

MORAL HAZARD IN HEALTH INSURANCE: DO DYNAMIC INCENTIVES MATTER?

Aviva Aron-Dine, Liran Einav, Amy Finkelstein, and Mark Cullen*

Abstract—Using data from employer-provided health insurance and Medicare Part D, we investigate whether health care utilization responds to the dynamic incentives created by the nonlinear nature of health insurance contracts. We exploit the fact that because annual coverage usually resets every January, individuals who join a plan later in the year face the same initial (“spot”) price of health care but a higher expected end-of-year (“future”) price. We find a statistically significant response of initial utilization to the future price, rejecting the null that individuals respond only to the spot price. We discuss implications for analysis of moral hazard in health insurance.

I. Introduction

THE size and rapid growth of the health care sector, and the pressure this places on public sector budgets, has created great interest among both academics and policymakers in possible approaches to reducing health care spending. On the demand side, the standard, long-standing approach to constraining health care spending is through consumer cost sharing in health insurance, such as deductibles and coinsurance. Not surprisingly therefore, a substantial academic literature is devoted to trying to quantify how the design of health insurance contracts affects medical spending. These estimates have important implications for the costs of alternative health insurance contracts and, hence, for the optimal design of private insurance contracts or social insurance programs.

Received for publication January 25, 2013. Revision accepted for publication December 8, 2014. Editor: Amitabh Chandra.

* Aron-Dine: U.S. Office of Management and Budget; Einav: Stanford University and NBER; Finkelstein: MIT and NBER; Cullen: Stanford University and NBER.

We are grateful to David Molitor, Ray Kluender, and James Wang for outstanding research assistance, and to Amitabh Chandra (the editor), Ben Handel, Kate Ho, Jeff Liebman, Simone Schaner, numerous seminar participants, and two anonymous referees for helpful comments and suggestions.

Earlier versions of this paper were circulated under the title “Moral Hazard in Health Insurance: How Important Is Forward Looking Behavior?” (Aron-Dine et al., 2012). The material in section IV on Medicare Part D was previously circulated as a subsection of Einav, Finkelstein, and Schrimpf (2013).

We gratefully acknowledge support from the NIA R01 AG032449 (L.E. and A.F.), the National Science Foundation Grant SES-0643037 (L.E.), the John D. and Catherine T. MacArthur Foundation Network on Socio-economic Status and Health, and Alcoa, Inc. (M.C.), and the U.S. Social Security Administration through grant 5 RRC08098400-04-00 to the NBER as part of the SSA Retirement Research Consortium (L.E. and A.F.). The findings and conclusions expressed are solely our own and do not represent the views of SSA, any agency of the federal government, or the NBER.

The Alcoa portion of the data was provided as part of an ongoing service and research agreement between Alcoa, Inc. and Stanford, under which Stanford faculty, in collaboration with faculty and staff at Yale University, perform jointly agreed-on ongoing and ad hoc research projects on workers’ health, injury, disability, and health care, and Mark Cullen serves as senior medical advisor for Alcoa, Inc.

A.A.-D. contributed to this paper while she was a graduate student at MIT. She received her PhD in June 2012. She is currently employed as associate director for economic policy at the Office of Management and Budget. The views in this paper do not represent those of the Office of Management and Budget or the White House.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00518.

One aspect of this literature that we find remarkable is the near consensus on the nature of the endeavor: the attempt to quantify the response of medical spending with respect to its (out-of-pocket) price to the consumer. For example, in their chapter on health insurance in *Handbook of Health Economics*, Cutler and Zeckhauser (2000) summarize about thirty studies of the impact of insurance on health care utilization that all report an estimate of “the” price elasticity of demand for medical care (with respect to the out-of-pocket price). A particularly famous and widely used estimate of this elasticity is the RAND Health Insurance Experiment’s estimate of -0.2 (Manning et al., 1987; Keeler & Rolph, 1988).

Why do we find this remarkable? Health insurance contracts in the United States are highly nonlinear. Trying to estimate the behavioral response to a single out-of-pocket price is therefore usually not a well-posed exercise: It begs the question, “With respect to which price?” In the context of such nonlinear budget sets, trying to characterize an insurance policy by a single price could produce very misleading inferences. Consider, for example, an attempt to extrapolate from an estimate of the effect of coinsurance on health spending to the effect of introducing a high-deductible health insurance plan. An individual who responds only to the current “spot” price of medical care would respond to the introduction of a deductible as if the price has sharply increased to 100%. However, an individual who takes dynamic incentives into account and has annual health expenditures that are likely to exceed the new deductible would experience little change in the effective marginal price of care and therefore might not change his or her behavior much. Indeed, once one accounts for the nonlinear contract design, even characterizing which insurance contract would provide greater incentives to economize on medical spending becomes a complicated matter.¹

In this paper we test whether individuals in fact take dynamic incentives into account in their medical consumption decisions. A fully rational, forward-looking individual who is not liquidity constrained should take into account only the future price and recognize that (conditional on the future price) the spot price applied to a particular claim

¹ Consider, for example, two plans with a coinsurance arm that is followed by an out-of-pocket maximum of \$5,000. Imagine that plan A has a 10% coinsurance rate and plan B has a 50% coinsurance rate. Which plan would induce less spending? The naive answer would be that plan B is less generous and would therefore lead to lower medical utilization. Yet the answer depends on the distribution of medical spending without insurance, as well as the extent to which individuals respond to the dynamic incentives created by the nonlinear budget set. For example, an individual who suffers a compound fracture early in the coverage period and spends \$10,000 on a surgery would effectively obtain full insurance coverage for the rest of the year under plan B, but would face a 10% coinsurance rate (with a remaining \$4,000 stop loss) under plan A. We would therefore expect this individual to have greater medical utilization under plan B.

is not relevant and should not affect utilization decisions. Yet although it is natural to expect that individuals should respond to these dynamic incentives, there are at least three reasons that individuals might not optimize with respect to the entire nonlinear budget set. First, they may be affected by an extreme form of present bias and behave as if they are completely myopic, discounting the future entirely and responding only to the current spot price of care. Second, individuals may be affected by a host of potential behavioral biases and may not be completely aware of, attentive to, or understand the nonlinear budget set created by their health insurance plan; it is plausible that in most situations, the spot price that appears on the medical bill is much more salient. Finally, individuals may factor in the future prices they will face but be affected by the spot price due to liquidity constraints. Consider, for instance, a potential, nonemergency \$150 physician visit early in the coverage year. Even if the individual expects medical consumption during the year to eventually put him way above the deductible, he may find it difficult to obtain the \$150 that would be required to pay for a nonemergency physician visit early in the year, while he is still within the deductible range.

An ideal experiment for whether dynamic incentives matter in individuals' medical utilization response to health insurance would keep the spot price of care constant while generating variation in the future price of care. Finding such variation is empirically challenging, which may explain why there has been relatively little work on this topic. The key empirical difficulty arises because the spot price and the future price often vary jointly. A low-spending individual faces both a high spot price (because all her spending falls below the deductible) and a high expected end-of-year price (because she does not expect to hit the deductible), while the opposite is true for a high-spending individual. Similarly, the types of variation that have most often been used to estimate the impact of health insurance on medical spending, such as variation in deductibles or coinsurance rates, will change the spot price and the future price jointly.

We propose, and implement, an empirical strategy that approximates the ideal experiment, generating (arguably) identifying variation in the future price of care, while holding the spot price constant. Our strategy relies on the observation that typical health insurance contracts offer coverage for a fixed duration and reset on prespecified dates, but they generate identifying variation when they get applied to individuals whose initial coverage period is shorter. This practice produces a relatively common situation in which similar individuals face the same spot price for their consumption decision but have substantially different expected end-of-year prices. To see this, consider an annual health insurance contract that runs from January 1 to December 31 and has an annual deductible. It is common practice that when individuals join the plan in the middle of the year, the deductible remains at its annual level and applies only until the end of the calendar year. As a result, those who join a plan later in the year have fewer months to spend past the deductible

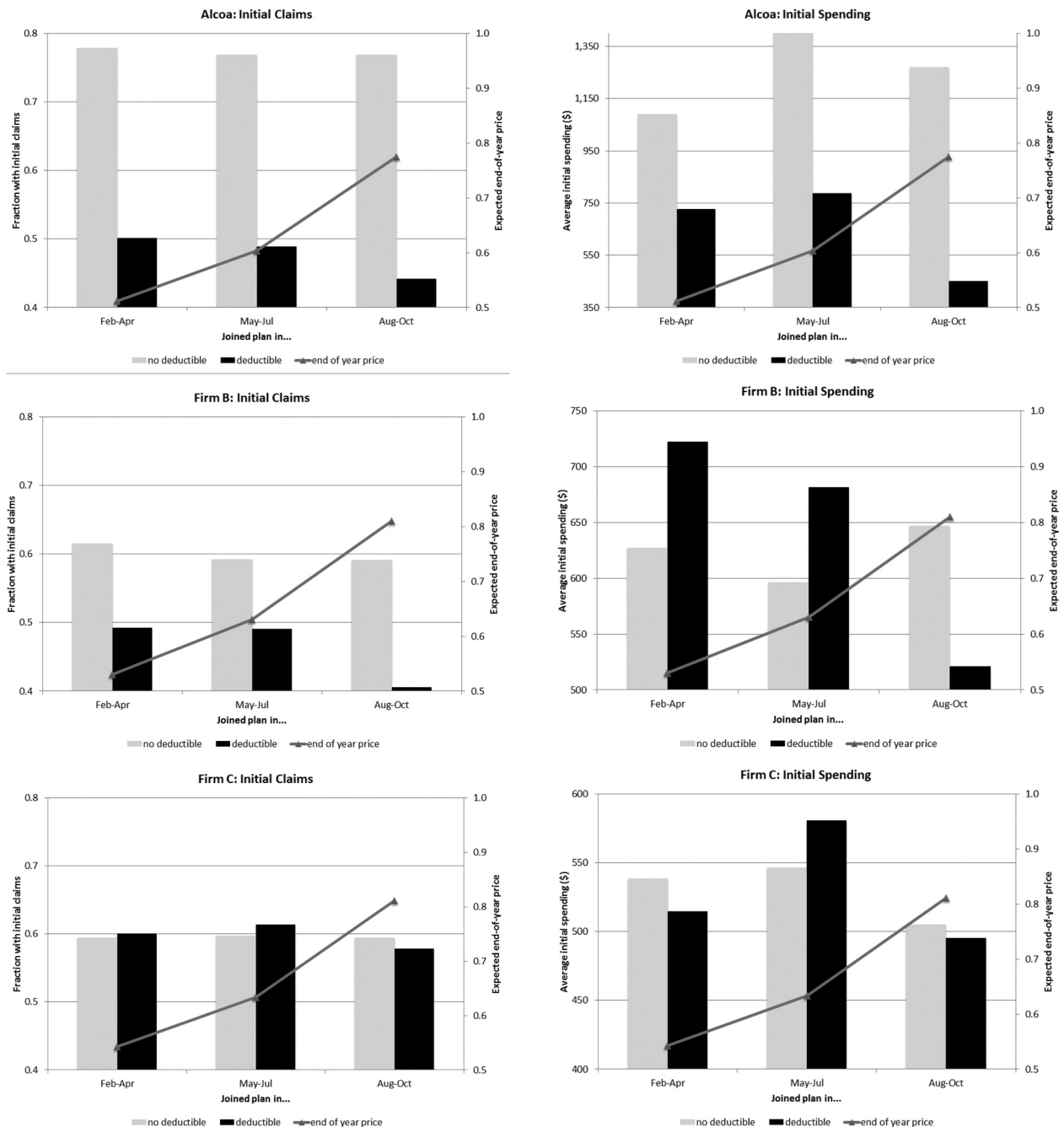
and thus face higher future prices. Initially, however, the spot price is the same regardless of when the individual joined the plan. Thus, all else equal, the expected end-of-year price is increasing with the calendar month in which individuals join a deductible plan, while the spot price is initially held fixed.

Our primary analysis applies this empirical strategy in the context of employer-provided health insurance in the United States, the source of over 85% of private health insurance coverage. We do so by comparing initial medical utilization across individuals who join the same deductible plan in the same firm but in different months of the year. To account for potential confounders such as seasonality in health care spending or seasonality in firm hiring, we use patterns of initial utilization by join month for individuals who join the same plan with no deductible, in which the future price hardly varies over the course of the year. To operationalize this strategy empirically, we draw on data from several large employers with information on their plan details as well as their employees' plan choices, demographics, and medical claims. Figure 1 (whose construction we describe in much more detail later in the paper) provides one way of summarizing this empirical exercise. It shows, separately for each firm, that as employees join a plan later in the year (and the expected end-of-year price rises for those in the deductible plan), initial medical utilization in the deductible plan tends to fall in both absolute terms and relative to the corresponding pattern in the no-deductible plan. We will use this basic design to present more formal analysis of a robust and statistically significant decline in initial medical utilization associated with a higher future price of medical care.

To further validate this finding and illustrate that it is not specific to the particular firms and their employees whose data we use, we also present an analogous finding in a different context and for a different population. Specifically, we take advantage of the fact that individuals newly eligible for Medicare, the public health insurance program for those aged 65 and over, can enroll in a Part D prescription drug plan in the month they turn 65, but the plan resets on January 1 regardless of when in the year they enrolled. Variation in birth month thus generates variation in contract duration and hence in expected end-of-year price among individuals in a given plan in their first year, and serves the same empirical purpose as variation in hire month did in our primary analysis. Using this design, we find once again that a higher future price is associated with a robust and statistically significant decline in initial prescription drug purchasing. This section of the paper also illustrates that the basic idea behind our empirical strategy, which we view as one of the contributions of the paper, is not unique to the specific data set we use but can be exploited in multiple settings.

Overall, the results consistently point to the fact that individuals respond to the future price of care, thus rejecting the null that individuals respond only to the spot price. That is, it appears that individuals understand something about the nature of their dynamic budget constraint and make their

FIGURE 1.—INITIAL MEDICAL UTILIZATION BY JOIN QUARTER



All utilization measures refer to utilization by the employee and any covered dependents. Initial claims (left panels) are defined as any claims within the first three months, and initial spending (right panels) is defined as the sum of all claim amounts of the claims that were made within the first three months. Expected end-of-year price is computed for the deductible plan only and corresponds to the end-of-year prices reported in table 2. Sample sizes by plan and join quarter are reported in table 2.

health care utilization decisions with at least some attention to dynamic considerations.

In the last section of the paper, we discuss some implications of these results for further work that aims to estimate and interpret moral hazard in health insurance. Our findings, while not necessarily surprising, should give pause to researchers pursuing the currently common practice in empirical work on moral hazard in health insurance of estimating an elasticity of demand for medical care with respect to a

single price. They suggest that taking into account dynamic incentives could be important for analyses of moral hazard and that summarizing highly nonlinear contracts with a single price could be prohibitively restrictive.

Yet an important limitation to our findings is that they fall short of pointing to a specific behavioral model or to a set of appropriate elasticities that would allow us to quantify the response of medical utilization to changes in the (nonlinear) contract design. In other words, while we have rejected

the null that individuals respond only to the spot price, our results do not necessarily imply that individuals respond only to the future price; we think that it is plausible that both prices could be important. With this in mind, in the last section, we describe what we see as potentially constructive uses of our findings for future work.

Our paper is related to several distinct literatures. Naturally our paper fits in the large empirical literature that tries to estimate moral hazard in health insurance or the price sensitivity of demand for medical care. For this literature, our findings highlight the importance of thinking about the entire budget set rather than about a single price. This point was emphasized in some of the early theoretical work on the impact of health insurance on health spending (Keeler, Newhouse, & Phelps, 1977; Ellis, 1986) but until recently has rarely been incorporated into empirical work. Several papers on the impact of health insurance on medical spending—Ellis (1986), Cardon and Hendel (2001), and more recently, Kowalski (2012), Dalton (2014), and our own work (Einav et al., 2013)—explicitly account for the nonlinear budget set, but do so under the (untested) assumption that individuals respond only to the future price of care.²

Outside the context of health insurance, a handful of papers address the question of whether individuals respond at all to the nonlinearities in their budget set and which single price may best approximate the nonlinear schedule to which individuals respond. This is the focus of Liebman and Zeckhauser (2004), Feldman and Katuscak (2006), and Saez (2010) in the context of the response of labor supply to the progressive income tax schedule, and of Borenstein (2009) and Ito (2014) in the context of residential electricity utilization. In most of these other contexts, as well as in our own previous work on moral hazard in health insurance (Einav et al., 2013), the analysis of demand in the presence of a nonlinear pricing schedule is static. This is partly because in most nonhealth contexts, information about intermediate utilization levels (within the billing or tax cycle) is not easy to obtain (for both consumers and researchers) and partly because dynamic modeling often introduces unnecessary complications in the analysis. In this sense, our current study, using the precise timing of medical utilization within the contract year, is virtually unique within this literature in its explicit focus on the dynamic aspect of medical utilization.³

² Nonlinear pricing schedules are not unique to health insurance. Indeed, a large literature, going back at least to Hausman (1985), develops methods that address the difficulties that arise in modeling selection and utilization under nonlinear budget sets, and applies these methods to other settings in which similar nonlinearities are common, such as labor supply (Burtless & Hausman, 1978; Blundell & MaCurdy, 1999; Chetty et al., 2011), electricity utilization (Reiss & White, 2005), and cellular phones (Grubb & Osborne, 2015; Yao et al., 2012).

³ One exception in this regard is Keeler and Rolph (1988), who, like us, test for dynamic incentives in health insurance contracts (but use a different empirical strategy and reach a different conclusion). More recent exceptions are Nevo, Turner, and Williams (2013), who analyze the effect of nonlinear pricing schedules in the context of residential broadband use, and our own work in the context of Medicare Part D (Einav, Finkelstein, & Schrimpf, 2015).

The focus on dynamic incentives relates more generally to empirical tests of forward-looking behavior, which plays a key role in many economic problems. From this perspective, a closely related work to ours is Chevalier and Goolsbee's (2009) investigation of whether durable goods consumers are forward looking in their demand for college textbooks (they find that they are). Despite the obvious difference in context, their empirical strategy is similar to ours. They use the fact that static, spot incentives remain roughly constant (as the pricing of textbook editions do not change much until the arrival of new editions), while dynamic incentives (the expected time until a new edition is released) change. A slightly cleaner aspect of our setting is that the constant spot prices and varying dynamic incentives are explicitly stipulated in the coverage contract rather than empirical facts that need to be estimated from data.

The rest of the paper proceeds as follows. Section II describes our research design and our data from the employer-provided health insurance context. Section III presents our main results. Section IV presents complementary analysis and evidence in a related context using data from Medicare Part D. The final section discusses some of the implications of our findings for empirical work on moral hazard in health insurance.

II. Approach and Data

A. Basic Approach

We test the null hypothesis that individuals' health care utilization decisions do not respond to dynamic incentives created by nonlinear health insurance contracts. In other words, we test whether their decisions can be approximated by a myopic assumption according to which they respond only to the spot price associated with a given health care purchase.

An ideal experiment would randomly assign individuals to settings in which the spot price is held fixed, while dynamic incentives vary, and examine initial health care decisions. If health care utilization decisions are well approximated by assuming that individuals respond only to the spot price, (initial) health care decisions would not change across settings. We focus throughout on initial health care decisions because variation in dynamic incentives will, by design, vary the out-of-pocket price associated with subsequent "noninitial" health care spending.

Our novel observation is that the institutional features of many health insurance contracts in the United States come close to approximating such an ideal experiment. Specifically, we use the fact that unlike other lines of insurance, such as auto insurance or home insurance, the annual coverage period of employer-provided health insurance is not customized to individual people and thus resets for all insureds on the same date (typically on January 1). This presumably reflects the need for market-wide synchronization to accommodate open enrollment periods, bidding by insurers,

and other related activities. (Presumably for similar reasons, the same institutional feature applies to Medicare Part D prescription drug plans. We discuss this in more detail below.)

As a result, employees' annual health insurance coverage begins (and ends, unless it is terminated due to job separation) on the same date for almost all individuals. Although all employees can choose to join a new plan for the subsequent year during the open enrollment period (typically in October or November), there are only two reasons employees can join a plan in the middle of the year: they are new hires or have a qualifying event that allows them to change plans in the middle of the year.⁴ In order to transition new employees (and occasionally existing employees who have a qualifying event) into the regular cycle, the common practice is to let employees choose from the regular menu of coverage options, to prorate linearly the annual premium associated with their choices, but to hold (at its annual level) the deductible amount constant. As a result, individuals who are hired at different points in the year, but are covered by the same (deductible) plan, face the same spot price (of one) but different dynamic incentives; all else equal, the probability that individuals would hit the (uniform) deductible is higher for longer coverage periods. Thus, as long as employees join the company at different times for reasons that are exogenous to their medical utilization behavior, variation in hire date (or in the timing of qualifying events) generates quasi-experimental variation in the dynamic incentives individuals face.

To illustrate, consider two identical employees who select a plan with a \$1,200 (annual) deductible and full coverage for every additional spending. The first individual is hired by the company in January and the second in October. The employee who is hired in January faces a regular, annual coverage, while the one who joins in October has only a three-month coverage horizon, after which coverage resets (on January 1 of the subsequent calendar year). The January joiner is therefore more likely to hit his deductible by the time the coverage resets. Therefore, in expectation, his end-of-year price is lower, although—crucially—just after they get hired, both January and October joiners have yet to hit their deductible, so their spot price is (initially) the same. If the employees respond to dynamic incentives, the January joiner has a greater incentive to utilize medical care on joining the plan; this would not be the case if the employees respond only to the spot price of care. Therefore, differences in (initial) spending cannot be attributed to differences in spot prices and must reflect dynamic considerations.

B. Data

With this strategy in mind, we obtained claim-level data on employer-provided health insurance in the United States. We limited our sample to firms that offered at least one PPO plan with a deductible (which would generate variation in the

dynamic incentives based on the employee's join month, as described in the previous section) and at least one PPO plan with no deductible. The relationship between initial utilization and join month in the no-deductible plan is used to try to control for other potential confounding patterns in initial medical utilization by join month (such as seasonal flu); in a typical no-deductible plan, the expected end-of-year price is roughly constant by join month, so, absent confounding effects that vary by join month, there are few dynamic incentives and the initial medical utilization of employees covered by a no-deductible plan should not systematically vary with join month.

The data come from two sources. The first is Alcoa, Inc., a large multinational producer of aluminum and related products. We have four years of data (2004–2007) on the health insurance options, choices, and medical insurance claims of its employees (and any insured dependents) in the United States. We study the two most common health insurance plans at Alcoa; one with a deductible for in-network expenditure of \$250 for single coverage (\$500 for family coverage) and one with no deductible associated with in-network spending. While Alcoa employed (and the data cover) about 45,000 U.S.-based individuals every year, the key variation we use in this paper is driven by midyear plan enrollment by individuals not previously insured by the firm, thus restricting our analysis to only about 7,000 unique employees (over the four years) who meet our sample criteria.⁵ Of the employees at Alcoa who join a plan midyear and did not previously have insurance at Alcoa that year, about 80% are new hires, and the other 20% are employees who were at Alcoa but uninsured at the firm, had a qualifying event that allowed them to change plans in the middle of the year, and chose to switch to Alcoa-provided insurance.

The Alcoa data are almost ideal for our purposes, with the important exception of sample size. To increase statistical power, we examined the set of firms (and plans) available through the NBER's files of Medstat's MarketScan database. The data on plan choices and medical spending are virtually identical in nature and structure to our Alcoa data (indeed, Alcoa administers its health insurance claims via Medstat); they include coverage and claim-level information from an employer-provided health insurance context, provided by a set of (anonymous) large employers.

We selected two firms that satisfied our basic criteria of being relatively large and offering both deductible and no-deductible options to their employees. Each firm has about 60,000 employees who join one of these plans in the middle of the year over the approximately six years of our data. This substantially larger sample size is a critical advantage over the Alcoa data. The main disadvantages of these data are that we cannot tell apart new hires from existing employees who are new to the firm's health coverage (presumably due to

⁴ Qualifying events include marriage, divorce, birth or adoption of a child, a spouse's loss of employment, or death of a dependent.

⁵ We restrict our analysis to employees who are not insured at the firm prior to joining a plan in the middle of the year because if individuals change plans within the firm (due to a qualifying event), the deductible would not reset.

TABLE 1.—SUMMARY STATISTICS

Employer	Plan	Years Offered	Midyear New Enrollees ^a	In-Network Deductible (\$)		Fraction Family	Fraction Female	Average Age	“Average” Enrollment Month ^b
				Single	Family				
Alcoa	A0	2004–2007	3,269	0	0	0.622	0.379	38.56	6.28
	A1	2004–2007	3,542	250	500	0.408	0.254	35.68	6.42
Firm B	B0	2001–2005	37,759	0	0	0.530	0.362	36.77	6.35
	B1	2001–2005	9,553	150	300	0.382	0.341	36.87	6.29
Firm C ^c	C0	1999–2002, 2004–2005	27,968	0	0	0.348	0.623	36.40	7.35
	C1	1999–2000	6,243	200	500	0.348	0.622	37.53	7.50
	C2	2001–2002	8,055	250	625	0.323	0.606	38.66	7.56
	C3	2004–2005	5,633	300	750	0.299	0.660	38.51	7.67

^aThe sample includes employees who enroll in February through October.

^bIn computing the average enrollee month, we number the join months from 2 (February) through 10 (October).

^cWe omit 2003 from the analysis since the plan documentation regarding the deductible plan was incomplete in that year.

qualifying events that allow them to join a health insurance plan in the middle of the year), and there is less demographic information on the employees.

Because employers in MarketScan are anonymous (and we essentially know nothing about them), we will refer to these two additional employers as firm B and firm C. We focus on two plans offered by firm B. We have five years of data (2001–2005) for these plans, during which firm B offered one plan with no in-network deductible and one plan that had a \$150 (\$300) in-network single (family) deductible. The data for firm C are similar, except that the features of the deductible plan have changed slightly over time. We have seven years of data for firm C (1999–2005), during which the firm continuously offered a no-deductible plan (in-network) alongside a plan with a deductible. The deductible amount increased over time, with a single (family) in-network deductible of \$200 (\$500) during 1999 and 2000, \$250 (\$625) during 2001 and 2002, and \$300 (\$750) during 2004 and 2005.

Table 1 summarizes the key features of the plans (and their enrollment) that are covered by our final data set. In all three firms, we limit our sample to employees who join a plan between February and October and did not have insurance at the firm immediately prior to this join date. We omit employees who join in January for reasons related to the way the data are organized that make it difficult to tell apart new hires who join the firm in January from existing employees. We omit employees who join in November or December because, as we discuss in more detail below, we use data from the first three months after enrollment to construct our measures of “initial” medical utilization.⁶ Table 1 also summarizes, by plan, the limited demographic information we observe on each covered employee: the type of coverage they chose (family or single) and the employee’s gender, age, and enrollment month.⁷

⁶In practice we observe the join month rather than the join date. Thus, throughout the paper, when we speak of the “first three months” after enrollment, more precisely we are using the first two to three months after enrollment. As long as the join day within the month is similar across months, the average time horizon should also be similar by join month.

⁷In each firm we lose roughly 15% to 30% of new plan joiners because of some combination of missing information about the employee’s plan, missing plan details, or missing claims data (because the plan is an HMO or a partially or fully capitated POS plan).

C. Measuring Dynamic Incentives

Table 2 describes the key variation we use in our empirical analysis. We use the expected end-of-year price to summarize the dynamic incentive and report it as a function of the time within the year an employee joined the plan.⁸ Specifically, we define the expected end-of-year price, or future price, p^f , as

$$p_{jm}^f = 1 - \Pr(\text{hit}_{jm}), \quad (1)$$

where $\Pr(\text{hit}_{jm})$ is the probability an employee who joins plan j in month m will hit (i.e., spend more than) the in-network deductible by the end of the year. We calculate $\Pr(\text{hit})$ as the fraction of employees in a given plan and join month who have spent more than the in-network deductible by the end of the year.⁹ For example, consider a plan with a \$500 deductible and full coverage for any medical expenditures beyond the deductible. If 80% of the employees who joined the plan in February have hit the deductible by the end of the year, the expected end-of-year price would be $0.8 \times 0 + 0.2 \times 1 = 0.2$. If only 40% of the employees who joined the plan in August have hit the deductible by the end of the year, their expected end-of-year price would be $0.4 \times 0 + 0.6 \times 1 = 0.6$. Thus, the future price is the average (out-of-pocket) end-of-year price of an extra dollar of in-network spending. It is a function of one’s plan j , join month m , and the annual spending of all the employees in one’s plan and join month.¹⁰

Table 2 summarizes the average future price for each plan based on the quarter of the year in which one joins the plan. For plans with no deductible (A0, B0, and C0), the future price is mechanically 0 (since everyone hits the zero deductible), regardless of the join month. For deductible plans, however,

⁸In this and all subsequent analyses, we pool the three different deductible plans in firm C that were offered at different times over our sample period.

⁹We calculate $\Pr(\text{hit})$ separately for employees with individual and family coverage (since both the deductible amount and spending patterns vary with the coverage tier), and therefore in all of our analyses, p^f varies with coverage tier. However, for conciseness, in the tables we pool coverage tiers and report the (weighted) average across coverage tiers within each plan.

¹⁰To the extent that individuals have private information about their future health spending, and thus about their expected end-of-year price, our “average” measure will introduce measurement error and thus bias us against finding a response to dynamic incentives.

TABLE 2.—VARIATION IN EXPECTED END-OF-YEAR PRICE

Employer	Plan	Deductible (Single/Family) (<i>N</i> = enrollees)	Expected End-of-Year Price ^a		
			February–April ^b	May–July ^b	August–October ^b
Alcoa	A0	0 (<i>N</i> = 3,269)	0.000 (<i>N</i> = 1,007)	0.000 (<i>N</i> = 981)	0.000 (<i>N</i> = 1,281)
	A1	250/500 (<i>N</i> = 3,542)	0.512 (<i>N</i> = 975)	0.603 (<i>N</i> = 1,114)	0.775 (<i>N</i> = 1,453)
Firm B	B0	0 (<i>N</i> = 37,759)	0.000 (<i>N</i> = 8,863)	0.000 (<i>N</i> = 15,102)	0.000 (<i>N</i> = 13,794)
	B1	150/300 (<i>N</i> = 9,553)	0.529 (<i>N</i> = 2,165)	0.630 (<i>N</i> = 4,175)	0.806 (<i>N</i> = 3,213)
Firm C	C0	0 (<i>N</i> = 27,968)	0.000 (<i>N</i> = 6,504)	0.000 (<i>N</i> = 6,158)	0.000 (<i>N</i> = 15,306)
	C1–C3 ^c	200–300/500–750 (<i>N</i> = 19,931)	0.543 (<i>N</i> = 4,001)	0.633 (<i>N</i> = 4,143)	0.811 (<i>N</i> = 11,787)

^aExpected end-of-year price is equal to the fraction of individuals who do not hit the deductible by the end of the calendar year (and therefore face a marginal price of 1). It is computed based on the plan's deductible level(s), the join month, and the annual spending of all the employees who joined that plan in that month. We compute it separately for family and single coverage within a plan and report the enrollment-weighted average.

^bMonth individuals joined plan.

^cIn firm C, we pool the three deductible plans (C1, C2, and C3) that are offered in different years.

the future price varies with the join month. Only a small fraction of the individuals who join plans late in the year (August through October) hit their deductible, so their future price is around 0.8 in all firms. In contrast, many more employees who join a deductible plan early in the year (February to April) hit their deductible, so for such employees, the future price is just over 0.5. Thus, early joiners who select plans with a deductible face an average end-of-year price that is about 30 percentage points lower than the end-of-year price faced by late joiners. Yet initially (just after they join), both types of employees have yet to hit their deductible, so they all face a spot price of one. Differences in initial spending between the groups therefore plausibly reflect their dynamic response to the future price. This baseline definition of the future price—the fraction of employees who join a given plan in a given month whose spending does not exceed the in-network deductible by the end of the calendar year—will be used as the key right-hand variable in much of our subsequent empirical work.

Our baseline measure of the future price abstracts from several additional characteristics of the plans, which are summarized in appendix table A1. First, it ignores any coinsurance features of the plans. Plans A0, A1, and C1 to C3 all have a 10% coinsurance rate, while plans B0 and C0 have a 0 coinsurance rate. The coinsurance rate for plan B1 is unknown (to us). Second, we use only the in-network plan features and assume that all spending occurs in network. In practice, each plan (including the no-deductible plan) has deductibles and higher consumer coinsurance rates for medical spending that occurs out of network.

There are two consequences of these abstractions, both of which bias any estimated impact of the future price on behavior toward 0. First, abstracting from these features introduces measurement error into the future price. Second, our analysis assumes that for the no-deductible plans, there is no variation in the future price for employees who join in different months (i.e., the spot price and the future price are always the same). In practice, both a positive in-network

coinsurance rate (prior to the stop-loss) and the existence of out-of-network deductibles in all of the no-deductible (in-network) plans mean that the future price also increases with the join month for employees in the no-deductible plans.

III. Main Results

A. Patterns of Initial Utilization by Join Month

We proxy for initial medical utilization with two measures. In both cases, the measures of utilization encompass the utilization of the employee and any covered dependents. Our main measure of initial utilization is an indicator for whether the individual had any claim over some initial duration (“any initial claim”). We use three months as our initial duration measure; results are qualitatively similar for shorter durations. Overall, about 58% of the sample has a claim within the first three months. As an additional measure, we also look at a measure of total spending (in dollars) over the initial three months. Average three-month spending in our sample is about \$600.

Figure 1 reports, for each firm separately, the pattern of initial medical utilization by join month for the deductible plan and the no-deductible plan. The panels on the left report the results for whether there is a claim in the first three months; the right panels report results for initial three-month spending. These statistics already indicate what appears to be a response to dynamic incentives. For the deductible plan, the probability of an initial claim and initial medical spending both generally tend to fall with join month (and expected end-of-year price), while there is generally no systematic relationship between join month and initial medical utilization in the corresponding no-deductible plan. This is exactly the qualitative pattern one would expect if individuals respond to dynamic incentives.

We operationalize this analysis a little more formally by regressing measures of initial utilization on join month. A unit of observation is an employee *e* who joins health insurance

TABLE 3.—RELATIONSHIP BETWEEN INITIAL MEDICAL UTILIZATION AND JOIN MONTH

Employer	Plan	Deductible (Single/Family) (<i>N</i> = enrollees)	Any Initial Claim ^a		Log Initial Spending ^b	
			Difference (1)	DD (2)	Difference (3)	DD (4)
Alcoa	A0	0 (<i>N</i> = 3,269)	−0.001 (0.002)		−0.003 (0.023)	
	A1	250/500 (<i>N</i> = 3,542)	−0.002 (0.004)	−0.001 (0.004)	−0.015 (0.021)	−0.012 (0.027)
Firm B	B0	0 (<i>N</i> = 37,759)	−0.006 (0.001)		−0.015 (0.007)	
	B1	150/300 (<i>N</i> = 9,553)	−0.016 (0.004)	−0.010 (0.003)	−0.091 (0.026)	−0.075 (0.025)
Firm C	C0	0 (<i>N</i> = 27,968)	−0.0004 (0.002)		−0.004 (0.013)	
	C1–C3	200–300/500–750 (<i>N</i> = 19,931)	−0.0043 (0.0021)	−0.0039 (0.0018)	−0.027 (0.012)	−0.022 (0.010)

The table reports coefficients (and standard errors in parentheses) from regressing a measure of initial medical care utilization (defined in the top row) on join month (which ranges from 2, for February, to 10, for October). Columns 1 and 3 report the coefficient on join month separately for each plan, based on estimating equation (2). The regressions also include an indicator variable for coverage tier (single versus family). Columns 2 and 4 report the difference-in-differences (DD) coefficient on the interaction of join month and having a deductible plan, separately for each firm, based on estimating equation (3). The regressions also include plan by coverage tier fixed effects and join month fixed effects. Standard errors are clustered on join month by coverage tier.

^aThe dependent variable is an indicator for at least one claim made by the employee or any covered family members within the first three months in the plan.

^bThe dependent variable is $\log(s + 1)$, where s is the total medical spending of the employee and any covered family members in their first three months in the plan.

plan j during calendar month m . As mentioned, we limit attention to employees who join new plans between February and October, so $m \in \{2, \dots, 10\}$. As a result, all of our comparisons of patterns of initial utilization by join month, both within and across plans, are among a set of employees who are new to their plan.

The simplest way by which we can implement our strategy is to look within a given health plan that has a positive deductible and regress a measure of initial medical utilization y_e on the join month m_e and possibly a set of controls x_e , so that

$$y_e = \beta_j m_e + x_e' \gamma + u_e. \quad (2)$$

Absent any confounding influences of join month on y_e , we would expect an estimate of $\beta_j = 0$ for deductible plans if individuals respond only to the spot price of care and $\beta_j < 0$ if individuals do not. We include an additional covariate for whether the employee has family (as opposed to single) coverage to account for the fact that the deductible varies within a plan by coverage tier (see table 1) and that there naturally exist large differences in average medical utilization in family versus single coverage plans.

We analyze two different dependent variables. “Any initial claim” is a binary variable for whether the insured had any claim in the first three months. “Log initial spending” is defined as $\log(s + 1)$, where s is total medical spending (in dollars) by the insured during his first three months in the plan. Given that medical utilization is highly skewed, the log transformation helps in improving precision and reducing the effect of outliers.¹¹ Columns 1 and 3 of table 3 report results from estimating equation (2) on these two dependent

variables, separately for each plan. The key right-hand-side variable is the join month, enumerated from 2 (February) to 10 (October). In plans that have a deductible (A1, B1, and C1–C3), dynamic considerations would imply a negative relationship between join month and initial medical utilization. The results show exactly this qualitative pattern.

B. Patterns of Initial Utilization by Join Month for Deductible versus No-Deductible Plans

If seasonality in medical utilization is an important factor, it could confound the interpretation of the estimated relationship we have just discussed. For example, if claim propensity in the spring is greater than claim propensity in the summer due to, say, seasonal flu, then we may incorrectly attribute the decline in spot utilization for late joiners as a response to dynamic incentives. To address such concerns about possible confounding differences across employees who join plans at different months of the year, we use as a control group employees within the same firm who join the health insurance plan with no deductible in different months. As discussed earlier, such employees are in a plan in which the spot price and future price are roughly the same, so that changes in their initial utilization over the year (or lack thereof) provide a way to measure and control for factors that influence initial utilization by join month that are unrelated to dynamic incentives.

Columns 1 and 3 of table 3, discussed earlier, also show the plan-level analysis of the relationship between initial medical utilization and join month for the no-deductible plans (A0, B0, and C0).¹² The coefficient on join month for the no-deductible plans tends to be much smaller than the

¹¹ While conceptually a concave transformation is therefore useful, we have no theoretical guidance as to the “right” functional form. Any transformation therefore (including the one we choose) is ad hoc, and we simply choose one that is convenient and easy to implement. We note, however, that Box-Cox analysis of the $s + 1$ variable suggests that a log transformation is appropriate.

¹² The table reports average effects across the year. Figure 1 provided the graphical analog. Table 2 shows that the number of observations (new employees) is not uniform over the year; thus the regression results, especially in the context of Firm C, are affected more strongly by the late-in-the-year utilization patterns.

below, when we discuss possible threats to this identifying assumption and attempt to examine its validity.

The above analyses suggest that individuals' initial medical utilization responds to their contract duration, and does so differentially by the nature of the contract. To more directly investigate this relationship, as well as to enhance comparability across firms, we map the contract duration of a given contract into a measure of the expected end-of-year price p^f defined earlier (recall equation [1] for a definition, and table 2 for summary statistics).¹³ We then analyze a variant of the difference-in-difference analysis (equation [3]) in which we replace the deductible interacted with the join month variable ($m_e D_j$) with the expected end-of-year price variable. The estimating equation is thus modified to

$$y_e = \tilde{\beta}' p_{im}^f + \tilde{\mu}_i + \tilde{\tau}_m + x_e' \tilde{\gamma}' + \tilde{v}_e, \quad (4)$$

where (as before) $\tilde{\mu}_j$ are plan (by coverage tier) fixed effects and $\tilde{\tau}_m$ are join-month fixed effects. This transformation also aids in addressing the likely nonlinear effect of join month on both expected end-of-year price and expected utilization. Table 2 indicates that, indeed, our measure of the end-of-year price varies nonlinearly over time.

The future price variable is constructed based on the observed spending patterns of people who join a specific plan (and coverage tier) in a specific month, while our dependent variable is a function of that spending for an individual in that plan and enrollment month. This raises three related issues for interpreting the coefficient on the future price in equation (4) as the causal effect of the future price on initial medical utilization. The first issue is the mechanical relationship between the future price (our key right-hand-side variable) and initial health care utilization (the dependent variable): higher initial spending of individuals who enroll in a given plan in a given month would make them more likely to hit the deductible by the end of the year, and thus mechanically make our measure of future price lower. This will bias our estimate of the impact of the future price on utilization away from 0. A second issue is a standard reflection bias concern. The future price is calculated based on the total spending of the set of people who enroll in a given plan in a given month, and the dependent variable is the initial spending of a given person who enrolled in the plan in that month. This problem is more acute the smaller is the sample size of people enrolling in a given plan in a given month (and thus the larger the contribution of the individual to the plan-month mean total spending). A final issue is the potential for common shocks. If there is

¹³ If individuals are risk neutral, this is the only moment of the dynamic incentives that should matter for their utilization decisions. In practice, individuals may not be risk neutral and other moments of the end-of-year price may affect initial utilization. Limiting our analysis to the impact of the expected end-of-year price can therefore bias us against rejecting the null of no response to dynamic incentives.

a shock to health or spending that is specific to individuals enrolling in a specific plan in a specific month (e.g., the flu hits particularly virulently those who enroll in a particular plan in a different month), this introduces an omitted variable that is driving both the calculated future price and initial drug use.

To address all three of these concerns, we estimate an instrumental variable version of equation (4) in which we instrument for the future price with a simulated future price. In principle, the join month interacted with the deductible would be a valid instrument (and in this sense table 3 can be seen as presenting a set of “reduced form” results). Our simulated future price is simply a nonlinear function of join month and deductible. Like the future price, the simulated future price is computed based on the characteristics of the plan (by coverage tier) chosen and the month joined. However, unlike the future price, which is calculated based on the spending of people who joined that plan (by coverage tier) that month, the simulated future price is calculated based on the spending of all employees in that firm and coverage tier in our sample who joined either the deductible or no-deductible plan, regardless of join month. Specifically, for every employee in our sample in a given firm and coverage tier (regardless of plan and join month), we compute their monthly spending for all months that we observe them during the year that they join the plan, creating a common monthly spending pool. We then simulate the future price faced by an employee in a particular plan and join month by drawing (with replacement) 110,000 draws of monthly spending from this common pool, for every month for which we need a monthly spending measure.¹⁴ For each simulation we then compute the expected end-of-year price based on the draws. By using a common sample of employee spending that does not vary with plan or join month, the instrument is “purged” of any potential variation in initial medical utilization that is correlated with plan and join month, in very much the same spirit as Currie and Gruber’s (1996) simulated Medicaid eligibility instrument. An additional attraction of this IV strategy is that it helps correct for any measurement error in our calculated future price (which would bias the coefficient toward 0). On net, therefore, the OLS estimate may be biased upward or downward relative to the IV estimate.

The first three rows of table 4 report, separately for each firm, the results of estimating equation (4) by OLS and IV. As would be expected, the first stage is very strong. The results consistently show a negative relationship between the future price and our measures of initial medical use. The results are statistically insignificant for Alcoa, but almost always statistically significant for firms B and C (where the sample sizes are much bigger). The IV estimates tend to be smaller than the OLS estimates, suggesting that on net, the OLS estimates

¹⁴ For the first month, we draw from the pool of first-month spending (since people may join the plan in the middle of the month, the first month’s spending has a different distribution from other months), whereas for all other months in the plan that year, we draw from the pool (across families and months) of non-first-month spending.

TABLE 4.—RELATIONSHIP BETWEEN INITIAL MEDICAL UTILIZATION AND EXPECTED END-OF-YEAR PRICE

Sample	N	Any Initial Claim ^a		Log Initial Spending ^b	
		OLS (1)	IV (2)	OLS (3)	IV (4)
Alcoa	6,811	−0.09 (0.07)	−0.03 (0.09)	−0.76 (0.51)	−0.33 (0.63)
Firm B	47,312	−0.22 (0.07)	−0.22 (0.06)	−1.73 (0.54)	−1.70 (0.51)
Firm C	47,899	−0.10 (0.04)	−0.09 (0.05)	−0.81 (0.37)	−0.37 (0.24)
Pooled	102,022	−0.14 (0.04)	−0.13 (0.04)	−1.08 (0.29)	−0.78 (0.27)

The table reports coefficients (and standard errors in parentheses) from regressing a measure of initial medical care utilization (defined in the top row) on the expected end-of-year price, computed for each plan (by coverage tier) and join month. Columns 1 and 3 report the OLS coefficient on expected end-of-year price (p^e) from estimating equation (4); these regressions include plan by coverage tier fixed effects and join month fixed effects. In the bottom row, we estimate this by pooling the data from all firms and plans, where, in addition to the plan-by-coverage tier and join month fixed effects, these regressions now also include firm-by-join-month fixed effects. Standard errors are clustered on join month by coverage tier by firm. Columns 2 and 4 report IV estimates of the same regressions, where we use a simulated end-of-year price as an instrument for the expected end-of-year price (see the text for details). For the pooled regression (bottom row), the coefficient on the instrument in the first stage is 0.56 (standard error 0.024); the F -statistic on the instrument is 524. For the individual firm regressions, the coefficient on the instrument in the first stage ranges between 0.49 and 0.71, is always statistically significant, and has an F -statistic above 400.

^aThe dependent variable is an indicator for at least one claim made by the employee or any covered family members within the first three months in the plan.

^bThe dependent variable is $\log(s + 1)$ where s is the total medical spending of the employee and any covered family members in their first three months in the plan.

are upward biased due to common shocks or endogenous spending discussed above.

Thus far, all of the analysis has been of single plans or pairs of plans within a firm. The use of future price (rather than join month) also allows us to more sensibly pool results across firms and summarize them with a single number, since the relationship between join month and future price will vary with both the level of the deductible and the employee population. In pooling the data, however, we continue to rely on only within-firm variation. That is, we estimate

$$y_e = \beta p_{jm}^f + \mu_j + \tilde{\tau}_{mf} + x_e' \tilde{\gamma} + \tilde{v}_e, \quad (5)$$

where $\tilde{\tau}_{mf}$ denotes a full set of join month by firm fixed effects. The bottom row of table 4 reports the results from this regression. Once again we report both OLS and IV results.

The effect of future price in this pooled regression is statistically significant for both dependent variables. The IV estimates indicate that a 10 cent increase in the future out-of-pocket price (for every dollar of total health care spending) is associated with a 1.3 percentage point (about 2.2%) decline in the probability of an initial medical claim and a 7.8% decline in initial medical spending. Given an average expected end-of-year price of about 70 cents (for every dollar of medical spending) for people in our sample who choose the deductible plan, the 2.2% decline in the probability of an initial claim suggests an elasticity of initial claiming with respect to the future price of about -0.16 . The 7.8% decline in initial medical spending suggests an elasticity of initial medical spending with respect to the future price of -0.56 .¹⁵

¹⁵ We note that the IV estimate of the impact of the future price on the first three months’ spending could be biased upward since the spot price may vary across individuals over the first three months; 17% of individuals in

We investigated the margin on which the individual's response to the future price occurs. About three-quarters of medical expenditures in our data represent outpatient spending; per episode, inpatient care is more expensive and perhaps less discretionary than outpatient care. Perhaps not surprisingly therefore, we find clear evidence of a response of outpatient spending to the future price, but we are unable to reject the null of no response of inpatient spending to the future price (appendix table A2 contains the results).

We also explored the robustness of our results to some of our modeling choices. The working paper version of this paper shows the robustness of our results to alternative functional forms for the dependent variables (Aron-Dine et al., 2012). In appendix table A3, we further show the robustness of our results to alternative choices of covariates regarding the firm and coverage tier fixed effects.

D. Assessing the Identifying Assumption

The key identifying assumption that allows us to interpret the results as a rejection of the null hypothesis that health care utilization responds only to the spot price of care is that there are no confounding differences in initial medical utilization among employees by their plan and join month. In other words, any differential patterns of initial medical utilization that we observe across plans by join month are caused by differences in expected end-of-year price. This identifying assumption might not be correct if for some reason individuals who join a particular plan in different months vary in their underlying propensity to use medical care.

In particular, one might be concerned that the same response to dynamic incentives that may lead to differential medical care utilization might also lead to differential selection into a deductible compared to a no-deductible plan on the basis of join month. However, if there are additional choice or preference parameters governing insurance plan selection that do not directly determine medical utilization, there may be no reason to expect such selection, even in the context of dynamically optimizing individuals. For example, if individuals anticipate the apparently large switching costs associated with subsequent reoptimization of plan choice (as in Handel, 2013), they might make their initial, midyear plan choice based on their subsequent optimal open enrollment

selection for a complete year. In such a case, we would not expect differential selection into plans by join month.

Ultimately, whether there is quantitatively important differential selection and its nature is an empirical question. The summary statistics in table 2 present some suggestive evidence that individuals may be (slightly) more likely to select the deductible plan relative to the no-deductible plan later in the year.¹⁶ Quantitatively, however, the effect is trivial; joining a month later is associated with only a 0.3 percentage point increase in the probability of choosing the deductible plan, or 0.9%.¹⁷ Consistent with a lack of strategic plan selection based on join month, we did not find differential selection in employees' plan choices in their second year at the firm out of the deductible plan relative to the no-deductible plan by join month.¹⁸ Nor did we find systematic differences in observable characteristics (i.e., age and gender) of individuals joining a deductible versus no-deductible plan by join month (shown in appendix table A4). Relatedly, row 2 of table 5 shows that our results are not sensitive to adding controls for the observable demographic characteristics of the employees: employee age, gender, and join year (see table 1).¹⁹

As another potential way to investigate the validity of the identifying assumption, we implement an imperfect placebo test by reestimating our baseline specification, equation (5), with the dependent variable as the "initial" medical utilization in the second year the employee is in the plan. In other words, we continue to define "initial medical utilization" relative to the join month (so that the calendar month in which we measure initial medical utilization varies in the same way as in our baseline specification across employees by join month), but we now measure it in the second year the employee is in the plan. For example, for employees who joined the plan in July 2004, we look at their medical spending from July through September 2005. In principle, when employees are in the plan for a full year, there should be no effect of join month (of the previous year) on their expected end-of-year price, and therefore no difference in initial utilization by join month across the plans. In practice, the test suffers from the problem that the amount of medical care consumed in the join year could influence (either positively or negatively) the amount consumed in the second year, either because of intertemporal substitution (which could generate negative serial correlation) or because medical care creates demand for further care (e.g., follow-up visits or further tests), which could generate positive serial correlation.

deductible plans spend past the deductible in the first three months. This is not a problem when the dependent variable is whether the individual has a first claim (since the spot price is the same for all individuals within a plan at the time of first claim). In practice, moreover, any upward bias is likely unimportant quantitatively. We estimate a virtually identical response to the future price when the dependent variable is based on two-month (instead of three-month) spending, even though the fraction hitting the deductible within the initial utilization period (and therefore the likely magnitude of the bias) drops by almost half. Moreover, there is no noticeable trend in the likelihood of hitting the deductible within the first three months by the join month. Hitting the deductible within a short time after enrollment therefore appears to be primarily determined by large and possibly nondiscretionary health shocks rather than an endogenous spending response to the future price.

¹⁶ Over the three join quarters shown in table 2, the share joining the deductible plan varies in Alcoa from 0.49 to 0.53 to 0.53, in firm B from 0.20 to 0.22 to 0.19, and in firm C from 0.38 to 0.40 to 0.44.

¹⁷ We regressed an indicator variable for whether the employee chose a deductible plan on the employee's join month (enumerated, as before, from 2 to 10), together with a dummy variable for coverage tier and firm fixed effects to parallel our main specification. The coefficient on join month is 0.0034 (SE 0.0018).

¹⁸ On average, only about 4% of employees change their plan in the second year, which is consistent with low rates of switching found in other work (Handel, 2013).

¹⁹ In keeping with the "within-firm" spirit of the entire analysis, we interact each of these variables with the firm fixed effects.

TABLE 5.—ROBUSTNESS AND SPECIFICATION CHECKS

Specification	N	Any Initial Claim		Initial Spending	
		Coefficient on p^f (1)	SE (2)	Coefficient on p^f (3)	SE (4)
1 Baseline	102,022	−0.126	(0.039)	−0.78	(0.27)
2 Control for demographics	102,014	−0.097	(0.034)	−0.66	(0.24)
3 Only those who remain for 2nd year	64,398	−0.131	(0.043)	−0.96	(0.33)
4 Dependent Variable measured in 2nd year	64,398	0.003	(0.037)	0.03	(0.28)
5 Dependent Variable measured January–March of 2nd year	64,398	−0.032	(0.035)	−0.33	(0.25)

The table reports results from alternative analyses of the relationship between initial medical utilization and expected end-of-year price. We report the coefficient on expected end-of-year price (p^f) from estimating equation (4) by IV; we use a simulated end-of-year price as an instrument for the expected end-of-year price (see text for more details). Row 1 shows the baseline results reported in the bottom row of table 4. Alternative rows report single deviations from this baseline as explained below. In row 2, we add controls for age, gender, and start year (as well as interactions of each of those with the firm fixed effects) to the baseline specification. In row 3, we estimate the baseline specification on a smaller sample of employees who remain in the firm for the entire subsequent year. In row 4, we estimate the baseline specification on this same sample, but defining the dependent variable based on utilization in the same three months of the subsequent year (i.e., their first full year in the firm). And in row 5, we estimate the baseline specification on this same sample but now define the dependent variable based on utilization in January to March of the first full year in the firm.

Row 3 of table 5 shows the baseline results limited to the approximately 60% of the employees who remain at the firm for the entire second year. We continue to find a statistically significant negative relationship between the future price and initial medical use in this smaller sample. For this subsample of employees, row 4 shows that when we measure initial utilization in the same three months of the second year, we find no statistically significant relationship between the future price and this alternative dependent variable; the point estimates are in fact positive and one to two orders of magnitude smaller than our baseline estimates. We interpret this as supportive of the identifying assumption.

Finally, in row 5 we investigate the extent to which the decrease in utilization in response to a higher future price represents intertemporal substitution of medical spending to the next year. Such intertemporal substitution would not be a threat to our empirical design—indeed, it might be viewed as evidence of another form of response to dynamic incentives—but it would affect the interpretation of our estimates and is of interest in its own right. We therefore rerun our baseline specification but now with the dependent variables measured in January to March of the second year. The results indicate that individuals who face a higher future price (and therefore consume less medical care) also consume less medical care in the beginning of the subsequent year, although the results are not statistically significant and are substantially smaller than our baseline estimates. This suggests that intertemporal substitution, in the form of postponement of care to the subsequent calendar year, is unlikely to be the main driver of the estimated decrease in care associated with a higher future price.

IV. Complementary Evidence from Medicare Part D

To further validate the main finding and the more general applicability of the empirical strategy, we provide complementary evidence of the role of dynamic incentives among elderly individuals enrolled in Medicare Part D prescription drug plans.

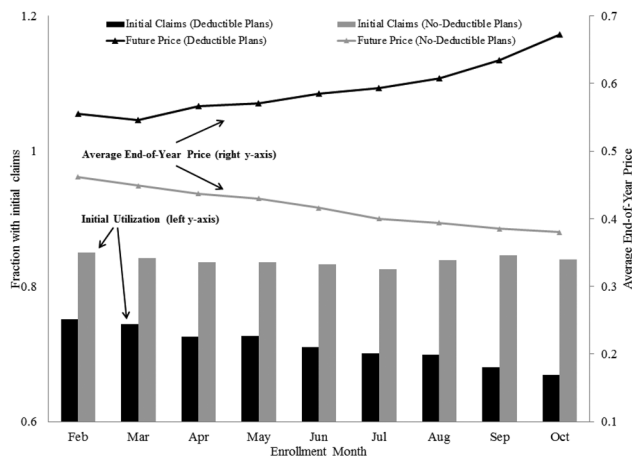
Medicare provides medical insurance to the elderly and disabled. Medicare Part D, which was introduced in 2006, provides prescription drug coverage. Enrollees in Part D can choose among different prescription drug plans, with different plan features and premiums, offered by private insurers. The key institutional feature that generates variation in the contract length is similar to the one we have exploited so far. Medicare Plan D provides annual coverage, which resets every January.²⁰ Most individuals become eligible for Part D coverage in the beginning of the month that they turn 65 (CMS, 2011), so during their first year in Medicare, individuals' coverage duration varies primarily due to variation in birth month.

Like employer-provided health insurance plans, prescription drug insurance plans in Part D create highly nonlinear budget sets, so that variation in the contract length in turn creates variation in the future price. In the spirit of the prior analysis, we can compare initial drug use for individuals who face the same spot price but different future prices for a reason that is plausibly unrelated to prescription drug use. Specifically, individuals who newly enroll in a given Part D plan when they turn 65 face the same initial spot price for drugs. However, because the insurance contract resets at the end of each calendar year, different individuals in the same Part D plan face different future prices depending on which month of the year they turn 65 and enroll in Part D. Furthermore, the sign and magnitude of the relationship between the month in which the individual joins the plan and the future price will vary depending on the type of plan.

Indeed, a particularly attractive feature of the Part D plans is that, unlike typical employer-provided health insurance contracts, in some plans, the future price is increasing in join month, while in others it is decreasing. For example, in the government-defined standard benefit design for 2008, the individual initially pays for all expenses out of pocket, until she has spent \$275, at which point she pays only 25%

²⁰ As in the employer-provided context, Medicare Part D choices are made during the open enrollment period in November and December of each year, and unless a specific qualifying event occurs, individuals cannot switch plans during the year.

FIGURE 2.—PROBABILITY OF INITIAL CLAIM AND EXPECTED END-OF-YEAR PRICE BY ENROLLMENT MONTH



The figure graphs the pattern of expected end-of-year price and of any initial drug claim by enrollment month for individuals in Medicare Part D during their first year of eligibility (once they turn age 65). We graph results separately for individuals in deductible and no-deductible plans. We calculate the expected end-of-year price separately for each individual based on plan and birth month, using all other individuals who enrolled in the same plan that month. The fraction with initial claim is measured as the share of individuals (by plan type and enrollment month) who had at least one claim over the first three months. The online appendix provides more details on the construction of variables used in this figure. $N = 137,536$ ($N = 108,577$ for no-deductible plans and $N = 28,959$ for deductible plans).

of subsequent drug expenditures until her total drug spending reaches \$2,510. At this point the individual enters the famed “doughnut hole,” or the gap, within which she must once again pay for all expenses out of pocket, until total drug expenditures reach \$5,726, the amount at which catastrophic coverage sets in and the marginal out-of-pocket price of additional spending drops substantially, to about 7%. In this plan, given the empirical distribution of drug spending, the end-of-year price is rising with enrollment month, since a shorter contract length means less chance to spend past the deductible into the lower-cost sharing arm. However, many individuals buy plans with no deductible, and here the expected end-of-year price tends to decline with join month since a shorter contract length creates less time to spend into the higher-cost sharing doughnut hole.

To operationalize the strategy, we use a 20% random sample of all Medicare Part D beneficiaries in 2007 through 2009. We observe the cost-sharing characteristics of each beneficiary’s plan, as well as detailed claim-level information on any prescription drugs purchased. We also observe basic demographic information (including age, gender, and eligibility for various programs tailored to low income individuals). Given the identification strategy, our analysis is limited to 65-year-olds who enroll between February and October. We make a number of other sample restrictions, including limiting to individuals who are in stand-alone prescription drug plans and are not dually eligible for Medicaid or other low-income subsidies. These restrictions are described in more detail in the appendix. Our analysis sample consists of about 138,000 beneficiaries.

Figure 2 summarizes the main empirical result graphically. We show the pattern of expected end-of-year price and initial drug use by enrollment month separately for beneficiaries

in two groups of plans: deductible and no-deductible plans. The expected end-of-year price depends on the cost-sharing features of the beneficiary’s plan, the number of months of the contract, and the individual’s expected spending. We measure initial drug use by whether the individual had a claim in the first three months of coverage.

Once again, the patterns of initial use by enrollment month present evidence against the null of no response to the dynamic incentives. In deductible plans, where the future price is increasing with enrollment month, initial utilization is decreasing with enrollment month. By contrast, in the no-deductible plan, where the future price is decreasing with enrollment month, the probability of an initial claim does appear to vary systematically with the enrollment month. A difference-in-difference comparison of the pattern of initial drug use by enrollment month for people in plans in which the future price increases with enrollment month relative to people in plans in which the future price decreases with enrollment month thus suggests that initial use is decreasing in the expected end-of-year price. The appendix provides more detailed and formal analyses that follow the same structure as the main analysis in the employer-provided context. We estimate a statistically significant elasticity of initial prescription drug claims with respect to the future price. Our estimated elasticity is about -0.25 , which is qualitatively similar to the -0.16 estimate for initial medical claims we estimated in the main analysis.

V. Discussion

Our results show that individuals’ decisions regarding medical utilization respond to the dynamic incentives associated with the nonlinear nature of health insurance contracts. This jointly indicates that individuals understand something about the nonlinear pricing schedule they face and that they take account of the future price in making current medical decisions. One clear implication of our results is that assuming that the spot price associated with a given medical treatment is the only relevant price may be problematic, and that ignoring the dynamic incentives associated with nonlinear contracts may miss an important component of the story in many contexts. But then what? How should we take into account the dynamic incentives in analyzing moral hazard effects of health insurance? In this section we offer some general discussion and guidance on the topic.

A. A Single Elasticity Estimate May Not Be Enough

A natural reaction to our findings might be that researchers should estimate the response of health care utilization to the future price. This is, however, unlikely to be a satisfactory solution. First, defining the future price is itself a somewhat delicate task; it would require modeling individuals’ expectations about their health shocks over the coming year and deciding what moments of the distribution of possible end-of-year prices are relevant to the individual’s decisions.

TABLE 6.—ILLUSTRATIVE EXERCISE USING THE RAND HIE DATA

Regressor	Dependent Variable					
	Any Initial Claim ^a		Log Annual Spending ^b		Annual Spending ^c	
	OLS	IV	OLS	IV	OLS	IV
Coinurance rate	−0.21 (0.13)	−0.25 (0.19)	−1.78 (0.73)	−1.65 (1.03)	−230 (855)	777 (936)
Share Hit MDE	0.02 (0.04)	0.01 (0.04)	0.20 (0.20)	0.21 (0.24)	525 (229)	699 (240)
Coinurance rate × Share Hit MDE	0.18 (0.33)	0.28 (0.46)	1.76 (1.78)	1.40 (2.53)	−561 (2,097)	−3122 (2,284)

The sample consists of 5,653 family-years (1,490 unique families) in the RAND data in one of the positive coinsurance plans or the free care plan. The table reports results estimating equation (6). The dependent variable is given in the column heading and the key independent variables in the left-hand column. “Share hit MDE” is the share of families in a given coinsurance and maximum dollar expenditure (MDE) plan who spend past the MDE during the year. In addition, because plan assignment in the RAND experiment was random only conditional on site and month of enrollment in the experiment, all regressions control for site and start month fixed effects (see Newhouse et al., 1993, appendix B for more details). All regressions cluster standard errors on the family. In the IV specifications, we instrument for the share of families in a given coinsurance and MDE plan who hit the MDE with the “simulated” share hitting the MDE; the simulated share is calculated as the share of the full ($N = 5,653$) sample, which, given their observed spending, would have hit the MDE if (counterfactually) assigned to the given plan; the coefficient on the instrument in the first stage is 1.05 (standard error 0.003); the F -statistic on the instrument is 120,000. Appendix table A9 provides more details on the plans in the RAND experiment, the distribution of the sample across the different plans, and the share of families who hit the MDE in each plan.

^aThe dependent variable is an indicator for at least one claim made by the employee or any covered family members in the first three months in the plan.

^bThe dependent variable is $\log(s + 1)$ where s is the total annual medical spending of the employee and any covered family members.

^cThe dependent variable is total annual medical spending of the employee and any covered family members.

will show, sample sizes and resultant power issues preclude us from estimating statistically significant results in this setting. Nonetheless, the example provides a useful template for how such variation, occurring naturally or through further randomized trials, could be used to better understand the impact of contract design on spending.

Appendix table A9 provides sample counts and various summary statistics for the RAND plans.²³ As the table shows, average MDEs were considerably higher in plans where the MDE was set equal to 10% or 15% of family income than in plans where the MDE was set to 5% of income. These differences generated corresponding differences in the share of families hitting the MDE and in expected end-of-year price (columns 5 and 6).

With no other assumptions than random assignment to plans, the experiment delivers estimates of spending as a function of the particular combinations of coinsurance rates and MDEs that families were randomized into. Appendix table A10 summarizes these outcomes by cell. As already mentioned, and illustrated in appendix table A10, the relatively small sample size of the RAND experiment makes it difficult to draw strong conclusions at this very granular level. But even if the results in appendix table A10 were less noisy, it seems unlikely that the objects of interest would be limited to precisely those combinations of coinsurance rates and MDEs that were present in the experiment. Presumably

the researcher would want to use the experiment in order to provide predictions for other contracts.

Consider, for example, the question of how medical spending would respond to changes in coinsurance rates in the context of contracts that had no MDE or a much higher MDE than observed in the experiment (which, as shown in appendix table A9, had about 20% to 30% of families hitting the MDE in a given year). The answer cannot be directly “read” out of the experimental results as it asks about contracts not observed in the experiment. However, with two sources of independent variation (in coinsurance rates and MDEs) the effect of contracts with higher MDEs—or more generally contracts with different combinations of coinsurance rates and MDEs—can be estimated by imposing a few additional assumptions.

We start by noting that the presence of MDEs likely influences the impact of coinsurance: an individual with higher coinsurance rates (who would presumably spend less initially) would likely hit the MDE faster and would therefore likely respond more to a change in the MDE. We therefore construct a new variable, *Share_Hit*, which is the share (within a coinsurance-MDE cell) that hits the MDE. We then project the experimental variation (appendix table A10) at the (coinsurance rate, MDE) level on variation at the (coinsurance rate, *Share_Hit*) level with the following functional for the regression we estimate:

$$y_{fj} = \eta_1 \times \text{coins}_j + \eta_2 \times \text{Share_Hit}_j + \eta_3 \times \text{coins}_j \times \text{Share_Hit}_j + v_{fj}, \quad (6)$$

where y_{fj} is a measure of medical utilization by family f in plan j , coins_j is the coinsurance rate of the plan the family was randomized into (which is either 0%, 25%, 50%, or 95%), and Share_Hit_j is as defined above (and shown in appendix table A9).

Table 6 shows these (illustrative and highly noisy) results. We focus our discussion on the point estimates. The first column in table 6 uses an indicator for initial claims, as earlier in the paper. The positive relationship between initial claims

maximum. They test whether spending is higher among families who are closer to hitting their out-of-pocket maximum, as would be expected, all else equal, if people respond to dynamic incentives. They make several modeling assumptions to try to address the (selection) problem that families with higher underlying propensities to spend are more likely to come close to hitting their out-of-pocket-maximum. They also assume that individuals have perfect foresight regarding all the subsequent medical expenses within a year associated with a given health shock. They conclude that they cannot reject the null of complete myopia with respect to future health shocks.

²³ The original RAND investigators have very helpfully made their data publicly available. We accessed the data through the Inter-University Consortium for Political and Social Research. Appendix table A9 omits RAND’s individual deductible plans, which had coinsurance rates of 0% for inpatient services and 100% or 95% for outpatient services, because there was no MDE variation within this coinsurance rate structure.

and the share who hit the MDE indicates a response to dynamic incentives. If only spot price mattered, initial claims should have been solely affected by the coinsurance rate. The second and third columns allow us to assess the effect of combination of contract design options on overall spending, using both log and level specifications. We see that a lower MDE or a lower coinsurance rate is associated with lower expenditure. This simply reflects an averaging of the experimental treatment effects. As we emphasized, however, our projection of the variables onto coinsurance rates and *Share_Hit* also allows us to use the experimental variation to predict the spending effect of other contracts, which are not precisely part of the experimental treatment. For example, the effect of a (counterfactual) linear contract could now be read off the estimates of η_1 in table 6, which captures the spending effect of the coinsurance rate conditional on eliminating the MDE (or, equivalently, setting *Share_Hit* = 0).

Such an approach could not always work, of course, and is limited to contract features for which there is experimental variation. If one would have liked, for instance, to investigate the response to adding a deductible, one would presumably need either additional data with variation in the deductible amount or to specify and estimate a more complete behavioral model from which one can obtain deeper behavioral primitives, as described earlier.

VI. Conclusion

In this paper we provide evidence that individuals respond to the dynamic incentives associated with the typical nonlinear nature of health insurance contracts in the United States. We do not view our results as particularly surprising. Like most other economists, we expect people to respond to incentives and are not terribly surprised when we find that they do. Yet our results do highlight what we view as a somewhat peculiar feature of the vast amount of work devoted to the estimation of moral hazard in health insurance. That is, although most of the literature focuses on estimating a single elasticity, our results suggest that it is unlikely that a single elasticity estimate can summarize the spending response to changes in health insurance. Our findings underscore that such an estimate is not conceptually well defined (there are likely at least two price elasticities that are relevant). As discussed in the last section, we expect future work in this area to make progress by either developing and estimating more complete models of the health utilization decisions or, alternatively, using independent variation in multiple contract features to estimate multiple elasticities that together can be used to approximate more credibly the spending effects of a richer set of (counterfactual) contract designs.

REFERENCES

- Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein, "The RAND Health Insurance Experiment, Three Decades Later," *Journal of Economics Perspectives* 27 (2013), 197–222.
- Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen, "Moral Hazard in Health Insurance: How Important Is Forward Looking Behavior?" NBER working paper 17802 (2012).
- Blundell, Richard, and Thomas MaCurdy, "Labor Supply: A Review of Alternative Approaches," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics* (Oxford: Elsevier North-Holland, 1999).
- Borenstein, Severin, "To What Electricity Price Do Consumers Respond? Residential Demand Elasticity under Increasing-Block Pricing," unpublished manuscript, Haas School of Business, UC Berkeley (2009).
- Burtless, Gary, and Jerry Hausman, "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment," *Journal of Political Economy* 86 (1978), 1103–1130.
- Cabral, Marika, "Claim Timing and Ex Post Adverse Selection," unpublished manuscript, Department of Economics, University of Texas, Austin (2013).
- Cardon, James H., and Igal Hendel, "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *RAND Journal of Economics* 32 (2001), 408–427.
- Chetty, Raj, John Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics* 126 (2011), 749–804.
- Chevalier, Judith, and Austan Goolsbee, "Are Durable Goods Consumers Forward-Looking? Evidence from College Textbooks," *Quarterly Journal of Economics* 124 (2009), 1853–1884.
- CMS, "Understanding Medicare Enrollment Periods," CMS product 11219 (2011).
- Currie, Janet, and Jonathan Gruber, "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *Quarterly Journal of Economics* 111 (1996), 431–466.
- Cutler, David M., and Richard J. Zeckhauser, "The Anatomy of Health Insurance" (pp. 563–643), in A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics* (Amsterdam: Elsevier, 2000).
- Dalton, Christina M., "Estimating Demand Elasticities Using Nonlinear Pricing," *International Journal of Industrial Organization* 37 (2014), 178–191.
- Einav, Liran, Amy Finkelstein, and Paul Schrimpf, "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D," NBER working paper 19393 (2013a).
- Einav, Liran, Amy Finkelstein, Stephen Ryan, Paul Schrimpf, and Mark R. Cullen, "Selection on Moral Hazard in Health Insurance," *American Economic Review* 103 (2013b), 178–219.
- Einav, Liran, Amy Finkelstein, and Paul Schrimpf, "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D," *Quarterly Journal of Economics* 130 (2015), 841–899.
- Ellis, Randall, "Rational Behavior in the Presence of Coverage Ceilings and Deductibles," *RAND Journal of Economics* 17 (1986), 158–175.
- Feldman, Naomi E., and Peter Katuscak, "Should the Average Tax Rate Be Marginalized?" CERGE-EI working paper 304 (2006).
- Grubb, Michael D., and Matthew Osborne, "Cellular Service Demand: Biased Beliefs, Learning and Bill Shock," *American Economic Review* 105 (2015), 234–271.
- Handel, Benjamin, "Adverse Selection and Inertia in Health Insurance Markets: The Nudging Hurts," *American Economic Review* 103 (2013), 2643–2682.
- Hausman, Jerry, "The Econometrics of Nonlinear Budget Sets," *Econometrica* 53 (1985), 1255–1282.
- Ito, Koichiro, "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review* 104 (2014), 537–563.
- Keeler, Emmett, Joseph P. Newhouse, and Charles Phelps, "Deductibles and the Demand for Medical Care Services: The Theory of a Consumer Facing a Variable Price Schedule under Uncertainty," *Econometrica* 45 (1977), 641–656.
- Keeler, Emmett B., and John E. Rolph, "The Demand for Episodes of Treatment in the Health Insurance Experiment," *Journal of Health Economics* 7 (1988), 337–367.
- Kowalski, Amanda, "Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance," NBER working paper 18108 (2012).
- Liebman, Jeffrey B., and Richard J. Zeckhauser, "Schmeduling," unpublished manuscript, Harvard Kennedy School, Harvard University (2004).
- Manning, Willard, Joseph Newhouse, Naihua Duan, Emmett Keeler, Arleen Leibowitz, and Susan Marquis, "Health Insurance and the Demand

- for Medical Care: Evidence from a Randomized Experiment,” *American Economic Review* 77 (1987), 251–277.
- Nevo, Aviv, John L. Turner, and Jonathan W. Williams, “Usage-Based Pricing and Demand for Residential Broadband,” unpublished manuscript, Department of Economics, Northwestern University (2013).
- Newhouse, Joseph P., and the Insurance Experiment Group, *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge, MA: Harvard University Press, 1993).
- Reiss, Peter C., and Matthew W. White, “Household Electricity Demand, Revisited,” *Review of Economic Studies* 72 (2005), 853–883.
- Saez, Emmanuel, “Do Taxpayers Bunch at Kink Points?” *American Economic Journal: Economic Policy* 2 (2010), 180–212.
- Yao, Song, Carl F. Mela, Jeongwen Chiang, and Yuxin Chen, “Determining Consumers’ Discount Rates with Field Studies,” *Journal of Marketing Research* 49 (2012), 822–841.