function of the policy instrument:

$$W(t) = \max_x U(x) + \sum_{m=1}^{M} \lambda_m G_m(x,t,T).$$

This specification nests competitive production because any equilibrium allocation can be viewed as the choice of a benevolent planner seeking to maximize total private surplus subject to technological constraints.

For example, in the single-agent Harberger model analyzed above,

$$\begin{aligned} U(x) &= u(x_1, \ldots, x_{J-1}) + x_J \\ G_1(x,t,T) &= T + Z - t_1 x_1 - c(x_1, \ldots, x_{J-1}) - x_J. \end{aligned} \tag{13}$$

A more general specification of preferences and constraints will yield a formula that is more robust but harder to implement empirically.

3.2. Express $\frac{dW}{dt}$ in Terms of Multipliers

Using the envelope conditions associated with optimization in the private sector, differentiate W to obtain

$$\frac{dW}{dt} = \sum_{m=1}^{M} \lambda_m \left\{ \frac{\partial G_m}{\partial T} \frac{dT}{dt} + \frac{\partial G_m}{\partial t} \right\}, \tag{14}$$

where λ_m denotes the Lagrange multiplier associated with constraint m in the agent's problem in Equation 12. In this equation, $\frac{dT}{dt}$ is known through the government's budget

constraint, and $\frac{\partial G_m}{\partial T}$ and $\frac{\partial G_m}{\partial t}$ can be calculated mechanically. For example, in the Harberger model, $T(t) = tx_1$ and hence $\frac{dT}{dt} = x_1 + t\frac{dx_1}{dt}$. Differentiating Equation 13 yields $\frac{dG_1}{dT} = 1$ and $\frac{dG_1}{dt} = -x_1$. It follows that $\frac{dW}{dt} = \lambda_1 t\frac{dx_1}{dt}$.

The critical unknowns are the λ_m multipliers. In the excess-burden application, λ_1 measures the marginal value of relaxing the budget constraint. In a social-insurance application, λ_1 could represent the marginal value of relaxing the constraint that limits the extent to which agents can transfer consumption across states. If λ_1 is small, there is little value to social insurance, whereas if it is large, $\frac{dW}{dt}$ could be large.

3.3. Substitute Multipliers by Marginal Utilities

The λ_m multipliers are recovered by exploiting restrictions from the agent's first-order-conditions. Optimization leads agents to equate marginal utilities with linear combinations of the multipliers:

$$u'(x_j) = -\sum_{m=1}^{M} \lambda_m \frac{\partial G_m}{\partial x_j}.$$

Inverting this system of equations generates a map from the multipliers into the marginal utilities. It is helpful to impose the following assumption on the structure of the constraints in order to simplify this mapping.

> **Assumption 1:** The tax t enters all the constraints in the same way as the good on which it is levied (x_1), and the transfer T enters all the constraints in the same way as the good in which it is paid (x_J). Formally, there exist functions $k_t(x,t,T)$, $k_T(x,t,T)$ such that
>
> $$\frac{\partial G_m}{\partial t} = k_t(x,t,T)\frac{\partial G_m}{\partial x_1} \forall m = 1,\dots,M$$
>
> $$\frac{\partial G_m}{\partial T} = -k_T(x,t,T)\frac{\partial G_m}{\partial x_J} \forall m = 1,\dots,M$$

Assumption 1 requires that x_1 and t enter every constraint interchangeably (up to a scale factor k_t).[12] That is, increasing t by \$1 and reducing x_1 by \$$k_t$ would leave all constraints unaffected. A similar interchangeability condition is required for x_J and T. In models with only one constraint per agent, Assumption 1 is satisfied by definition. In the Harberger model (in which the only constraint is the budget constraint), k_t corresponds to the mechanical increase in expenditure caused by a \$1 increase in t (\$$x_1$) versus a \$1 increase in x_1 (\$$p_1 + t$). Hence, $k_t = \frac{x_1}{p_1+t}$ in that model. Because increasing the transfer by \$1 affects the budget constraint in the same way as reducing consumption of x_J by \$1, $k_T = 1$.

Models in which the private-sector choices are second-best efficient subject to the resource constraints in the economy typically satisfy the conditions in Assumption 1. This is because the fungibility of resources ensures that the taxed good and tax rate enter all constraints in the same way (see Chetty 2006a for details). The sufficient-statistic approach can be implemented in models that violate Assumption 1 (see Section 5 for an

---

[12]If the tax t is levied on multiple goods $(x_1,\dots,x_t)$ as in Feldstein (1999), the requirement is that it enters the constraints in the same way as the combination of all the taxed goods, i.e., $\frac{\partial G_i}{\partial t} = \sum_{i=1}^{t} k_t(x,t,T)\frac{\partial G_i}{\partial x_i}$.

example), but the algebra is much simpler when this assumption holds. This is because the conditions in Assumption 1 permit direct substitution into Equation 14 to obtain

$$\frac{dW}{dt} = \sum_{m=1}^{M} \lambda_m \left\{ -k_T \frac{\partial G_m}{\partial x_J} \frac{dT}{dt} + k_t \frac{\partial G_m}{\partial x_1} \right\} = -k_T \frac{dT}{dt} \sum_{m=1}^{M} \lambda_m \frac{\partial G_m}{\partial x_J} + k_t \sum_{m=1}^{M} \lambda_m \frac{\partial G_m}{\partial x_1}$$

$$\frac{dW}{dt} = k_T \frac{dT}{dt} u'(x_J(t)) - k_t u'(x_1(t)). \tag{15}$$

This expression captures a simple and general intuition: Increasing the tax t is equivalent to reducing consumption of x_1 by k_t units, which reduces the agent's utility by $k_t u'(x_1(t))$. The additional transfer that the agent gets from the tax increase is $\frac{dT}{dt} k_T$ units of good x_J, which raises his utility by $k_T \frac{dT}{dt} u'(x_J(t))$. Because k_T, k_t, and $\frac{dT}{dt}$ are known based on the specification of the model, this expression distills local welfare analysis to recovering a pair of marginal utilities.[13]

In models with heterogeneity, the aggregate welfare gain is a function of a pair of average marginal utilities across agents. In dynamic models, the welfare gain is also a function of a pair of average marginal utilities, but with the mean taken over the lifecycle for a given agent. This result is obtained using envelope conditions when differentiating the value function.

3.4. Recover Marginal Utilities from Observed Choices

The final step in obtaining an empirically implementable expression for $\frac{dW}{dt}$ is to back out the two marginal utilities. One way to do this is to make assumptions about the relevant marginal utilities based on surveys or external evidence, such as measures of the value of an additional year of healthy life. This is implicitly the approach taken in the cost-benefit analyses sometimes reported in reduced-form studies (e.g., Levitt 1997).

Sufficient-statistic studies recover the marginal utilities from choice data using the structure of the model specified in step 1. There is no canned procedure for this step. The applications below provide several illustrations. The trick that is typically exploited is that the marginal utilities are elements in first-order conditions for various choices. As a result, they can be backed out from the comparative statics of behavior. For instance, in the single-agent Harberger model above, the assumption of no income effects implies $u'(x_J) = 1$. To identify $u'(x_1)$, exploit the first-order condition for x_1, which is $u'(x_1) = p_1 + t$. Plugging in these expressions and the other parameters above into Equation 15, we obtain Equation 4:

$$\frac{dW(t)}{dt} = 1 \cdot (x_1 + t\frac{dx_1}{dt}) - \frac{x_1}{p_1 + t} \cdot (p_1 + t) = t\frac{dx_1(t)}{dt}.$$

3.5. Empirical Implementation

Let us suppose the sufficient-statistic formula one derives has the following form:

$$\frac{dW}{dt}(t) = f\left(\frac{\partial x_1}{\partial t}, \frac{\partial x_1}{\partial Z}, t\right). \tag{16}$$

When implementing this expression empirically, researchers should keep two issues in mind. First, the relevant derivatives may require holding different variables fixed depending upon

---

[13]In many applications, steps 2 and 3 are consolidated into a single step because the constraints can be substituted directly into the objective function.

the application. For example, the Harberger formula in Equation 4 calls for the measurement of the total derivative $\frac{dx_1}{dt}$, which incorporates general equilibrium effects and price changes in all markets. In contrast, many reduced-form empirical studies explicitly seek to identify the partial derivative $\frac{\partial x_1}{\partial t}$, holding prices in other markets constant. These studies typically compare the behavior of a small group of individuals treated by a tax change with other unaffected individuals. Such studies do not recover the sufficient statistic of interest for the Harberger policy question. The elasticities they estimate, however, are directly relevant for assessing the efficiency cost of a tax increase on a small subset of consumers that does not affect equilibrium prices in other markets. The general lesson is that the experiment used to identify the relevant elasticities must be matched to the policy question being asked.[14] In situations in which one cannot credibly identify the elasticity called for by the formula, it may be possible to make progress using approximations (e.g., a tax levied on a small market has negligible effects on prices in other markets in equilibrium).

The second issue arises from the fact that policy changes of interest are never infinitesimal. The ideal way to implement Equation 16 to assess the efficiency costs of a discrete policy change is to estimate the inputs as nonparametric functions of the policy instrument t. With estimates of $\frac{\partial x_1}{\partial t}(t)$ and $\frac{\partial x_1}{\partial Z}(t)$, one can integrate Equation 16 between any two tax rates t_1 and t_2 that lie within the support of observed policies to evaluate the welfare gain ΔW for a policy change of interest. This procedure is similar in spirit to Heckman & Vytlacil's (2001, 2005) recommendation that researchers estimate a complete schedule of MTEs and then integrate that distribution over the desired range to obtain policy-relevant treatment effects. In the present case, the marginal welfare gain at t depends on the MTE at t; the analysis of nonmarginal changes requires estimation of the MTE as a function of t. Predictions about welfare gains from policies outside the observed support can be made by extrapolating $\frac{\partial x_1}{\partial t}(t)$ out of sample. That is, one can effectively obtain sufficient statistics for out-of-sample welfare analysis under assumptions about the functional form of $\frac{\partial x_1}{\partial t}(t)$.

In most applications, there is insufficient power to estimate $x_1(t)$ nonparametrically. Instead, typical reduced-form studies estimate local average treatment effects (Angrist et al. 2000), such as the effect of a discrete change in the tax rate from t_1 to t_2 on demand: $\frac{\Delta x_1}{\Delta t} = \frac{x_1(t_2)-x_1(t_1)}{t_2-t_1}$. The estimate of $\frac{\Delta x_1}{\Delta t}$ permits inference about the effect of raising the tax rate from t_1 to t_2 on welfare. To see this, consider the Harberger model, in which $\frac{dW}{dt}(t) = t\frac{dx_1}{dt}(t)$. A researcher who has estimated $\frac{\Delta x_1}{\Delta t}$ has two options. This first is to bound the average welfare gain over the observed range:

$$W(t_2) - W(t_1) = \int_{t_1}^{t_2}\frac{dW}{dt}\,dt = \int_{t_1}^{t_2}t\frac{dx_1}{dt}(t)dt \Rightarrow t_1\frac{\Delta x_1}{\Delta t} > \overline{dW/dt} > t_2\frac{\Delta x_1}{\Delta t}. \qquad (17)$$

Intuitively, the excess burden of taxation depends on the slope of the $x_1(t)$ between t_1 and t_2, multiplied by the height of the Harberger trapezoid at each point. When one observes only the average slope between the two tax rates, bounds on excess burden can be obtained by setting the height to the lowest and highest points over the interval.

---

[14]A related problem is that reduced-form studies often estimate the combined effect of multidimensional changes, such as welfare reforms that affect both time limits and financial incentives to work. Such estimates cannot be plugged directly into sufficient-statistic formulas. In such cases, structure may be needed to obtain elasticities with respect to each dimension of the change.

The second option is to use an approximation to $x_1(t)$ to calculate $\overline{dW/dt}$. For instance, if one can estimate only the first-order properties of $x_1(t)$, making the approximation that $\frac{dx_1}{dt}$ is constant over the observed range implies

$$\overline{dW/dt} \simeq \frac{t_1 + t_2}{2}\frac{\Delta x_1}{\Delta t}.$$

If $x_1(t)$ is linear, the average height of the trapezoid and $\frac{\Delta x_1}{\Delta t}$ exactly determine excess burden. If one has adequate data and variation to estimate the higher-order terms of $x_1(t)$, these estimates can be used to fit a higher-order approximation to $x_1(t)$ to obtain a more accurate estimate of $\overline{dW/dt}$.

The same two options are available in models in which $\frac{dW}{dt}$ is a function of more than one behavioral response, as in Equation 16. Bounds may be obtained using the estimated treatment effects $\left(\frac{\Delta x_1}{\Delta Z}, \frac{\Delta x_1}{\Delta t}\right)$ by integrating $\frac{dW}{dt}$ and setting the other parameters at their extrema as in Equation 17. Under a linear approximation to demand $\left(\frac{dx_1}{dt}, \frac{dx_1}{dZ} \text{ constant}\right)$, treatment effects can be mapped directly into the marginal welfare gain: $\frac{dW(t)}{dt} = f(\frac{\Delta x_1}{\Delta t}, \frac{\Delta x_1}{\Delta Z}, t)$. If $\frac{dW(t)}{dt}$ can only be estimated accurately at the current level of t, one can at least determine the direction in which the policy instrument should be shifted to improve welfare.

The bottom line is that the precision of a sufficient-statistic formula is determined by the precision of the information available about the sufficient statistics as a function of the policy instrument. In the applications discussed below, the data and variation available only permit estimation of first-order properties of the inputs, and the authors are therefore constrained to calculating a first-order approximation of $\overline{dW/dt}$. The potential error in this linear approximation can be assessed using the bounds proposed above or using a structural model.

3.6. Model Evaluation

Although sufficient-statistic formulas do not require full specification of the model, they do require some modeling assumptions; it is impossible to make theory-free statements about welfare. It is important to assess the validity of these assumptions to ensure that the formula's results are accurate. Unfortunately, because this step is not necessary to calculate $\frac{dW(t)}{dt}$, it is often not implemented in existing sufficient-statistic studies.

The model can be evaluated in two ways. First, one can test qualitative predictions that would falsify the assumptions that are central for deriving the sufficient-statistic formula. For example, Harberger's formula assumes that individuals treat prices and taxes identically, making choices based only on the total price of the good $(p + t)$. This assumption can be tested by comparing price and tax elasticities of demand (Chetty et al. 2009). Second, one should identify at least one vector of structural parameters ω that is consistent with the sufficient statistics estimated in step 5. If the empirical estimates of the sufficient statistics are internally consistent with the model, at least one ω must fit the estimated statistics. This is not a stringent test, because the idea of the sufficient-statistic approach is that there will be multiple values of ω consistent with the sufficient statistics. However, there are cases in which the estimated high-level elasticities may not be consistent with any underlying vector of primitives, rejecting the assumptions of the model (see Chetty 2006a,b for an example).

The next three sections show how a variety of recent papers in public economics can be interpreted as applications of this six-step rubric. Each application illustrates different

strengths and weaknesses of the sufficient-statistic approach and demonstrates the techniques that are helpful in deriving such formulas.

4. APPLICATION 1: INCOME TAXATION

Since the seminal work of Mirrlees (1971) and others, there has been a large structural literature investigating the optimal design of income-tax and transfer systems. Several studies have simulated optimal tax rates in calibrated versions of the Mirrlees model (see Tuomala 1990 for a survey). A related literature uses microsimulation methods to calculate the effects of changes in transfer policies on behavior and welfare. The most recent structural work in this area has generalized the Mirrlees model to dynamic settings and simulated the optimal design of tax policies in such environments using calibrated models. Parallel to this literature, a large body of work in labor economics has investigated the effects of tax and transfer programs on behavior using program-evaluation methods (see **Table 1** for examples of structural and reduced-form studies).

Recent work in public economics has shown that the elasticities estimated by labor economists can be mapped into statements about optimal tax policy in the models that have been analyzed using structural methods. This sufficient-statistic method has been widely applied in the context of income taxation in the past decade, with contributions by Feldstein (1995, 1999), Piketty (1997), Diamond (1998), Saez (2001), Gruber & Saez (2002), Goulder & Williams (2003), Chetty (2009), and others. All these papers can be embedded in the general framework proposed above. I focus on two papers here in the interest of space: Feldstein (1999) and Saez (2001).

4.1. Feldstein (1999)

Traditional empirical work on labor supply did not incorporate the potential effects of taxes on choices other than hours of work. For instance, income taxes could affect an individual's choice of training, effort, or occupational choice. Moreover, individuals may be induced to shelter income from taxation by evading or avoiding tax payments (e.g., taking fringe benefits, underreporting earnings). Although some studies have attempted to directly examine the effects of taxes on each of these margins, it is difficult to account for all potential behavioral responses to taxation by measuring each channel separately.

Feldstein proposed an elegant solution to the problem of calculating the efficiency costs of taxation in a model with multidimensional labor-supply choices. His insight is that the elasticity of taxable income with respect to the tax rate is a sufficient statistic for calculating deadweight loss. Feldstein considered a model in which an individual makes J labor-supply choices $(x_1, \ldots, x_J)$ that generate earnings. Let w_j denote the wage paid for choice j and $\psi_j(x_j)$ denote the disutility of labor supply through margin x_j.[15] In addition, let us suppose that the agent can shelter e of earnings from the tax authority (via sheltering or evasion) by paying a cost $g(e)$. Total taxable income is $TI = \sum_{j=1}^{J} w_j x_j - e$. Let

---

[15]Feldstein takes the wage rates w_j as fixed, implicitly assuming that the demand for each type of labor supply is perfectly elastic. Allowing for downward-sloping labor demand curves in a competitive market does not affect the formula he derives for the same reason that endogenous prices do not affect the Harberger formula in Equation 7. With endogenous wages, the sufficient statistic for deadweight loss is the total derivative $\frac{dTI}{dt}$, which incorporates all equilibrium wage responses to the tax change.

$c = (1 - t)TI + e$ denote consumption. For simplicity, assume that utility is linear in c to abstract from income effects. Feldstein demonstrated that it is straightforward to allow for income effects. As in the Harberger model, we calculate the excess burden of the tax by assuming that the government returns the tax revenue to the agent as a lump-sum transfer $T(t)$. Using the notation introduced in Section 3, we can write this model formally as

$$u(c, x, e) = c - g(e) - \sum_{j=1}^{J} \psi_j(x_j)$$
$$T(t) = t \cdot TI$$
$$G_1(c, x, t) = T + (1 - t)TI + e - c.$$

Social welfare is

$$W(t) = \left\{ (1 - t)TI + e - g(e) - \sum_{j=1}^{J} \psi_j(x_j) \right\} + t \cdot TI. \tag{18}$$

To calculate the marginal excess burden $\frac{dW}{dt}$, totally differentiate Equation 18 to obtain

$$\frac{dW}{dt} = TI + t\frac{dTI}{dt} - TI + (1 - t)\frac{dTI}{dt} + \frac{de}{dt}(1 - g'(e)) - \sum_{j=1}^{J} \psi_j'(x_j)\frac{dx_j}{dt}$$

$$= \frac{dTI}{dt} + \frac{de}{dt}(1 - g'(e)) - \sum_{j=1}^{J} \psi_j'(x_j)\frac{dx_j}{dt}. \tag{19}$$

This equation is an example of the marginal utility representation in Equation 15 given in step 3 of the rubric in Section 3. To recover the marginal utilities (step 4), Feldstein exploited the first-order conditions

$$g'(e) = t$$
$$\psi_j'(x_j) = (1 - t)w_j \Rightarrow \sum_{j=1}^{J} \psi_j'(x_j)\frac{dx_j}{dt} = \sum_{j=1}^{J}(1 - t)w_j\frac{dx_j}{dt} = (1 - t)\frac{d(TI + e)}{dt}, \tag{20}$$

where the last equality follows from the definition of TI. Plugging these expressions into Equation 19 and collecting terms yield the following expression for the marginal welfare gain from raising the tax rate from an initial rate of t:

$$\frac{dW(t)}{dt} = t\frac{dTI(t)}{dt}. \tag{21}$$

A simpler, but less instructive, derivation of Equation 21 is to differentiate Equation 18, recognizing that behavioral responses have no first-order effect on private surplus (the term in curly brackets) because of the envelope conditions. This immediately yields $\frac{dW}{dt} = -TI + TI + t\frac{dTI}{dt}$.

Equation 21 shows that we simply need to measure how taxable income responds to changes in the tax rate to calculate the deadweight cost of income taxation. There is no need to determine whether TI changes because of hours responses, changes in occupation, or avoidance behaviors to calculate efficiency costs. Intuitively, the agent supplies labor on every margin $(x_1, \ldots, x_J)$ up to the point at which his marginal disutility of earning another dollar through that margin equals $1 - t$. The marginal social value of earning an

extra dollar net of the disutility of labor is therefore t for all margins. Likewise, the agent optimally sets the marginal cost of reporting 1 less to the tax authority [$g'(e)$] equal to the marginal private value of doing so (t). Hence, the marginal social costs of reducing earnings (via any margin) and reporting less income via avoidance are the same at the individual's optimal allocation. This makes it irrelevant which mechanism underlies the change in *TI* for efficiency purposes.

The main advantage of identifying $\frac{dTI(t)}{dt}$ as a sufficient statistic is that it permits inference about efficiency costs without requiring the identification of the potentially complex effects of taxes on numerous labor supply, evasion, and avoidance behaviors. Moreover, data on taxable income are available on tax records, facilitating the estimation of the key parameter $\frac{dTI}{dt}$. Feldstein (1995) implemented Equation 21 by estimating the changes in reported taxable income around the Tax Reform Act of 1986, implicitly using the linear approximation described in step 5 of the rubric. He concluded based on these estimates that the excess burden of taxing high-income individuals is very large, possibly as large as \$2 per \$1 of revenue raised. This result has been influential in policy discussions by suggesting that top income-tax rates should be lowered (e.g., see Joint Econ. Comm. 2001). Subsequent empirical work motivated by Feldstein's result has found smaller values of $\frac{dTI}{dt}$, and the academic debate about the value of the taxable income elasticity remains active.

The sixth step of the rubric, model evaluation, has only been partially implemented in the context of Feldstein's formula. Slemrod (1995) and several other authors have found that the large estimates of $\frac{dTI}{dt}$ are driven primarily by evasion and avoidance behaviors $\left(\frac{de}{dt}\right)$. However, these structural parameters [$g(e), \psi_j(x_j)$] of the model have not been evaluated directly. Chetty (2009) gave an example of the danger in not investigating the structural parameters. He argued that the marginal social cost of tax avoidance may not be equal to the tax rate at the optimum, violating the first-order condition (Equation 20) that is critical to derive Equation 21, for two reasons. First, some of the costs of evasion and avoidance constitute transfers, such as the payment of fines for tax evasion, rather than resource costs. Second, there is considerable evidence that individuals overestimate the true penalties for evasion. Using a sufficient-statistic approach analogous to that above, Chetty relaxed the $g'(e) = t$ restriction and obtained the following generalization of Feldstein's formula:

$$\frac{dW(t)}{dt} = t\left\{\mu(t)\frac{dTI(t)}{dt} + (1 - \mu(t))\frac{dLI(t)}{dt}\right\}, \qquad (22)$$

where $LI = \sum_{j=1}^{J} w_j x_j$ represents total earned income and $\mu(t) = \frac{g'(e(t))}{t}$ measures the gap between social marginal costs of avoidance and the tax rate. Intuitively, deadweight loss is a weighted average of the taxable income elasticity $\left(\frac{dTI}{dt}\right)$ and the total earned income elasticity $\left(\frac{dLI}{dt}\right)$, with the weight determined by the resource cost of sheltering. If avoidance does not have a large resource cost, changes in e have little efficiency cost, and thus it is only $\frac{dLI}{dt}$—the real labor-supply response—that matters for deadweight loss.

Not surprisingly, implementing Chetty's more general formula requires the identification of more parameters than Feldstein's formula. The most difficult parameter to identify is $g'(e)$, which is a marginal utility. By leaving $g'(e)$ in the formula, Chetty does not complete step 4 of the rubric above; as a result, further work is required to implement Equation 22. Gorodnichenko et al. (2009) provided a method of recovering $g'(e)$ from consumption behavior. Their insight is that real resource costs expended on evasion should

be evident in consumption data; thus, the gap between income and consumption measures can be used to infer $g'(e)$. Implementing this method to analyze the efficiency costs of a large reduction in income-tax rates in Russia, these authors found that $g'(e)$ is quite small and that $\frac{dTI}{dt}$ is substantial, whereas $\frac{dLI}{dt}$ is not. They showed that Feldstein's formula substantially overestimates the efficiency costs of taxation relative to Chetty's more general measure. Intuitively, reported taxable incomes are highly sensitive to tax rates, but the sensitivity is driven by avoidance behavior that has little social cost at the margin and hence does not reduce the total size of the pie significantly.

This literature underscores the point that sufficient-statistic approaches are not model free. It is critical to evaluate the structure of the model, even though the formula for $\frac{dW}{dt}$ can be implemented without the last step of the rubric. In the taxable-income application, estimating $g'(e)$ has value instead of simply assuming that $g'(e) = t$, given plausible concerns that this condition does not hold in practice.

4.2. Saez (2001)

Harberger and Feldstein studied the efficiency effects and optimal design of a linear tax. Much of the literature on optimal income taxation has focused on nonlinear income-tax models and the optimal progressivity of such systems. Mirrlees (1971) formalized this question as a mechanism design problem and provided a solution in differential equations that are functions of primitive parameters. The Mirrlees solution offers little intuition into the forces that determine optimal tax rates. Building on the work of Diamond (1998), Saez (2001) expressed the optimality conditions in the Mirrlees model in terms of empirically estimable sufficient statistics.

Saez analyzed a model in which individuals choose hours of work, l, and have heterogeneous wage rates w distributed according to a distribution $F(w)$. Wage rates (skills) are unobservable to the government. Let pretax earnings be denoted by $z = wl$. For simplicity, I again restrict attention to the case without income effects, as in Diamond (1998).

Saez began by analyzing the optimal tax rate on top incomes. He considered a model in which the government levies a linear tax τ on earnings above a threshold $\bar{z}$ and characterizes the properties of the optimal tax rate τ^* as $\bar{z} \to \infty$. For a given $\bar{z}$, individuals maximize utility

$$u(c, l) = c - \psi(l)$$
$$\text{s.t. } G_1(c, l) = (1 - \tau)\max(wl - \bar{z}, 0) + \bar{z} - c = 0,$$

where $c(w, \tau)$ and $l(w, \tau)$ denote an agent's optimal choices as a function of his wage and the tax rate, and $z(w, \tau) = wl(w, \tau)$ denote the optimized earnings function. Let $z_m(\bar{z}) = E[wl(w, \tau)|z(w, \tau) > \bar{z}]$ denote the mean level of earnings for individuals in the top bracket. Let $\bar{w}$ represent the wage threshold that corresponds to an earnings threshold of $\bar{z}$ when the tax rate is τ: $\bar{w}l(\bar{w}, \tau) = \bar{z}$. The tax revenue generated by the top bracket tax is $R = \tau(z_m(\bar{z}) - \bar{z})$. The planner uses this tax revenue to fund a project that has a (normalized) value of \$1 per dollar spent.

The social planner's objective is to maximize a weighted average of individual's utilities, where the weights $\tilde{G}(u)$ are social-welfare weights that reflect the redistributive preferences of the planner:

$$W = \left\{ \int\limits_0^\infty \tilde{G}(u(c(w,\tau), wl(w,\tau)))dF(w) \right\} + \tau(z_m(\bar{z}) - \bar{z}).$$

In this equation, the first term (in curly brackets) represents private surplus and the second term reflects government revenue. To calculate $\frac{dW}{d\tau}$, we observe that individuals with incomes below $\bar{z}$ are unaffected by the tax increase. We normalize the measure of individuals in the top bracket to 1. Utility maximization implies that behavioral responses $\left(\frac{\partial l}{\partial \tau}\right)$ have no first-order effect on private surplus, as $wu_c(w,\tau) = \psi'(l(w,\tau))$. Using this envelope condition, we obtain

$$\frac{dW}{d\tau}(\tau) = -\int\limits_{\bar{w}}^\infty \tilde{G}_u(u)(z(w,\tau) - \bar{z})dF(w) + \left[(z_m - \bar{z}) + \tau\frac{dz_m}{d\tau} \right]$$

$$= -(z_m(\bar{z}) - \bar{z})\bar{g} + \left[(z_m(\bar{z}) - \bar{z}) + \tau\frac{dz_m}{d\tau} \right], \qquad (23)$$

where $\bar{g} = \int_{\bar{w}}^\infty \tilde{G}_u(u)(z - \bar{z})dF(w) / \int_{\bar{w}}^\infty (z - \bar{z})dF(w)$ denotes the mean marginal social-welfare weight placed by the planner on individuals in the top tax bracket. The parameter $\bar{g}$ measures the social value of giving \$1 more income to individuals in the top bracket relative to the value of public expenditure. If $\bar{g} = 1$, the government weighs the consumption of the individuals it taxes and public expenditure equally, and Equation 23 collapses to the Harberger formula for excess burden in Equation 4. When $\bar{g} < 1$, the first term in Equation 23 captures the welfare loss to individuals in the top tax bracket from having to pay more taxes. The second term reflects the gain in revenue to the government, which consists of two familiar components: the mechanical gain in revenue and the offset due to the behavioral response.

Equation 23 illustrates that three parameters are together sufficient statistics for the welfare gain of increasing top income-tax rates: (a) the effect of tax rates on earnings $\left(\frac{dz_m}{d\tau}\right)$, which quantifies the distortions created by the tax; (b) the shape of the earnings distribution $[z_m(\bar{z})]$, which measures the mass of individuals whose behavior is distorted by the marginal tax, and (c) the marginal social-welfare weight ($\bar{g}$), which measures the planner's redistributive preferences. We note that Saez does not implement step 4 of the rubric (recovery of marginal utilities from observed behavior) because he views the relevant marginal utility in this case ($\bar{g}$) as a feature of the planner's social-welfare function that is external to the choice environment. Thus, $\bar{g}$ is determined by the shape of the earnings distribution and the (exogenous) specification of the social-welfare function (e.g., utilitarian or Rawlsian).

The advantage of Equation 23 relative to a structural approach is that one does not need to identify preferences (ψ) or the shape of the skill distribution $F(w)$ to calculate $dW/d\tau$. Moreover, one can permit arbitrary heterogeneity across skill types in preferences without changing the formula. The disadvantage of Equation 23 is that $z_m, \bar{g}$, and $\frac{dz_m}{d\tau}$ are endogenous to τ: The level of earnings and the weight the social planner places on top earners presumably decrease with τ, whereas $\frac{dz_m}{d\tau}$ may vary with τ depending upon the shape of the $\psi(l)$ function. Hence, $\frac{dW}{d\tau}(\tau)$ measures only the marginal welfare gain at a given tax rate τ and must be estimated at all values of τ to calculate the tax rate τ^* that maximizes W. To simplify empirical implementation and derive an explicit formula for the optimal tax rate, Saez observed that the ratio $\frac{z_m(\bar{z})}{\bar{z}}$ is approximately constant in the upper

tail of the empirical distribution of earnings in the United States: That is, the upper tail of the income distribution is well described by a Pareto distribution. A Pareto distribution with parameter a has $\frac{z_m(\bar{z})}{\bar{z}} = \frac{a}{a-1}$ for all $\bar{z}$. Hence, Equation 23 can be expressed as

$$\frac{d\mathrm{W}}{dt} = \frac{(1 - \bar{g})}{a - 1}\bar{z} + \tau\frac{dz_m}{d\tau}.$$

The optimal top-bracket tax rate τ satisfies $\frac{d\mathrm{W}}{d\tau}(\tau) = 0$, implying

$$\frac{\tau^*}{1 - \tau^*} = \frac{1 - \bar{g}}{a\varepsilon}, \tag{24}$$

where $\varepsilon = \frac{dz_m}{d(1-\tau)}\frac{1-\tau}{z_m}$ denotes the taxable income elasticity in the top bracket. In the Mirrlees model, a and ε converge to constants (invariant to τ) in the limit as $\bar{z} \to \infty$. Equation 24 is therefore an explicit formula for the optimal asymptotic top income-tax rate if the social-welfare weight $\bar{g}$ is taken as exogenous. For example, one plausible assumption is that $\bar{g} \to 0$ as $\bar{z} \to \infty$. Saez exploited this property of Equation 24 to calculate optimal top income-tax rates using reduced-form estimates of the taxable income elasticity for high incomes (Gruber & Saez 2002) and a Pareto parameter of $a = 2$ consistent with the earnings distribution in the United States. He found that optimal top income-tax rates are generally above 50% when the formula is calibrated using plausible elasticities.

Building on this sufficient-statistic approach, Saez characterized the optimal tax rate at any income level z in a nonlinear tax system. Let E denote the fixed amount of government expenditure that must be financed through taxation, and $T(z)$ denote the total tax paid by an individual who earns income z, so that net of tax income is $z - T(z)$. Let $\varepsilon(z) = \frac{dz}{d(1-\tau)}\frac{1-\tau}{z}$ denote the earnings elasticity at income level z and $h(z)$ be the density of the earnings distribution at z. Finally, $\tilde{G}(u(z))$ denotes the weight that the planner places on an individual with earnings z and $g(z) = \tilde{G}_u \cdot u_c(z)$ is the corresponding marginal social-welfare weight.

The government chooses the schedule $T(Z)$ that maximizes social welfare

$$W(T(z)) = \int\limits_0^\infty \tilde{G}(u(c(w, T), wl(w, T)))dF(w)$$

subject to resource and incentive-compatibility constraints:

$$G_1(c, z, T) = \int\limits_0^\infty z(w, T)dF(w) - \int\limits_0^\infty c(w, T)dF(w) - E = 0$$

$$G_2(c, z, T) = (1 - T'(z))w - \psi'(l(w)) = 0.$$

Exploiting envelope conditions and perturbation arguments as above, the first-order conditions for the optimal tax rates can be expressed in terms of sufficient statistics. In the case without income effects, the optimal tax schedule satisfies the following condition at all z:

$$\frac{T(z)}{1 - T(z)} = \frac{1}{\varepsilon(z)zh(z)}\int\limits_z^\infty (1 - g(z'))h(z')dz'. \tag{25}$$

Equation 25 depends on the same three parameters as Equation 23: the taxable income elasticity, the shape of the earnings distribution, and the social-welfare weights. It is again important to recognize that all three of these parameters are endogenous to the tax regime

itself, and hence Equation 25 is not an explicit formula for optimal taxation. The techniques used to obtain the explicit formula for the asymptotic top income-tax rate in Equation 24 cannot be applied at an arbitrary income level z because there are no analogous limit convergence results. Hence, Equation 25 can only be used to evaluate how perturbing the existing tax system $T(z)$ would affect welfare.

To go further and calculate the optimal tax schedule, Saez made additional assumptions about the model's structure. He assumed that the elasticity $\varepsilon(z)$ is constant, which pins down the functional form of the utility $u(c, l)$. Given this utility function and an elasticity estimate taken from the literature, he inferred the skill distribution $F(w)$ from the empirically observed distribution of incomes in the current tax regime. Having identified all the primitives of the model, he simulated the optimal tax schedule in the calibrated model. The resulting optimal income-tax schedule is inverse-U shaped, with a large lump-sum grant to nonworkers and marginal rates ranging from 50% to 80%.

This exercise illustrates the costs and benefits of placing more structure on the problem. By assuming a one-parameter constant-elasticity utility, one can point identify the primitives from the data and calculate the optimal tax schedule. However, the strong assumptions required for structural identification of the model reduce our confidence in the optimal tax-schedule calculations. We have greater confidence in the top tax-rate calculations based on the sufficient-statistic formula in Equation 24 or marginal welfare-gain calculations based on Equation 25.

5. APPLICATION 2: SOCIAL INSURANCE

Programs such as unemployment insurance (UI), health insurance, social security, workers compensation, and disability insurance account for the majority of government expenditure in many countries. Starting with seminal contribution of Wolpin (1987), a large literature has studied the optimal design of social-insurance programs in dynamic structural models. Parallel to this literature, a large body of reduced-form empirical work has investigated the impacts of social-insurance programs on health expenditures, unemployment durations, consumption, disability claims, and so on (see **Table 1** for examples of structural and reduced-form studies).

In the context of social insurance, an important harbinger to the sufficient-statistic approach is the work of Baily (1978), who showed that the optimal level of unemployment benefits can be expressed as a function of a small set of parameters in a static model. Baily's result was viewed as having limited practical relevance because of the strong assumptions made in deriving the formula. However, recent work has shown that the parameters Baily identified are actually sufficient statistics for the welfare analysis of social insurance in a rich class of dynamic models. Studies in this literature include Gruber (1997), Chetty (2006a), Shimer & Werning (2007), Chetty (2008), Einav et al. (2008a), and Chetty & Saez (2008). I now embed these papers in the general framework above, focusing primarily on the first four papers.[16]

---

[16]The tax and social-insurance problems are closely related because social insurance is effectively state-contingent taxation. Rather than levying taxes on the basis of income, taxes and transfers are levied on the basis of a state (e.g., joblessness, sickness, injury, and disability). Conversely, redistributive taxation is social insurance against uncertain skill realizations behind the veil of ignorance.

5.1. Baily (1978) and Chetty (2006a)

For simplicity, consider a static model with two states: high and low. Let w_h denote the individual's income in the high state, and $w_l < w_h$ is income in the low state. Let A denote wealth, c_h consumption in the high state, and c_l consumption in the low state. The low state can be thought of as corresponding to job loss, injury, disability, natural disaster, and so on. The agent can control the probability of being in the bad state by exerting effort e at a cost $\psi(e)$. For instance, effort could reflect searching for a job, taking precautions to avoid injury, or locating a house away from areas prone to natural disasters. Choose units of e so that the probability of being in the high state is given by $p(e) = e$.

Individuals may have some ability to insure against shocks through informal private-sector arrangements, such as transfers between relatives. To model such informal private-insurance arrangements, we suppose that the agent can transfer $\$b_p$ between states at a cost $q(b_p)$, so that increasing consumption by b_p in the low state requires payment of a premium $\frac{1-e}{e} b_p + q(b_p)$ in the high state. The loading factor $q(b_p)$ can be interpreted as the degree of incompleteness in private insurance. If $q(b_p) = 0$, private-insurance markets are complete; if $q(b_p) = \infty$, there is no capacity for private insurance.

The government pays a benefit b in the low state that is financed by an actuarially fair tax $t(b) = \frac{1-e}{e} b$ in the high state. The model can be formally specified using the notation in Section 3 as follows:[17]

$$U(c_l, c_h, e) = eu(c_h) + (1-e)u(c_l) - \psi(e), \ t(b) = \frac{1-e}{e} b$$

$$G_1(c_l, c_h, t) = c_h + \frac{1-e}{e} b_p + q(b_p) + t - w_h - A$$

$$G_2(c_l, c_h, t) = c_l - b_p - b - w_l - A. \tag{26}$$

Substituting the constraints into the utility function yields social welfare as a function of the government benefit level:

$$W(b) = eu\left(A + w_h - \frac{1-e}{e} b_p - q(b_p) - t(b)\right) + (1-e)u(A + w_l + b_p + b) - \psi(e).$$

Differentiating this expression and using the envelope conditions for b_p and e give

$$\frac{dW(b)}{db} = (1-e)u'(c_l) - \frac{dt}{db} eu'(c_h) = (1-e)\left\{ u'(c_l) - \left(1 + \frac{\varepsilon_{1-e,b}}{e}\right) u'(c_h) \right\},$$

where $\varepsilon_{1-e,b} = \frac{d(1-e)}{db} \frac{b}{1-e}$ denotes the elasticity of the probability of being in the bad state (which can be measured as the unemployment rate, rate of health insurance claims, etc.) with respect to the benefit level. This elasticity measures the total effect of an increase in benefits on e, taking into account the tax increase needed to finance the higher level of benefits.

In tax models with quasilinear utility, the welfare gain measure $\frac{dW}{dt}$ is a money metric. Because the curvature of utility is an essential feature of the social-insurance problem, we need a method of converting $\frac{dW}{db}$ to a money metric. An intuitive metric is to normalize the

---

[17]I follow the convention in the social-insurance literature of specifying the problem in terms of the transfer benefit b rather than the tax t.

welfare gain from a 1 (balanced budget) increase in the size of the government insurance program by the welfare gain from raising the wage bill in the high state by 1:

$$M_W(b) = \frac{\frac{dW}{db}(b)/(1-e)}{\frac{dW}{dw_h}(b)/e} = \frac{u'(c_l) - u'(c_h)}{u'(c_h)} - \frac{\varepsilon_{1-e,b}}{e}. \tag{27}$$

This expression, which is Baily's (1978) formula, corresponds to the marginal utility expression obtained after the third step of the rubric in Section 3. The first term in Equation 27 measures the gap in marginal utilities between the high and low states, which quantifies the welfare gain from transferring an additional dollar from the high to low state. The second term measures the cost of transferring this additional dollar due to behavioral responses.

Chetty (2006a) established that the parameters in Equation 27 are sufficient statistics in that they are adequate to calculate $M_W(b)$ in a general class of dynamic models that nest existing structural models of insurance. He analyzed a dynamic model in which transitions from the good state to the bad state follow an arbitrary stochastic process. Agents make J choices and are subject to M constraints. The choices could include variables such as reservation wages, savings behavior, labor supply, or human capital investments. Subject to a regularity condition analogous to Assumption 1, Chetty showed that Equation 27 holds in this general model, with the difference in marginal utilities replaced by the difference between the average marginal utilities in the high and low states over the agent's life. This result distills the calculation of welfare gains in complex dynamic models to two parameters: the gap in average marginal utilities and the elasticity that enters the government's budget constraint $\varepsilon_{1-e,b}$. The identification of parameters such as asset limits or the degree of private insurance $[q(b_p)]$ is not required. This permits calculation of $\frac{dW}{db}$ without the assumptions made in the structural studies for tractability, such as no private insurance or no borrowing (Hansen & İmrohoroğlu 1992, Hopenhayn & Nicolini 1997).

Equation 27 cannot be directly implemented because the gap in marginal utilities remains to be recovered from choice data (step 4 of the rubric). The recent literature has proposed the use of three types of choice data to recover the marginal utility gap: consumption (Gruber 1997), liquidity and substitution effects in effort (Chetty 2008), and reservation wages (Shimer & Werning 2007).

5.2. Gruber (1997)

Taking a quadratic approximation to the utility function, Gruber observed that

$$\frac{u'(c_l) - u'(c_h)}{u'(c_l)} = \gamma \frac{\Delta c}{c_h}(b), \tag{28}$$

where $\gamma = -\frac{u''(c_h)}{u'(c_h)} c_h$ is the coefficient of relative risk aversion evaluated at c_h and $\Delta c = c_h - c_l$. He posited that the effect of UI benefits on consumption is linear (an assumption that should ideally be evaluated using a structural simulation):

$$\frac{\Delta c}{c_b}(b) = \alpha + \beta b.$$

In this specification, α measures the consumption drop that would occur in the absence of government intervention, and β measures the slope of the consumption function with respect to the benefit level. Combining this equation with Equations 27 and 28 yields the following expression for the marginal welfare gain from increasing the benefit level:

$$M_W(b) = (\alpha + \beta b)\gamma - \frac{\varepsilon_{1-e,b}}{e}. \qquad (29)$$

Building on work by Hamermesh (1982), Gruber estimated the consumption-smoothing effect of UI benefits by exploiting changes in UI benefit laws across states in the United States coupled with panel data on consumption. He estimated $\alpha = 0.24$ and $\beta = -0.28$, and then calibrated the welfare gain from raising UI benefits using estimates of $\varepsilon_{1-e,b}$ from Meyer (1990). He found that at conventional levels of risk aversion ($\gamma < 2$), increasing the UI benefit level above the levels observed in his data (roughly 50% of the wage) would lead to substantial welfare losses.

Gruber proceeded to solve for the b^* such that $\frac{dW}{db}(b^*) = 0$ in Equation 29 and found that b^* is close to zero. These calculations of the optimal benefit level assume that $\frac{\Delta c}{c}$ is linear in b, and $\gamma(b)$ and $\varepsilon_{1-e,b}(b)$ do not vary with b. This application of the sufficient-statistic formula, which is not guided by a structural model, could be inaccurate because it uses ad hoc assumptions to make predictions about counterfactuals that are well out of sample. Equation 29 should not be used to make statements about global optima unless one can estimate the sufficient statistics for a range of different benefit levels. Lacking such estimates, a more coherent method of inferring b^* would be to calibrate a structural model to match the sufficient statistics and simulate the optimal b^* in that model, as in Saez (2001).

A difficulty with Equation 29 is that risk aversion (γ) is known to vary substantially across contexts (Rabin 2000, Chetty & Szeidl 2007). Because Gruber's results are highly sensitive to the assumed value of γ, more recent studies have sought alternative techniques for recovering the gap in marginal utilities that do not require an estimate of γ.

5.3. Chetty (2008)

Chetty (2008) demonstrated that the gap in marginal utilities in Equation 27 can be backed out from the comparative statics of effort choice. To see this, we observe that the first-order condition for effort is

$$\psi'(e) = u(c_b) - u(c_l). \qquad (30)$$

Now let us consider the effect of an exogenous cash grant (such as a severance payment to job losers) on effort, holding fixed the private-insurance level b_p and the UI tax t:

$$\partial e/\partial A = \{u'(c_b) - u'(c_l)\}/\psi''(e) \leq 0. \qquad (31)$$

The effect of increasing the benefit level on effort (again holding b_p and t fixed) is

$$\partial e/\partial b = -u'(c_l)/\psi''(e). \qquad (32)$$

Combining Equations 31 and 32, we see that the ratio of the liquidity effect $(\partial e/\partial A)$ to the substitution effect $(\partial e/\partial w_b = \partial e/\partial A - \partial e/\partial b)$ recovers the gap in marginal utilities:

$$\frac{u'(c_l) - u'(c_b)}{u'(c_b)} = \frac{-\partial e/\partial A}{\partial e/\partial A - \partial e/\partial b}.$$

Plugging this into Equation 27 yields the following expression for the welfare gain from increasing the benefit level:

$$M_W(b) = \frac{-\partial e/\partial A}{\partial e/\partial A - \partial e/\partial b} - \frac{\varepsilon_{1-e,b}}{e}. \qquad (33)$$

The intuition for this formula is that the gap between marginal utilities in the good and bad states can be inferred from the extent to which effort is affected by liquidity versus moral hazard. In a model with perfect consumption smoothing $(c_b = c_l)$, the liquidity effect $\partial e/\partial A = 0$ because a cash grant raises $u(c_b)$ and $u(c_l)$ by the same amount. Chetty showed that similar to Baily's formula, Equation 33 holds in a general dynamic search model because of envelope conditions in the agent's value functions.

An issue that arises in the empirical implementation of Equation 33 is that $\frac{\partial e}{\partial b}$ must be measured holding the tax t fixed, whereas the elasticity $\varepsilon_{1-e,b}$ must be measured while permitting t to vary. Instead of attempting to estimate both parameters, Chetty used numerical simulations to show that the effect of a UI benefit increase on job-finding rates is virtually identical regardless of UI taxes being held fixed. This is because the fraction of unemployed individuals is quite small, making UI tax rates very low.

Chetty implemented Equation 33 by estimating the effects of unemployment benefits and severance payments on search intensity using hazard models for unemployment durations. He found that the welfare gains from raising the unemployment benefit level are small but positive, suggesting that the current benefit level is slightly below the optimum given the concavity of $W(b)$. He concluded that the optimal benefit level is close to the current wage-replacement rate of approximately 50%.

5.4. Shimer & Werning (2007)

Shimer & Werning (2007) inferred the gap in marginal utilities from the comparative statics of reservation wages instead of effort in a model of job search. They considered a model in which the probability of finding a job, e, is determined by the agent's decision to accept or reject a wage offer rather than by search intensity. Wage offers are drawn from a distribution $F(w)$. If the agent rejects the job offer, he receives income of $w_l + b$ as in the model above. For simplicity, assume that the agent has no private insurance $(q = \infty)$; allowing $q < \infty$ complicates the algebra but does not affect the final formula. The remainder of the model is specified as in Equation 26.

The agent rejects any net-of-tax wage offer $w - t$ below his outside option $w_l + b$. Therefore, $e = 1 - F(w_l + b + t)$, and the agent's expected value upon job loss is

$$W(b) = eE[u(w - t)|w - t > w_l + b] + (1 - e)u(w_l + b).$$

Even though the microeconomic choices of accepting or rejecting wage offers are discrete, the welfare function is smooth because of aggregation, as in Equation 10.

Shimer & Werning's insight is that $\frac{dW}{db}$ can be calculated using information on the agent's reservation wage. Suppose we ask the agent what wage he would be willing to accept with certainty prior to the start of job search.[18] Define the agent's reservation wage prior to job search as the wage $\bar{w}_0$ that would make the agent indifferent about accepting a job immediately to avoid having to take a random draw from the wage-offer distribution. The reservation wage $\bar{w}_0$ satisfies

$$u(\bar{w}_0 - t) = W(b).$$

The government's problem is to

$$\max W(b) = \max u(\bar{w}_0 - t)$$
$$\Rightarrow \max \bar{w}_0 - t. \tag{34}$$

Differentiating Equation 34 gives a sufficient-statistic formula:[19]

$$M_W(b) = \frac{d\bar{w}_0}{db} - \frac{dt}{db} = \frac{d\bar{w}_0}{db} - \frac{1-e}{e}\left(1 + \frac{1}{e}\varepsilon_{1-e,b}\right).$$

Intuitively, $\frac{d\bar{w}_0}{db}$ encodes the marginal value of insurance because the agent's reservation wage directly measures his expected value when unemployed. Shimer & Werning implemented Equation 34 using an estimate of $\frac{d\bar{w}_0}{db}$ from Feldstein & Poterba (1984) and found a large, positive value for $M_W(b)$ at current benefit levels. However, they cautioned that the credibility of existing reservation-wage elasticity estimates is questionable, particularly in view of evidence that UI benefit levels have little impact on subsequent wage rates (e.g., Card et al. 2007, van Ours & Vodopivec 2008).

The multiplicity of formulas for $M_W(b)$ illustrates a general feature of the sufficient-statistic approach: Because the model is not fully identified by the inputs to the formula, there are generally several representations of the formula for welfare gains.[20] This flexibility allows the researcher to apply the representation most suitable for his application given the available variation and data. For example, in analyzing disability insurance, it may be easiest to implement Chetty's (2008) formula as the available variation permits the identification of liquidity and moral hazard effects (Autor & Duggan 2007).

5.5. Inefficiencies in Private Insurance

An important assumption made in all three formulas above is that the choices within the private sector are constrained Pareto efficient; that is, total surplus is maximized in the private sector subject to the constraints. In practice, private-insurance contracts are likely

---

[18]Shimer & Werning studied a stationary dynamic model with constant absolute risk aversion utility where the reservation wage is fixed over time, in which case it does not matter at what point of the spell the reservation wage is elicited.

[19]This corresponds to equation 12 in Shimer & Werning (2007), where the unemployment rate is $u = 1 - e$. The slight difference between the formulas (the $\frac{1}{1-u}$ factor in the denominator) arises because Shimer & Werning write the formula in terms of a partial-derivative-based elasticity. Here, $\varepsilon_{1-e,b}$ is the elasticity including the UI tax response needed to balance the budget; in Shimer & Werning's notation, it is holding the tax fixed.

[20]All three formulas hold in models that allow both reservation wage and search intensity choices. Chetty's (2006a) generalization of Baily's formula nests the model with stochastic wages. If agents control the arrival rate of offers via search effort, the first-order condition for search effort remains the same as in Equation 30, with $Eu'(c_b)$ replacing $u'(c_b)$. It follows that Chetty's (2008) formula also holds with stochastic wages.

to be second-best inefficient as well because of adverse selection and moral hazard in private markets. In this case, the envelope condition invoked in deriving Equation 27 is violated because of externalities on the private insurer's budget constraint that are not taken into account by the individual.

Recent work by Einav et al. (2008b) and Chetty & Saez (2008) identified sufficient statistics for the welfare gains from social insurance in environments with adverse selection and moral hazard in private-insurance markets. Einav et al. developed a method of characterizing the welfare gain from government intervention that uses information about insurance purchase decisions. They showed that the demand curve for private insurance and the average cost of providing insurance as a function of the price are together sufficient statistics for welfare analysis. Einav et al. implemented their method using quasi-experimental price variation in health insurance policies and found that the welfare gains from government intervention in health insurance markets are small.

Chetty & Saez focused on ex-post behaviors, namely how marginal utilities vary across the high and low states, as in Baily's formula. They developed a simple extension to Gruber's (1997) implementation of the formula that includes two more parameters: the size of the private-insurance market and the crowdout of private insurance by public insurance. Intuitively, the government exacerbates the moral hazard distortion created by private insurance, and must therefore take into account the amount of private insurance and degree of crowdout to calculate the welfare gains from intervention. Chetty & Saez applied their formula to analyze health insurance, showing that naively applying Equation 27 dramatically overstates the welfare gains from government intervention in this case.

These examples illustrate that the sufficient-statistic approach can be extended to environments in which private-sector choices are not second-best efficient. One has to deviate slightly from the general rubric to handle such second-best inefficiencies because the incentive compatibility constraints for private insurance violate Assumption 1. These constraints lead to additional terms in the sufficient-statistic formula, increasing the number of moments that need to be estimated for welfare analysis.

6. APPLICATION 3: BEHAVIORAL MODELS

There is now considerable reduced-form evidence that individuals' behavior deviates systematically from the predictions of neoclassical perfect optimization models [see **Table 1** for a few examples and DellaVigna's (2009) review for many more]. In light of this evidence, an important new challenge is normative analysis in models where agents' choices deviate from perfect optimization. The budding literature on this topic has proposed some structural approaches to this issue, primarily in the context of time discounting. An early example is Feldstein's (1985) model of social security with myopic agents, in which individuals have a higher discount rate than the social planner. Feldstein numerically calculated the welfare gains from social security policies under various assumptions about the primitives. More recently, a series of papers have used calibrations of Laibson's (1997) β-δ model of hyperbolic discounting to make numerical predictions about optimal policy for agents who are impatient (see **Table 1**). Another set of studies has modeled the behavioral patterns identified in earlier work—such as ironing and spotlighting effects in response to nonlinear price schedules—and simulated optimal tax policy in such models (Liebman & Zeckhauser 2004, Feldman & Katuščák 2006).

The difficulty with the structural approach in behavioral applications is that there are often multiple positive models that can explain deviations from rationality, and each model can lead to different welfare predictions. The sufficient-statistic approach can be useful in such situations because welfare analysis does not require full specification of the positive model underlying observed choices (Bernheim & Rangel 2009). In applications in which agents optimize, the main benefit of the sufficient-statistic approach is that it simplifies identification. If one had unlimited power to identify primitives, there would be no advantage to using the sufficient-statistic approach in such models. In nonoptimizing models, however, the sufficient-statistic strategy has value even if identification is not a problem because there is no consensus alternative to the neoclassical model.

Given the infancy of this area, there is currently little work applying sufficient-statistic approaches to behavioral models. However, this is a fertile area for further research, as illustrated by Chetty et al.'s (2009) recent analysis of the welfare consequences of taxation when agents optimize imperfectly with respect to taxes. These authors presented evidence that the effect of commodity taxes on demand depends on whether the tax is included in posted prices. Taxes that are not included in posted prices (and are hence less salient to consumers) induce smaller demand reductions. There are various psychological and economic theories that could explain why salience affects behavioral responses to taxation, including bounded rationality, forgetfulness, and cue theories of attention. Chetty et al. therefore developed a sufficient-statistic approach to welfare analysis that is robust to specifications of the positive theory of tax salience.

Chetty et al. characterized the efficiency costs and incidence of taxation in a two-good model analogous to the Harberger model presented in Section 2. Let x denote the taxed good and y the numeraire. The demand for x as a function of its pretax price and tax rate is denoted by $x(p,t)$. I assume here that utility is quasilinear in y and production is constant returns to scale. These assumptions simplify the exposition by eliminating income effects and changes in producer prices; Chetty et al. showed that similar results are obtained when these assumptions are dropped.

The agent's true ranking of the consumption bundles (x,y) is described by a smooth, quasiconcave utility function

$$U(x,y) = u(x) + y$$
$$= u(x) + Z - (p+t)x,$$

where the second line imposes that the allocation (x,y) the agent chooses must satisfy the true budget constraint $(p+t)x + y = Z$. Chetty et al. departed from the traditional Harberger analysis by dropping the assumption that the consumption bundle (x,y) is chosen to maximize $U(x,y)$. Instead, they took the demand function $x(p,t)$ as an empirically estimated object generated by a model unknown to the policy maker, permitting $\frac{\partial x}{\partial p} \neq \frac{\partial x}{\partial t}$.

To calculate excess burden, assume that tax revenue is returned to the agent as a lump sum. Then, under the assumption that an individual's utility is a function purely of her ultimate consumption bundle, social welfare is given by

$$W(p,t) = \{u(x) + Z - (p+t)x\} + T(t),$$

where $T(t) = tx(t)$. In nonoptimizing models, one must deviate from step 2 of the rubric in Section 3 at this point because the envelope condition used to derive Equation 14 does not hold. Instead, totally differentiate the social-welfare function to obtain

$$\frac{dW}{dt} = [u'(x) - p]\frac{dx}{dt}. \tag{35}$$

This marginal-utility-based expression captures a simple intuition. An infinitesimal tax increase reduces consumption of x by $\frac{dx}{dt}$. The loss in surplus from this reduction in consumption of x is given by the difference between willingness to pay for x [$u'(x)$] and the cost of producing x, which equals the price p under the constant-returns-to-scale assumption.

Equation 35 applies in any model, irrespective of how $x(p,t)$ is chosen. Given that $\frac{dx}{dt}$ can be estimated empirically, the challenge in calculating $\frac{dW}{dt}$ (which reflects the main challenge in behavioral public economics more generally) is the recovery of the true preferences [$u'(x)$]. In neoclassical models, we use the optimality condition $u'(x) = p + t$ to recover marginal utility and immediately obtain the Harberger formula in Equation 4. As we do not know how x is chosen, we cannot use this condition here. Chetty et al. tackled this problem by making the following assumption.

> **Assumption 2:** When tax-inclusive prices are fully salient, the agent chooses the same allocation as a fully optimizing agent:
>
> $$x(p,0) = \arg\max_X u(x) + Z - px.$$

This assumption requires that the agent only makes mistakes with respect to taxes, and not fully salient prices. To see why this assumption suffices to calculate welfare, let $P(x) = x^{-1}(p,0)$ denote the agent's inverse-price-demand curve. Assumption 2 implies that $P(x) = u'(x)$ via the first-order condition for $x(p,0)$. Plugging this into Equation 35 yields

$$\frac{dW}{dt} = [P(x) - p]\frac{dx}{dt}.$$

This formula for $\frac{dW}{dt}$ can be implemented using an estimate of the inverse-price-demand curve $P(x)$. To simplify implementation, Chetty et al. made the approximation that demand $x(p,t)$ is linear in both arguments to obtain

$$\frac{dW}{dt} = \left[\frac{dp}{dx} \cdot (x(p,t) - x(p,0))\right]\frac{dx}{dt} = \left[\frac{dp}{dx} \cdot \frac{dx}{dt} t\right]\frac{dx}{dt} = t\theta\frac{dx}{dt}, \tag{36}$$

where $\theta = \frac{dx}{dt} / \frac{dx}{dp}$ measures the degree of underreaction to the tax relative to the price. This expression, which nests the Harberger formula as the case in which $\theta = 1$, shows that the price and tax elasticities of demand are together sufficient statistics to calculate excess burden in behavioral economics models. Intuitively, the tax-demand curve ($\frac{dx}{dt}$) is used to determine the actual effect of the tax on behavior. Then, the price-demand curve ($\frac{dx}{dp}$) is used to calculate the effect of that change in behavior on welfare. The price-demand curve can be used to recover the agent's preferences and calculate welfare changes because it is (by assumption) generated by optimizing behavior.

Because it does not rely on a specific structural model, Equation 36 accommodates all errors in optimization with respect to taxes and is hence easily adaptable to other applications. For example, confusion between average and marginal income-tax rates (Liebman & Zeckhauser 2004, Feldman & Katuščák 2006) or overestimation of estate tax rates (Slemrod 2006) can be handled using the same formula, without requiring knowledge of individuals' tax perceptions or rules of thumb. Any such behaviors are reflected in the empirically observed tax and price elasticities.

In sum, one can make progress in behavioral welfare economics by making assumptions that narrow the class of models under consideration without fully specifying one particular model. One is effectively forced to make stronger assumptions about the class of models in exchange for relaxing the full optimization assumption. These stronger assumptions make it especially important to implement step 6 of the rubric (model evaluation) in behavioral models. For instance, identifying the structural reasons for why tax salience matters would cast light on the plausibility of Assumption 2.

7. CONCLUSION: OTHER APPLICATIONS

The literature reviewed in this article has focused on identifying sufficient statistics for normative (welfare) analysis. Sufficient statistics can also be used to answer positive (descriptive) questions. A simple example is predicting the effect of a tax change on tax revenue. One only needs to estimate the elasticity of equilibrium quantity with respect to the tax rate to answer this question. Another example is the literature on capitalization effects (e.g., Summers 1981, Roback 1982, Greenstone & Gallagher 2008), which shows that changes in asset prices are sufficient statistics for distributional incidence in dynamic equilibrium models. The techniques relevant for positive analysis differ from those discussed in this article, as one cannot exploit envelope conditions in most positive applications. However, the general concept is still to formulate answers to questions in terms of a few elasticities instead of a full primitive structure.

Although the sufficient-statistic approach has been most widely applied in the public economics literature, it is natural to apply this approach in other areas. I conclude by briefly discussing applications in other fields.

7.1. Macroeconomics

A central debate in macroeconomics is whether households adhere to the permanent-income hypothesis. The structural approach to answering this question, taken, for example, by Scholz et al. (2006), is to specify a dynamic model of optimization and test whether observed consumption and savings patterns are consistent with those predicted by the model. The sufficient-statistic counterpart to this approach is to isolate one moment, such as the drop in consumption at retirement or the sensitivity of behavior to cash on hand, that is adequate to test between models (e.g., see Aguiar & Hurst 2005, Card et al. 2007). Similarly, models of business cycles and growth can be distinguished simply by identifying the labor wedge (Shimer 2008).

7.2. Labor

Labor economists have studied the effects of minimum wages on employment and wages extensively using reduced-form methods. Such evidence can be mapped into statements about optimal policy using a sufficient-statistic approach (Lee & Saez 2008). Another potential application is to the analysis of returns to schooling. Many studies have investigated the effects of schooling on behaviors such as job mobility and occupation choice. A sufficient-statistic approach would intuitively suggest that examining effects on total

earnings is adequate to measure the benefits of additional schooling in a model in which agents optimize.

7.3. Development

Starting with Townsend (1994), a large literature in development has studied risk-sharing arrangements. Although it is informative to understand the mechanisms through which shocks are smoothed, identifying consumption fluctuations and risk aversion is sufficient to make inferences about the welfare costs of shocks (Chetty & Looney 2006). Hence, quasi-experimental evidence such as that of Gertler & Gruber's (2002) study of health shocks in Indonesia sheds light on welfare and optimal policy even though it does not fully characterize the structure of insurance and risk-sharing networks. More generally, one may be able to give precise answers to policy questions using estimates from randomized experiments coupled with sufficient-statistic formulas derived from standard structural models.

7.4. Industrial Organization

Weyl & Fabinger (2009) showed that estimates of the pass through of cost shocks are sufficient statistics for questions ranging from the effects of mergers on markups to the effects of price caps on market structure. Researchers who seek to develop sufficient-statistic approaches to other problems in industrial organization (IO) must confront two challenges. First, many questions of interest concern discrete changes such as antitrust policy or the introduction of a new product. Second, much of the IO literature focuses on models of strategic interaction rather than price theory. In strategic games, small changes in exogenous parameters can lead to jumps in behavior. Because the techniques used in this article rely on smoothness, one may need to develop different techniques to apply the sufficient-statistic approach to IO problems. These are interesting topics for further research at the interface of structural and reduced-form methods.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

Aguiar M, Hurst E. 2005. Consumption vs. expenditure. *J. Polit. Econ.* 113:919–48
Amador M, Werning I, Angeletos G-M. 2006. Commitment vs. flexibility. *Econometrica* 74:365–96

Anderson PM, Meyer BD. 1997. Unemployment insurance takeup rates and the after-tax value of benefits. *Q. J. Econ.* 112:913–37

Angeletos G-M, Laibson D, Repetto A, Tobacman J, Weinberg S. 2001. The hyperbolic consumption model: calibration, simulation, and empirical evaluation. *J. Econ. Perspect.* 15:47–68

Angrist JD, Graddy K, Imbens G. 2000. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Rev. Econ. Stud.* 67:499–527

Angrist JD, Krueger AB. 1991. Does compulsory school attendance affect schooling and earnings? *Q.J. Econ.* 106:979–1014

Ashraf N, Karlan D, Yin W. 2006. Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines. *Q. J. Econ.* 121:635–72

Auerbach AJ. 1985. The theory of excess burden and optimal taxation. In *Handbook of Public Economics*, Vol. 1, ed. Alan J. Auerbach, MS Feldstein, pp. 67–127. Amsterdam: Elsevier Sci.

Autor DH, Duggan MG. 2003. The rise in the disability rolls and the decline in unemployment. *Q. J. Econ.* 118:157–205

Autor DH, Duggan MG. 2007. Distinguishing income from substitution effects in disability insurance. *Amer. Econ. Rev. Pages Proc.* 97:119–24

Baily MN. 1978. Some aspects of optimal unemployment insurance. *J. Public Econ.* 10:379–402

Ballard CL, Shoven JB, Whalley J. 1985. The total welfare cost of the United States tax system: a general equilibrium approach. *Natl. Tax J.* 38:125–40

Bernheim BD, Rangel A. 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *Q. J. Econ.* 124(1):51–104

Berry S. 1994. Estimating discrete-choice models of product differentiation. *RAND J. Econ.* 25:242–62

Berry S, Levinsohn J, Pakes A. 1995. Automobile prices in market equilibrium. *Econometrica* 63:841–90

Blau FD, Kahn LM. 2007. Changes in the labor supply behavior of married women: 1980–2000. *J. Labor Econ.* 25:393–438

Blundell R, Duncan A, McCrae J, Meghir C. 2000. The labour market impact of the working families' tax credit. *Fiscal Studies* 21:75–103

Blundell R, Duncan A, Meghir C. 1998. Estimating labor supply responses using tax reforms. *Econometrica* 66:827–62

Blundell R, Pistaferri L, Preston I. 2008. Consumption inequality and partial insurance. *Am. Econ. Rev.* 98(5):1887–921

Card D. 1990. The impact of the Mariel Boatlift on the Miami labor market. *Ind. Labor Relat. Rev.* 43:245–57

Card D, Chetty R, Weber A. 2007. Cash-on-hand and competing models of intertemporal behavior: new evidence from the labor market. *Q. J. Econ.* 122:1511–60

Chetty R. 2006a. A general formula for the optimal level of social insurance. *J. Public Econ.* 90:1879–901

Chetty R. 2006b. A new method of estimating risk aversion. *Am. Econ. Rev.* 96:1821–34

Chetty R. 2008. Moral hazard vs. liquidity and optimal unemployment insurance. *J. Polit. Econ.* 116:173–234

Chetty R. 2009. Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *Am. Econ. J.* In press

Chetty R, Looney A. 2006. Consumption smoothing and the welfare consequences of social insurance in developing economies. *J. Public Econ.* 90:2351–56

Chetty R, Looney A, Kroft K. 2009. Salience and taxation: theory and evidence. *Am. Econ. Rev.* In press

Chetty R, Saez E. 2008. *Optimal taxation and social insurance with endogenous private insurance*. NBER Work. Pap. 14403.

Chetty R, Saez E. 2009. *Teaching the tax code: earnings responses to an experiment with EITC recipients*. NBER Work. Pap. 14836

Chetty R, Szeidl A. 2007. Consumption commitments and risk preferences. *Q. J. Econ.* 122:831–77

Dawkins C, Srinivasan TN, Whalley J. 2001. Calibration. In *Handbook of Econometrics*, Vol. 5, ed. J Heckman, E Leamer, pp. 3653–703. Amsterdam: North Holland

Deaton A. 2009. *Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development*. NBER Work. Pap. 14690

Deaton A, Muellbauer J. 1980. An almost ideal demand system. *Am. Econ. Rev.* 70:312–26

DellaVigna S. 2009. Psychology and economics: evidence from the field. *J. Econ. Lit.* 47:315–72

DellaVigna S, Paserman MD. 2005. Job search and impatience. *J. Labor Econ.* 23:527–88

Diamond PA. 1998. Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *Am. Econ. Rev.* 88:83–95

Einav L, Finkelstein A, Cullen M. 2008a. *Estimating welfare in insurance markets using variation in prices*. NBER Work. Pap. 14414

Einav L, Finkelstein A, Schrimpf P. 2008b. *The welfare cost of asymmetric information: evidence from the U.K. annuity market*. NBER Work. Pap. W1322

Eissa N, Liebman JB. 1996. Labor supply response to the earned income tax credit. *Q. J. Econ.* 111:605–37

Feldman NE, Katuščák P. 2006. *Should the average tax rate be marginalized?* CERGE Work. Pap. 304

Feldstein MS. 1985. The optimal level of social security benefits. *Q. J. Econ.* 100:303–20

Feldstein MS. 1995. The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act. *J. Polit. Econ.* 103:551–72

Feldstein MS. 1999. Tax avoidance and the deadweight loss of the income tax. *Rev. Econ. Stat.* 81:674–80

Feldstein MS, Poterba J. 1984. Unemployment insurance and reservation wage. *J. Public Econ.* 23:141–67

Finkelstein A. 2007. The aggregate effects of health insurance: evidence from the introduction of Medicare. *Q. J. Econ.* 122:1–37

Genesove D, Mayer C. 2001. Loss aversion and seller behavior: evidence from the housing market. *Q. J. Econ.* 116:1233–60

Gertler PJ, Gruber J. 2002. Insuring consumption against illness. *Am. Econ. Rev.* 92:51–70

Golosov M, Tsyvinski A. 2006. Designing optimal disability insurance: a case for asset testing. *J. Polit. Econ.* 114:257–69

Golosov M, Tsyvinski A. 2007. Optimal taxation with endogenous insurance markets. *Q. J. Econ.* 122:487–534

Goolsbee A. 2000. What happens when you tax the rich? Evidence from executive compensation. *J. Polit. Econ.* 108:352–78

Gorodnichenko Y, Martinez-Vazquez J, Peter KS. 2009. Myth and reality of flat tax reform: micro estimates of tax evasion response and welfare effects in Russia. *J. Polit. Econ.* 117(3):504–54

Goulder LH, Williams RC. 2003. The substantial bias from ignoring general equilibrium effects in estimating excess burden, and a practical solution. *J. Polit. Econ.* 111:898–927

Greenstone M, Gallagher J. 2008. Does hazardous waste matter? Evidence from the housing market and the superfund program. *Q. J. Econ.* 123:951–1003

Gruber J. 1997. The consumption smoothing benefits of unemployment insurance. *Am. Econ. Rev.* 87:192–205

Gruber J, Saez E. 2002. The elasticity of taxable income: evidence and implications. *J. Public Econ.* 84:1–32

Gruber J, Wise D. 1999. *Social Security and Retirement Around the World*. Chicago: Univ. of Chicago Press

Hamermesh DS. 1982. Social insurance and consumption: an empirical inquiry. *Am. Econ. Rev.* 72:101–13

Hansen G, İmrohoroğlu A. 1992. The role of unemployment insurance in an economy with liquidity constraints and moral hazard. *J. Polit. Econ.* 100:118–42

Harberger AC. 1964. The measurement of waste. *Am. Econ. Rev.* 54:58–76

Hausman JA. 1981. Exact consumer's surplus and deadweight loss. *Am. Econ. Rev.* 71:662–76

Hausman JA, Newey WK. 1995. Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica* 63:1445–76

Heckman JJ, Vytlacil E. 2001. Policy relevant treatment effects. *Am. Econ. Rev.* 91:107–11

Heckman JJ, Vytlacil E. 2005. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73:669–738

Heckman JJ, Vytlacil E. 2007. Econometric evaluation of social programs. In *Handbook of Econometrics*, Vol. 6, ed. J Heckman, E Leamer, pp. 4779–874. Amsterdam: North Holland

Hines JR Jr. 1999. Three sides of Harberger triangles. *J. Econ. Perspect.* 13:167–88

Hopenhayn H, Nicolini JP. 1997. Optimal unemployment insurance. *J. Polit. Econ.* 105:412–38

Hoynes HW. 1996. Welfare transfers in two-parent families: labor supply and welfare participation under the AFDC-UP program. *Econometrica* 64:295–332

Imbens G. 2009. *Better LATE than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009)*. NBER Work. Pap. 14896

Imbens G, Wooldridge J. 2008. *Recent developments in the econometrics of program evaluation*. NBER Work. Pap. 14251

İmrohoroğlu A, İmrohoroğlu S, Joines D. 2003. Time inconsistent preferences and social security. *Q. J. Econ.* 118:745–84

Joint Econ. Comm. 2001. Economic benefits of personal income tax rate reductions. U.S. Congr., April

Keane M, Moffitt R. 1998. A structural model of multiple welfare program participation and labour supply. *Int. Econ. Rev.* 39:553–89

Keane M, Wolpin K. 1997. The career decisions of young men. *J. Polit. Econ.* 105:473–522

Koopmans TC. 1953. Identification problems in economic model construction. In *Studies of Econometric Method*, ed. WC Hood, TC Koopmans, pp. 27–48. Cowles Comm. Res. Econ. Monogr. No. 14. New York: Wiley

Laibson D. 1997. Golden eggs and hyperbolic discounting. *Q. J. Econ.* 62:443–77

Lalive R, van Ours J, Zweimuller J. 2006. How changes in financial incentives affect the duration of unemployment. *Rev. Econ. Stud.* 73:1009–38

Lalonde RJ. 1986. Evaluating the econometric evaluations of training programs with experimental data. *Am. Econ. Rev.* 76:604–20

Lee D, Saez E. 2008. *Optimal minimum wage in competitive labor markets*. NBER Work. Pap. 14320

Lentz R. 2009. Optimal unemployment insurance in an estimated job search model with savings. *Rev. Econ. Dynam.* 12:37–57

Levitt SD. 1997. Using electoral cycles in police hiring to estimate the effect of police on crime. *Am. Econ. Rev.* 87:270–90

Liebman JB, Zeckhauser RJ. 2004. *Schmeduling*. Work. Pap., Kennedy School Gov., Harvard Univ.

Lumsdaine R, Stock J, Wise D. 1992. Three models of retirement: computational complexity versus predictive ability. In *Topics in the Economics of Aging*, ed. D Wise, pp. 19–60. Chicago: Univ. of Chicago Press

Madrian BC, Shea DF. 2001. The power of suggestion: inertia in 401(k) participation and savings behavior. *Q. J. Econ.* 116:1149–87

Marschak J. 1953. Economic measurements for policy and prediction. In *Studies of Econometric Method*, ed. WC Hood, TC Koopmans, pp. 1–26. Cowles Comm. Res. Econ., Monogr. No. 14. New York: Wiley

Meyer BD. 1990. Unemployment insurance and unemployment spells. *Econometrica* 58:757–82

Meyer BD, Rosenbaum DT. 2001. Welfare, the earned income tax credit, and the labor supply of single mothers. *Q. J. Econ.* 116:1063–114

Mirrlees JA. 1971. An exploration in the theory of optimum income taxation. *Rev. Econ. Stud.* 38:175–208

Piketty T. 1997. La redistribution fiscale contre le chômage. *Rev. Fr. Econ.* 12:157–203

Rabin M. 2000. Risk aversion and expected-utility theory: a calibration theorem. *Econometrica* 68:1281–92

Roback J. 1982. Wages, rents, and the quality of life. *J. Polit. Econ.* 90:1257–78

Rosenzweig MR, Wolpin KI. 2000. Natural "natural experiments" in economics. *J. Econ. Lit.* 38:827–74

Rust J, Phelan C. 1997. How Social Security and Medicare affect retirement behavior in a world of incomplete markets. *Econometrica* 65:781–832

Saez E. 2001. Using elasticities to derive optimal income tax rates. *Rev. Econ. Stud.* 68:205–29

Scholz K, Seshadri A, Khitatrakun S. 2006. Are Americans saving 'optimally' for retirement? *J. Polit. Econ.* 114:607–43

Shapiro JM. 2005. Is there a daily discount rate? Evidence from the food stamp nutrition cycle. *J. Public Econ.* 89:303–25

Shimer R. 2008. Convergence in macroeconomics: the labor wedge. *Am. Econ. J. Macroeconomics* 1:280–97

Shimer R, Werning I. 2007. Reservation wages and unemployment insurance. *Q. J. Econ.* 122:1145–85

Shoven JB. 1976. The incidence and efficiency effects of taxes on income from capital. *J. Polit. Econ.* 84:1261–83

Slemrod JB. 1995. Income creation or income shifting? Behavioral responses to the Tax Reform Act of 1986. *Am. Econ. Rev.* 85:175–80

Slemrod JB. 2006. The role of misconceptions in support for regressive tax reform. *Natl. Tax J.* 59:57–75

Small KA, Rosen HS. 1981. Applied welfare economics with discrete choice models. *Econometrica* 49:105–30

Stone R. 1954. Linear expenditure systems and demand analysis: an application to the pattern of British demand. *Econ. J.* 64:511–27

Summers L. 1981. Taxation and corporate investment: a q-theory approach. *Brookings Pap. Econ. Act.* 1:67–127

Townsend R. 1994. Risk and insurance in village India. *Econometrica* 62:539–91

Train K. 2003. *Discrete Choice Methods with Simulation.* Cambridge, UK: Cambridge Univ. Press

Tuomala M. 1990. *Optimal Income Tax and Redistribution.* Oxford, UK: Clarendon

van Ours JC, Vodopivec M. 2008. Shortening the potential duration of unemployment benefits does not affect the quality of post-unemployment jobs: evidence from a natural experiment. *J. Public Econ.* 92:684–95

Weinzierl M. 2008. *The surprising power of age-dependent taxes.* Work. Pap., Harvard Univ.

Weyl GE, Fabinger M. 2009. *Pass-through as an economic tool.* Work. Pap., Harvard Univ.

Wolpin KI. 1987. Estimating a structural search model: the transition from school to work. *Econometrica* 55:801–17

Contents

Errata

An online log of corrections to *Annual Review of Economics* articles may be
found at http://econ.annualreviews.org