THE

QUARTERLY JOURNAL
OF ECONOMICS

Vol. CXXV August 2010 Issue 3

ESTIMATING WELFARE IN INSURANCE MARKETS USING VARIATION IN PRICES*

LIRAN EINAV
AMY FINKELSTEIN
MARK R. CULLEN

We provide a graphical illustration of how standard consumer and producer theory can be used to quantify the welfare loss associated with inefficient pricing in insurance markets with selection. We then show how this welfare loss can be estimated empirically using identifying variation in the price of insurance. Such variation, together with quantity data, allows us to estimate the demand for insurance. The same variation, together with cost data, allows us to estimate how insurers' costs vary as market participants endogenously respond to price. The slope of this estimated cost curve provides a direct test for both the existence and the nature of selection, and the combination of demand and cost curves can be used to estimate welfare. We illustrate our approach by applying it to data on employer-provided health insurance from one specific company. We detect adverse selection but estimate that the quantitative welfare implications associated with inefficient pricing in our particular application are small, in both absolute and relative terms.

*We are grateful to Felicia Bayer, Brenda Barlek, Chance Cassidy, Fran Filpovits, Frank Patrick, and Mike Williams for innumerable conversations explaining the institutional environment of Alcoa, to Colleen Barry, Susan Busch, Linda Cantley, Deron Galusha, James Hill, Sally Vegso, and especially Marty Slade for providing and explaining the data, to Tatyana Deryugina, Sean Klein, Dan Sacks, and James Wang for outstanding research assistance, and to Larry Katz (the Editor), three anonymous referees, Kate Bundorf, Raj Chetty, Peter Diamond, Hanming Fang, David Laibson, Jonathan Levin, Erzo Luttmer, Jim Poterba, Dan Silverman, Jonathan Skinner, and numerous seminar participants for helpful comments. The data were 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-upon 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. We gratefully acknowledge support from the NIA (R01 AG032449), the National Science Foundation Grant SES-0643037 (Einav), the Alfred P. Sloan Foundation (Finkelstein), and John D. and Catherine T. MacArthur Foundation Network on Socioeconomic Status and Health, and Alcoa, Inc. (Cullen).

The Quarterly Journal of Economics, August 2010

I. Introduction

The welfare loss from selection in private insurance markets is a classic result in economic theory. It provides, among other things, the textbook economic rationale for the near-ubiquitous government intervention in insurance markets. Yet there has been relatively little empirical work devoted to quantifying the inefficiency that selection causes in a particular insurance market, or the welfare consequences of potential policy interventions in that market. This presumably reflects not a lack of interest in this important topic, but rather the considerable challenges posed by empirical welfare analysis in markets with hidden information.

Recently, there have been several attempts to estimate the welfare costs of private information in particular insurance markets, specifically annuities (Einav, Finkelstein, and Schrimpf 2010) and health insurance (Bundorf, Levin, and Mahoney 2008; Carlin and Town 2009; Lustig 2009). These papers specify and estimate a structural model of insurance demand that is derived from the choices of optimizing agents, and recover the underlying (privately known) information about risk and preferences. This allows rich, out-of-sample, counterfactual welfare analysis. However, it requires the researcher to make critical assumptions about the nature of both the utility function and individuals' private information. These modeling choices can have nontrivial effects on the welfare estimates. Moreover, they are often specific to the particular market studied, making it difficult to compare welfare estimates meaningfully across markets or to readily adapt these approaches from one context to another.

Our objective in this paper is therefore to propose a complementary approach to empirical welfare analysis in insurance markets. We make fewer assumptions about the underlying primitives, yet impose enough structure to allow meaningful welfare analysis. These fewer assumptions come at the cost of limiting our welfare analyses to only those associated with the pricing of existing contracts.

We start in Section II by showing how standard consumer and producer theory—familiar to any student of intermediate micro—can be applied to welfare analysis of insurance markets with selection. As emphasized by Akerlof (1970) and Stiglitz (1987), among others, the key feature of markets with selection is that firms' costs depend on which consumers purchase their products. As a result, insurers' costs are endogenous to price. Welfare analysis therefore requires not only knowledge of how demand varies with price, but also information on how changes in price affect the costs of insuring the (endogenous) market participants. We use these insights to provide a particular graphical representation of the welfare cost of inefficient pricing arising from selection. We view these graphs as providing helpful intuition, and therefore as an important contribution of the paper. The graphs illustrate, among other things, how the qualitative nature of the inefficiency depends on whether the selection is adverse or advantageous.

Our graphical analysis also suggests a straightforward empirical approach to the welfare analysis of pricing in insurance markets. Section III shows how our framework translates naturally into a series of estimating equations, and discusses the data requirements. The key observation is that the same pricing variation that is needed to estimate the demand curve (or willingness to pay) in any welfare analysis—be it the consequences of tax policy, the introduction of new goods, or selection in insurance markets—can also be used to estimate the cost curve in selection markets, that is, how costs vary as the set of market participants endogenously changes. The slope of the estimated cost curve provides a direct test of the existence and nature of selection that—unlike the widely used " bivariate probit" test for asymmetric information (Chiappori and Salanie 2000)—is not affected by the existence (or lack thereof) of moral hazard. Specifically, rejection of the null hypothesis of a constant (i.e., horizontal) marginal cost curve allows us to reject the null hypothesis of no selection, whereas the sign of the slope of the marginal cost curve tells us whether the resultant selection is adverse (if marginal cost is increasing in price) or advantageous (if marginal cost is decreasing in price).

Most importantly, with both the demand and cost curves in hand, welfare analysis of inefficient pricing caused by any detected selection is simple and familiar. In the same vein, the estimates lend themselves naturally to welfare analysis of a range of counterfactual public policies that change the prices of existing contracts. These include insurance mandates, subsidies or taxes for private insurance, and regulation of the prices that private insurers can charge.

Our approach has several attractive features. First, it does not require the researcher to make (often difficult-to-test) assumptions about consumers' preferences or the nature of *ex ante* information. As long as we accept revealed preference, the demand and cost curves are sufficient statistics for welfare analysis of the pricing of existing contracts. In this sense, our approach is similar in spirit to Chetty (2008) and Chetty and Saez (2010), who show how key *ex post* behavioral elasticities are sufficient statistics for welfare analysis of the optimal level of public insurance benefits (see also Chetty [2009] for a more general discussion of the use of sufficient statistics for welfare analysis).

Second, our approach is relatively straightforward to implement, and therefore potentially widely applicable. In particular, although cost data are often quite difficult to obtain in many product markets (so that direct estimation of the cost curve is often a challenge), direct data on costs tend to be more readily available in insurance markets, because they require information on accident occurrences or insurance claims, rather than insight into the underlying production function of the firm. In addition, the omnipresent regulation of insurance markets offers many potential sources for the pricing variation needed to estimate the demand and cost curves. Third, the approach is fairly general, as it does not rely on specific institutional details; as a result, estimates of the welfare cost of adverse selection in different contexts may be more comparable.

These attractive features are not without cost. As mentioned already, the chief limitation of our approach is that our analysis of the welfare cost of adverse selection is limited to the cost associated with inefficient pricing of a fixed (and observed) set of contracts. Our approach therefore does not allow us to capture the welfare loss that adverse selection may create by distorting the set of contracts offered, which in many settings could be large.[1] At the end of Section III, we discuss in some detail the settings where this limitation may be less prohibitive.

Analysis of the welfare effects of distortions in the contract space due to selection—or of counterfactual public policies that introduce new contracts—requires modeling and estimating the structural primitives underlying the demand and cost curves, and it is in this sense that we view our approach as complementary to a full model of these primitives. We note, however, that although such richer counterfactuals are feasible with a more complete model of the primitives, in practice the existing papers (mentioned

1. A related limitation is that our approach forces us to rely on uncompensated (Marshallian) demand for welfare analysis. To account for income effects, we would need either to assume them away (by assuming constant absolute risk aversion) or to impose more structure and specify a full model of primitives that underlies the demand function.

above) that fully modeled these primitives have primarily confined themselves to welfare analyses of the pricing of existing contracts, as we do in this paper. This presumably reflects both researchers' (understandable) caution in taking their estimates too far out of sample, and the considerable empirical and theoretical challenges to modeling the endogenous contract response (Einav, Finkelstein, and Levin 2010). Perhaps similar reasons may also explain why many (although not all) government interventions in insurance markets tend to focus on the pricing of contracts, through taxes and subsidies, regulations, or mandates.

The last part of the paper (Section IV) provides an illustration of our approach by applying it to the market for employer-provided health insurance in the United States, a market of substantial interest in its own right. The existing empirical evidence on this market is consistent with asymmetric information (see Cutler and Zeckhauser [2000] for a review). However, until recently there has been relatively little empirical work on the welfare consequences of the detected market failure. Cutler and Reber (1998) are a notable exception. Like us, they analyze selection in employer-provided health insurance, and, like us, they estimate the demand curve. A key distinction, however, is that although they provide important and novel evidence of the existence of adverse selection in the market, they do not estimate the cost curve, which is crucial for welfare analysis.

We utilize rich individual-level data from Alcoa, Inc., a large multinational producer of aluminum and related products. We observe the health insurance options, choices, and medical insurance claims of its employees in the United States. We use the fact that, due to Alcoa's organizational structure, employees doing similar jobs in different sections of the company are faced with different prices for otherwise identical sets of coverage options. We verify that pricing appears orthogonal to the characteristics of the employees that the managers setting these prices can likely observe. Using this price variation, we estimate that marginal cost is increasing in price, and thus detect adverse selection in this market. However, we estimate the welfare costs associated with the inefficient pricing created by adverse selection to be small. Specifically, we estimate that in a competitive market the annual efficiency cost of this selection would be just below \$10 per employee, or about 3% of the total surplus at stake from efficient pricing. By way of comparison, this estimated welfare cost is an order of magnitude smaller than our estimate of the deadweight loss that would arise from monopolistic pricing in this market. We also estimate that the social cost of public funds for the price subsidy that would be required to move from the (inefficient) competitive equilibrium to the efficient outcome is about five times higher than our estimate of the welfare gain from achieving the efficient allocation. These results are robust across a range of alternative specifications.

It is extremely important to emphasize that there is no general lesson in our empirical findings. Our estimates are specific to our population and to the particular health insurance choices they face. Nonetheless, at a conceptual level, our findings highlight the importance of moving beyond detection of market failures to quantifying their welfare implications. Our particular findings provide an example of how it is possible for adverse selection to exist, and to impair market efficiency, without being easily remediable through standard public policies.

II. Theoretical Framework

II.A. Model

Setup and Notation. We consider a situation in which a given population of individuals are allowed to choose from exactly two available insurance contracts, one that offers high coverage (contract H) and one that offers less coverage (contract L). As we discuss in more detail below, it is conceptually straightforward to extend the analysis to more than two contracts, but substantially complicates the graphical presentation. To further simplify the exposition, we assume that contract L is no insurance and is available for free, and that contract H is full insurance. These are merely normalizations and straightforward to relax; indeed we do so in our empirical application.

A more important assumption is that we take the characteristics of the insurance contracts as given, although we allow the price of insurance to be determined endogenously. As we discuss in more detail in Section III, this seems a reasonable characterization of many insurance markets; it is often the case that the same set of contracts are offered to observably different individuals, with variation across individuals only in the pricing of the contracts, and not in offered coverage. Our analysis is therefore in the spirit of Akerlof (1970) rather than Rothschild and Stiglitz (1976), who endogenize the level of coverage as well.

We define the population by a distribution $G(\zeta)$, where ζ is a vector of consumer characteristics. A key aspect of the analysis is that we do not specify the nature of ζ; it could describe multidimensional risk factors, consumers' *ex ante* risk perception, and/or preferences. We denote the (relative) price of contract H by p, and denote by $v^H(\zeta_i, p)$ and $v^L(\zeta_i)$ consumer i's (with characteristics ζ_i) utility from buying contracts H and L, respectively. Although not essential, it is natural to assume that $v^H(\zeta_i, p)$ is strictly decreasing in p and that $v^H(\zeta_i, p = 0) > v^L(\zeta_i)$. Finally, we denote the expected monetary cost associated with the insurable risk for individual i by $c(\zeta_i)$. For ease of exposition, we assume that these costs do not depend on the contract chosen, that is, that there is no moral hazard. We relax this assumption in Section II.D, where we show that allowing for moral hazard does not substantively affect the basic analysis.

Demand for Insurance. We assume that each individual makes a discrete choice of whether to buy insurance or not. Because we take as given that there are only two available contracts and their associated coverages, demand is only a function of the (relative) price p. We assume that firms cannot offer different prices to different individuals. To the extent that firms can make prices depend on observed characteristics, one should think of our analysis as applied to a set of individuals that vary only in unobserved (or unpriced) characteristics. We assume that if individuals choose to buy insurance they buy it at the lowest price at which it is available, so it is sufficient to characterize demand for insurance as a function of the lowest price p.

Given the above assumptions, individual i chooses to buy insurance if and only if $v^H(\zeta_i, p) \geq v^L(\zeta_i)$. We can define $\pi(\zeta_i) \equiv \max\left\{p : v^H(\zeta_i, p) \geq v^L(\zeta_i)\right\}$, which is the highest price at which individual i is willing to buy insurance. Aggregate demand for insurance is therefore given by

$$D(p) = \int 1(\pi(\zeta) \geq p)\, dG(\zeta) = \Pr(\pi(\zeta_i) \geq p), \tag{1}$$

and we assume that the underlying primitives imply that $D(p)$ is strictly decreasing, continuous, and differentiable.

Supply and Equilibrium. We consider $N \geq 2$ identical risk-neutral insurance providers, who set prices in a Nash equilibrium (*à la* Bertrand). Although various forms of imperfect competition may characterize many insurance markets, we choose to focus on the case of perfect competition as it represents a natural benchmark for welfare analysis of the efficiency cost of selection; under perfect competition, symmetric information leads to efficient outcomes, so that any inefficiency can be attributed to selection and does not depend on the details of the pricing model. We note, however, that it is straightforward to replicate the theoretical and empirical analysis for any other *given* model of the insurance market, including models of imperfect competition.

We further assume that when multiple firms set the same price, individuals who decide to purchase insurance at this price choose a firm randomly. We also assume that the only costs of providing contract H to individual i are the insurable costs $c(\zeta_i)$.[2] The foregoing assumptions imply that the average (expected) cost curve in the market is given by

$$\text{(2)} \quad \mathrm{AC}(p) = \frac{1}{D(p)} \int c(\zeta) 1(\pi(\zeta) \geq p)\, dG(\zeta) = E(c(\zeta) \mid \pi(\zeta) \geq p).$$

Note that the average cost curve is determined by the costs of the sample of individuals who endogenously choose contract H. The marginal (expected) cost curve[3] in the market is given by

$$\text{(3)} \qquad \mathrm{MC}(p) = E(c(\zeta) \mid \pi(\zeta) = p).$$

In order to straightforwardly characterize equilibrium, we make two further simplifying assumptions. First, we assume that there exists a price $\overline{p}$ such that $D(\overline{p}) > 0$ and $\mathrm{MC}(p) < p$ for every $p > \overline{p}$. In words, we assume that it is profitable (and efficient, as we will see soon) to provide insurance to those with the highest willingness to pay for it.[4] Second, we assume that if there exists $\underline{p}$ such that $\mathrm{MC}(\underline{p}) > \underline{p}$, then $\mathrm{MC}(p) > p$ for all $p < \underline{p}$. That is, we assume that $\mathrm{MC}(p)$ crosses the demand curve at most once.[5] It

2. Note that $c(\zeta_i)$ reflects only direct insurer claims (i.e., payout) costs, and not other administrative (production) costs of the insurance company. We discuss in Section III.B how such additional costs can be incorporated into the analysis.

3. Note that there could be multiple marginal consumers. Because price is the only way to screen in our setup, all these consumers will together average (point-by-point) to form the marginal cost curve.

4. This assumption seems to hold in our application. Bundorf, Levin, and Mahoney (2008) make the interesting observation that there are contexts where it may not hold.

5. In the most basic economic framework of insurance the difference between $\pi(\zeta)$ and $\mathrm{MC}(\zeta)$ is the risk premium, and is positive for risk-averse individuals. If all individual are risk-averse, $\mathrm{MC}(\zeta)$ will never cross the demand curve. In practice, however, there are many reasons for such crossing. Those include, among others,

is easy to verify that these assumptions guarantee the existence and uniqueness of equilibrium. In particular, the equilibrium is characterized by the lowest break-even price, that is,

$$p^* = \min\{p : p = \mathrm{AC}(p)\}. \tag{4}$$

II.B. Measuring Welfare

We measure consumer surplus by the certainty equivalent. The certainty equivalent of an uncertain outcome is the amount that would make an individual indifferent between obtaining this amount for sure and obtaining the uncertain outcome. An outcome with a higher certainty equivalent therefore provides higher utility to the individual. This welfare measure is attractive as it can be measured in monetary units. Total surplus in the market is the sum of certainty equivalents for consumers and profits of firms. Throughout we ignore any income effects associated with price changes.[6]

Denote by $e^H(\zeta_i)$ and $e^L(\zeta_i)$ the certainty equivalents for consumer i of an allocation of contract $H$and L, respectively; under the assumption that all individuals are risk-averse, the willingness to pay for insurance is given by $\pi(\zeta_i) = e^H(\zeta_i) - e^L(\zeta_i) > 0$. We can write consumer welfare as

$$\mathrm{CS} = \int [(e^H(\zeta) - p)1(\pi(\zeta) \geq p) + e^L(\zeta)1(\pi(\zeta) < p)]\, dG(\zeta) \tag{5}$$

and producer welfare as

$$\mathrm{PS} = \int (p - c(\zeta))1(\pi(\zeta) \geq p)\, dG(\zeta). \tag{6}$$

Total welfare will then be given by

$$\begin{aligned} \mathrm{TS} = \mathrm{CS} + \mathrm{PS} = \int & [(e^H(\zeta) - c(\zeta))1(\pi(\zeta) \geq p) \\ & + e^L(\zeta)1(\pi(\zeta) < p)]\, dG(\zeta). \end{aligned} \tag{7}$$

---

loading factors on insurance, moral hazard, and horizontal product differentiation. As a result, it may not be socially efficient for all individuals to have insurance, even if they are all risk-averse.

6. In a textbook expected-utility framework, this is equivalent to assuming that the utility function exhibits constant absolute risk aversion (CARA). When the premium changes are small relative to the individual's income (as in the choice we study in our empirical application below), it seems natural to view CARA as a reasonable approximation. An alternative would be to fully specify the underlying utility function, from which income effects can be derived. This is one additional limitation of our simpler approach.

It is now easy to see that it is socially efficient for individual i to purchase insurance if and only if

$$\pi(\zeta_i) \geq c(\zeta_i). \tag{8}$$

In other words, in a first-best allocation individual i purchases insurance if and only if his willingness to pay is at least as great as the expected social cost of providing the insurance to him.[7]

In many contexts (including our application below), price is the only instrument available to affect the insurance allocation. In such cases, achieving the first best may not be feasible if there are multiple individuals with different $c(\zeta_i)$'s who all have the same willingness to pay for contract H (see footnote 3). It is therefore useful to define a constrained efficient allocation as the one that maximizes social welfare subject to the constraint that price is the only instrument available for screening. Using our notation, this implies that it is (constrained) efficient for individual i to purchase contract H if and only if

$$\pi(\zeta_i) \geq E(c(\widetilde{\zeta}) \mid \pi(\widetilde{\zeta}) = \pi(\zeta_i)), \tag{9}$$

that is, if and only if $\pi(\zeta_i)$ is at least as great as the expected social cost of allocating contract H to *all* individuals with willingness to pay $\pi(\zeta_i)$. We use this constrained efficient benchmark throughout the paper, and hereafter refer to it simply as the efficient allocation.[8]

II.C. Graphical Representation

We use the framework sketched above to provide a graphical representation of adverse and advantageous selection. Although the primary purpose of doing so is to motivate and explain the empirical estimation strategy, an important ancillary benefit of these graphs is that they provide what we believe to be helpful intuition for the efficiency costs of different types of selection in insurance markets.

7. Implicit in this discussion is that insurer claims $c(\zeta_i)$ represent the full social cost associated with allocating insurance to individual i. To the extent that this is not the case, for example, due to positive or negative externalities associated with insurance or imperfections in the production of the underlying good that is being insured, our measure of welfare would have to be adjusted accordingly.

8. See Greenwald and Stiglitz (1986), who analyze efficiency in an environment with a similar constraint. See also Bundorf, Levin, and Mahoney (2008), who investigate the efficiency consequences of relaxing this constraint. In a symmetric-information case, the first best could be achieved by letting prices fully depend on $\pi(\zeta_i)$ and $c(\zeta_i)$.

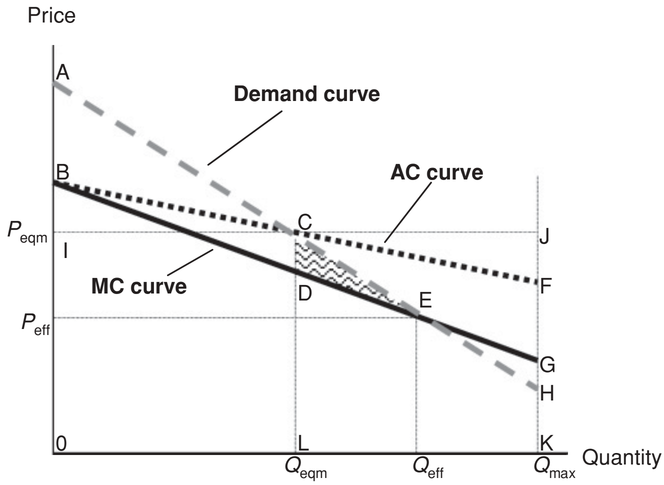

FIGURE I

Efficiency Cost of Adverse Selection

This figure represents the theoretical efficiency cost of adverse selection. It depicts a situation of adverse selection because the marginal cost curve is downward-sloping (i.e., increasing in price, decreasing in quantity), indicating that the people who have the highest willingness to pay also have the highest expected cost to the insurer. Competitive equilibrium is given by point C (where the demand curves intersects the average cost curve), whereas the efficient allocation is given by point E (where the demand curve intersects the marginal cost curve). The (shaded) triangle CDE represents the welfare cost from underinsurance due to adverse selection.

Adverse Selection. Figure I provides a graphical analysis of adverse selection. The relative price (or cost) of contract H is on the vertical axis. Quantity (i.e., share of individuals in the market with contract H) is on the horizontal axis; the maximum possible quantity is denoted by Q_{max}. The demand curve denotes the relative demand for contract H. Likewise, the average-cost (AC) curve and marginal-cost (MC) curve denote the average and marginal *incremental* costs to the insurer from coverage with contract H relative to contract L.

The key feature of adverse selection is that the individuals who have the highest willingness to pay for insurance are those who, on average, have the highest expected costs. This is represented in Figure I by drawing a downward-sloping MC curve. That is, marginal cost is increasing in price and decreasing in quantity. As the price falls, the marginal individuals who select contract H have lower expected cost than inframarginal individuals, leading to lower average costs. The essence of the private-information problem is that firms cannot charge individuals based on their (privately known) marginal costs, but are instead restricted to charging a uniform price, which in equilibrium implies average-cost pricing. Because average costs are always higher than marginal costs, adverse selection creates underinsurance, a familiar result first pointed out by Akerlof (1970). This underinsurance is illustrated in Figure I. The equilibrium share of individuals who buy contract H is Q_{eqm} (where the AC curve intersects the demand curve), whereas the efficient number is $Q_{\text{eff}} > Q_{\text{eqm}}$ (where the MC curve intersects the demand curve).

The welfare loss due to adverse selection is represented by the shaded region CDE in Figure I. This represents the lost consumer surplus from individuals who are not insured in equilibrium (because their willingness to pay is less than the average cost of the insured population) but whom it would be efficient to insure (because their willingness to pay exceeds their marginal cost). One could similarly evaluate and compare welfare under other possible allocations. For example, mandating that everyone buy contract H generates welfare equal to the area ABE minus the area EGH. This can be compared to welfare at the competitive equilibrium (area ABCD), welfare at the efficient allocation (area ABE), welfare from mandating everyone to buy contract L (normalized to zero), or the welfare effect of policies that subsidize (or tax) the equilibrium price. The relative welfare rankings of these alternatives are an open empirical question. A primary purpose of the proposed framework is to develop an empirical approach to assessing welfare under alternative policy interventions (including the no-intervention option).

Advantageous Selection. The original theory of selection in insurance markets emphasized the possibility of adverse selection and the resultant efficiency loss from underinsurance (Akerlof 1970; Rothschild and Stiglitz 1976). Consistent with this theory, the empirical evidence points to several insurance markets, including health insurance and annuities, in which the insured have higher average costs than the uninsured. However, a growing body of empirical evidence suggests that in many other insurance markets, including life insurance and long-term care insurance, there exists “advantageous selection”: Those with more insurance have lower average costs than those with less or no insurance. Cutler, Finkelstein, and McGarry (2008) provide a review of the evidence of adverse and advantageous selection in different insurance markets.

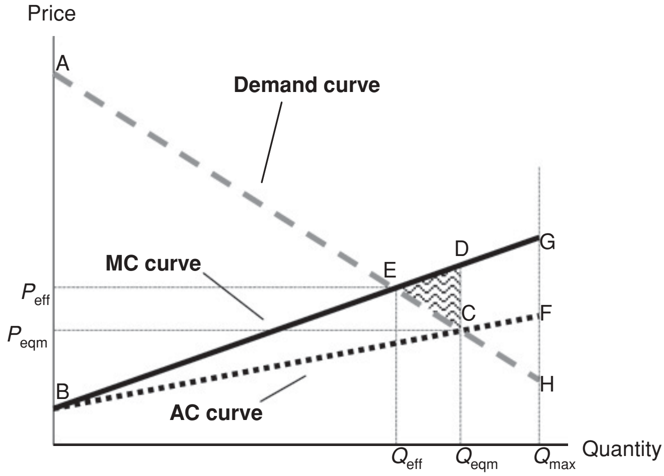

FIGURE II
Efficiency Cost of Advantageous Selection

This figure represents the theoretical efficiency cost of advantageous selection. It depicts a situation of advantageous selection because the marginal cost curve is upward-sloping, indicating that the people who have the highest willingness to pay have the lowest expected cost to the insurer. Competitive equilibrium is given by point C (where the demand curve intersects the average cost curve), whereas the efficient allocation is given by point E (where the demand curve intersects the marginal cost curve). The (shaded) triangle CDE represents the welfare cost from overinsurance due to advantageous selection.

Our framework makes it easy to describe the nature and consequences of advantageous selection. Figure II provides a graphical representation. In contrast to adverse selection, with advantageous selection the individuals who value insurance the most are those who have, on average, the lowest expected costs. This translates to upward-sloping MC and AC curves. Once again, the source of market inefficiency is that consumers vary in their marginal cost, but firms are restricted to uniform pricing, and at equilibrium, price is based on average cost. However, with advantageous selection, the resultant market failure is one of overinsurance rather than underinsurance (i.e., $Q_{eff} < Q_{eqm}$ in Figure II), as was pointed out by de Meza and Webb (2001), among others. Intuitively, insurance providers have an additional incentive to reduce price, as the inframarginal customers whom they acquire as a result are relatively good risks. The resultant welfare loss is given by the shaded area CDE, and represents the excess of MC over willingness to pay for individuals whose willingness to pay exceeds the average costs of the insured population. Once again, we can also easily evaluate welfare of different situations in Figure II, including mandating contract H (the area ABE minus the area EGH), mandating contract L (normalized to zero), competitive equilibrium (ABE minus CDE), and efficient allocation (ABE).

Sufficient Statistics for Welfare Analysis. These graphical analyses illustrate that the demand and cost curves are sufficient statistics for welfare analysis of equilibrium and nonequilibrium pricing of existing contracts. In other words, different underlying primitives (i.e., preferences and private information, as summarized by ζ) have the same welfare implications if they generate the same demand and cost curves. [9]

This in turn is the essence of our empirical approach. We estimate the demand and cost curves but remain agnostic about the underlying primitives that give rise to them. As long as individuals' revealed choices can be used for welfare analysis, the precise *source* of selection is not germane for analyzing the efficiency consequences of the resultant selection, or the welfare consequences of public policies that change the equilibrium price.

The key to any counterfactual analysis that uses the approach we propose is that insurance contracts are taken as given, and only their prices vary. Thus, for example, the estimates generated by our approach can be used to analyze the effect of a wide variety of standard government interventions in insurance markets that change the price of insurance. These include mandatory insurance coverage, taxes and subsidies for insurance, regulations that outlaw some of the existing contracts, regulation of the allowable price level, and regulation of allowable pricing differences across observably different individuals. However, more structure and assumptions would be required if we were to analyze the welfare effects of introducing insurance contracts not observed in the data.

II.D. Incorporating Moral Hazard

Thus far we have not explicitly discussed any potential moral-hazard effect of insurance. This is because moral hazard does not fundamentally change the analysis, but only complicates the presentation. We illustrate this by first discussing the baseline case in which we define a contract H to be full coverage and

9. Note that we have placed no restrictions in Figure I or II on the nature of the underlying consumer primitives ζ_i. Individuals may well differ on many unobserved dimensions concerning their information and preferences. Nor have we placed any restriction on the nature of the correlation across these primitives.

contract L to be no coverage. Here, moral hazard has no effect on the welfare analysis. We then discuss the slight modification needed when we allow contract L to include some partial coverage.

With moral hazard, the expected insurable cost for individual i is now a function of his contract choice, because coverage may affect behavior. We therefore define two (rather than one) expected monetary costs for individual i. We denote by $c^H(\zeta_i)$ individual i's expected insurable costs under contract H relative to contract L when he behaves as if covered by contract H. Similarly, we define $c^L(\zeta_i)$ to be individual i's expected insurable costs under contract H relative to contract L when he behaves as if covered by contract L. That is, $c^j(\zeta_i)$ always measures the incremental insurable costs under contract H compared to contract L, whereas the superscript j denotes the underlying behavior, which depends on coverage. We assume throughout that $c^H(\zeta_i) \geq c^L(\zeta_i)$; this inequality will be strict if and only if moral hazard exists. As a result, we now have two marginal cost curves, MC^H and MC^L, and two corresponding average cost curves, AC^H and AC^L (with MC^H and AC^H always weakly higher than MC^L and AC^L, respectively).

In contrast to the selection case, a social planner generally has no potential comparative advantage over the private sector in ameliorating moral hazard (i.e., in encouraging individuals to choose socially optimal behavior). Our welfare analysis of selection therefore takes any moral hazard effect as given. We investigate the welfare cost of the inefficient pricing associated with selection or the welfare consequences of particular public policy interventions given any existing moral-hazard effects, just as we take as given other features of the environment that may affect willingness to pay or costs.

To explicitly recognize moral hazard in our foregoing equilibrium and welfare analysis, one can simply replace $c(\zeta_i)$ everywhere above with $c^H(\zeta_i)$, and obtain the same results. Recall, as emphasized earlier, that the cost curve is defined based on the costs of individuals who endogenously buy contract H (see equation (2)); in the new notation their costs are given by $c^H(\zeta_i)$ because they are covered by contract H (and behave accordingly). Thus, $c^L(\zeta_i)$ is largely irrelevant. The intuition from the firm perspective is clear: the insurer's cost is only affected by the behavior of insured individuals, and not by what their behavior would be if they were not insured. From the consumer side, $c^L(\zeta_i)$ does matter. However, it matters only because it is one of the components that affect the willingness to pay for insurance. As we showed already, willingness to pay (π) and cost to the insurer (c^H) are sufficient statistics for the equilibrium and welfare analysis. Both can be estimated without knowledge of $c^L(\zeta_i)$. Therefore, as long as moral hazard is taken as given, it is inconsequential to break down the willingness to pay for insurance into a part that arises from reduction in risk and a part that arises from a change in behavior.

The one substantive difference once we allow for moral hazard is that the assumption that contract L involves no coverage is no longer inconsequential. Once contract L involves some partial coverage, it is no longer the case that all potential moral-hazard effects of contract H on insurable expenditures are internalized by the provider of contract H through their impact on c^H. To see this, we first note that when contract L involves some coverage, the market equilibrium can be thought of as one in which firms offering contract H compete only on the incremental coverage in excess of L.[10] Welfare analysis of the allocation of contract H must now account for the potential negative externality that coverage by contract H inflicts on the insurer providing contract L (through increased cost). This conceptual point does not pose practical difficulties for our framework. With estimates of the moral hazard effect, the welfare gain of providing contract H to individual i is simply smaller by the amount of the increased insurable costs for the provider of contract L that are associated with the change of behavior. As we discuss in more detail in Section III, our approach points to a natural way by which moral hazard can be estimated (and therefore incorporated into the welfare analysis when contract L involves some partial coverage).

III. Estimation

III.A. *The Basic Framework*

Applying our framework to estimating welfare in an insurance market requires data that allow estimation of the demand curve $D(p)$ and the average cost curve $\mathrm{AC}(p)$. The marginal cost curve can be directly backed out from these two curves and does

10. One natural example is that of contract L as the public health insurance program Medicare and contract H as the supplemental private Medigap insurance that covers some of the costs not covered by Medicare.

not require further estimation. To see this, note that

(10)
$$\mathrm{MC}(p) = \frac{\partial \mathrm{TC}(p)}{\partial D(p)} = \frac{\partial(\mathrm{AC}(p)\cdot D(p))}{\partial D(p)} = \left(\frac{\partial D(p)}{\partial p}\right)^{-1} \frac{\partial(\mathrm{AC}(p)\cdot D(p))}{\partial p}.$$

With these three curves—$D(p)$, $\mathrm{AC}(p)$, and $\mathrm{MC}(p)$—in hand, we can straightforwardly compute welfare under various allocations, as illustrated in Figures I and II.

As is standard, estimating the demand curve requires data on prices and quantities (i.e., coverage choices), as well as identification of price variation that can be used to trace out the demand curve. This price variation has to be exogenous to unobservable demand characteristics. To estimate the $\mathrm{AC}(p)$ curve we need, in addition, data on the expected costs of those with contract H, such as data on subsequent risk realization and how it translates to insurer costs. With such data we can then use the same variation in prices to trace out the $\mathrm{AC}(p)$ curve. Because expected cost is likely to affect demand, any price variation that is exogenous to demand is also exogenous to insurable cost. That is, we do not require a separate source of variation.

With sufficient price variation, no functional form assumptions are needed for the prices to trace out the demand and average cost curves. For example, if the main objective is to estimate the efficiency cost of inefficient pricing arising from selection, then price variation that spans the range between the market equilibrium price (point C in Figures I and II) and the efficient price (point E) allows us to estimate the welfare cost of the inefficient pricing associated with selection (area CDE) without making any restrictions on the shape of the demand or average cost curves. With pricing variation that does not span these points, the area CDE can still be estimated, but will require some extrapolation based on functional form assumptions.

III.B. Extensions

As mentioned, the basic framework we described in Section II made a number of simplifying assumptions for expositional purposes that do not limit the ability to apply this approach more broadly. It is straightforward to apply the approach to the case where contract H provides less than full coverage and/or where contract L provides some coverage. We discuss a specific example of this in our application below. In such settings we must simply be clear that the cost curve of interest is derived from the average *incremental* costs to the insurance company associated with providing contract H rather than providing contract L. For the welfare analysis, we must also be sure to incorporate any moral-hazard effects of contract H on the costs to the insurers providing contract L. We discussed above conceptually how to adjust the welfare analysis; later in this section we describe how to estimate the moral-hazard effect of contract H.

Likewise, although it was simpler to present the graphical analysis with only two coverage options, the approach naturally extends to more than two contracts. The data requirements would simply extend to having price, quantity, and costs for each contract, as well as pricing variation across all relevant relative prices, so that the entire demand and average cost systems can be estimated. Specifically, with N available contracts, one could normalize one of these contracts to be the reference contract, define incremental costs (and price) of each of the other contracts relative to the reference contract, and estimate a system $D(p)$ and $\mathrm{AC}(p)$, where demand, prices, and average costs are now $(N-1)$-dimensional vectors. As in the two-contract case, competitive equilibrium (defined by each contract breaking even) will be given by the vector of prices that solves $p = \mathrm{AC}(p)$. From the estimated systems $D(p)$ and $\mathrm{AC}(p)$ one can also back out the system of marginal costs $\mathrm{MC}(p)$, which defines the marginal costs associated with each price vector. We can then solve $p = \mathrm{MC}(p)$ for the efficient price vector and integrate $D(p) - \mathrm{MC}(p)$ over the (multidimensional) difference between the competitive and the efficient price vectors to obtain the welfare cost of the inefficient pricing associated with selection.[11]

Finally, we note that the estimated demand and cost curves are sufficient statistics for welfare analysis of equilibrium allocations of existing contracts generated by models other than the one we have sketched. This includes, for example, welfare analysis of other equilibria, such as those generated by imperfect competition rather than our benchmark of perfect competition. It also

11. Although conceptually straightforward, implementation of our approach with more than two contracts will likely encounter, in practice, a number of subtle issues. For example, with multiple contracts the system $\mathrm{AC}(p) = p$ or $\mathrm{MC}(p) = p$ may have more scope for multiple or no solutions, and the definition of "adverse selection" or "advantageous selection" may now be more subtle (see Einav, Finkelstein, and Levin [2010] for more discussion of this latter point). In addition, from an empirical standpoint, estimating entire demand and cost systems may be more challenging (e.g., in terms of the variation required) than estimating one-dimensional demand and cost curves.

includes welfare analysis of markets with other production functions, which may include fixed or varying administrative costs of selling more coverage, rather than our benchmark of no additional costs beyond insurable claims. This is because, as the discussion of estimation hopefully makes clear, we do not use assumptions about the equilibrium or the production function to estimate the demand and cost curves. An assumption of a different equilibrium simply requires calculation of welfare relative to a different equilibrium point (point C in the graphs). Similarly, if one has external information (or beliefs) about the nature of the production function, one can use this to shift or rotate the estimated cost curve, and calculate the new equilibrium and efficient points.

III.C. A Direct Test of Selection

Although the primary focus of our paper is on estimating the welfare cost of inefficient pricing associated with selection, our proposed approach also provides a direct test for the existence and nature of selection. This test is based on the slope of the estimated marginal-cost curve. A rejection of the null hypothesis of a constant marginal-cost curve allows us to reject the null of no selection.[12] Moreover, the sign of the slope of the estimated marginal-cost curve informs us of the nature of any selection; a downward-sloping marginal-cost curve (i.e., a cost curve declining in quantity and increasing in price) indicates adverse selection, whereas an upward-sloping curve indicates advantageous selection. This is a useful test, because detecting the existence of selection is a necessary precursor to analysis of its welfare effects.

Importantly, our "cost curve" test of selection is unaffected by the existence (or lack thereof) of moral hazard. This is a distinct improvement over the influential " bivariate probit" (a.k.a. "positive correlation") test of Chiappori and Salanie (2000), which has been widely used in the insurance literature. This test, which compares realized risks of individuals with more and less insurance coverage, jointly tests for the existence of either selection or moral hazard (but not for each separately). Identifying price variation—which is not required for the "positive correlation" test—is the key to our distinct test for selection. It allows us to analyze how the

12. Using the terminology we defined in Section II.B, a flat marginal-cost curve implies that the equilibrium outcome is constrained efficient. It does not, however, imply that the equilibrium is first-best. Finkelstein and McGarry (2006) present evidence on an insurance market that may exhibit a flat cost curve (no selection) but does not achieve the first-best allocation.

risk characteristics of the sample that select a given insurance contract vary with the price of that contract.

To see why our cost curve test is not affected by any potential moral hazard, note that the AC curve is estimated using the sample of individuals who choose to buy contract H at a given price. As we vary price we vary this sample, but everyone in the sample always has the same coverage. Because by construction the coverage of individuals in the sample is fixed, our estimate of the slope of the cost curve (our test of selection) is not affected by moral hazard (which determines how costs are affected as coverage changes). Of course, part of the selection reflected in the slope of the cost curve may reflect selection based on differences across individuals in the anticipated impact of coverage on costs (i.e., the moral hazard effect of coverage). We still view this as a selection effect, representing selection into contracts based on the anticipated incentive effects of these contracts.

III.D. Estimating Moral Hazard

Our framework also allows us to test for and quantify moral hazard. One way to measure moral hazard is by the difference between $c^H(\zeta_i)$—individual i's expected insurable cost when covered by contract H—and $c^L(\zeta_i)$—individual i's expected insurable cost when covered by contract L. That is, $c^H(\zeta_i) - c^L(\zeta_i)$ is the moral hazard effect from the insurer's perspective, or the increased cost to the insurer from providing contract H that is attributable to the change in behavior of covered individuals. We already discussed how identifying price variation can be used to estimate the AC and MC curves, which we denote by AC^H and MC^H when moral hazard is explicitly recognized. With data on the costs of the uninsured (or less insured, if contract L represents some partial coverage), we can repeat the same exercise to obtain an estimate for AC^L and MC^L. That is, we can use the same identifying price variation to estimate demand for contract L and to estimate the AC^L curve from the (endogenously selected) sample of individuals who choose contract L. We can then back out the MC^L curve analogously to the way we back out the MC^H curve, using of course the demand curve for contract L and the AC^L curve (rather than the demand for contract H and the AC^H curve) in translating average costs into marginal costs (see equation (10)). The (point-by-point) vertical difference between MC^H and MC^L curves provides an estimate of moral hazard. A test of whether this difference is positive is a direct test for moral hazard, which is valid whether adverse selection is present or not.[13]

Of course, it is not a new observation that an exogenous shifter of insurance coverage (which in our context comes from pricing) facilitates the estimation of moral hazard. However, our proposed approach to estimating moral hazard (compared to, say, a more standard instrumental-variable framework) allows us to estimate (with sufficiently rich price variation) heterogeneous moral-hazard effect and to see how moral hazard varies across individuals with different willingness to pay $\pi(\zeta_i)$ or different expected costs $c^H(\zeta_i)$.

III.E. Applicability

In the next section we turn to a specific application of our proposed framework, which illustrates the mechanics of the approach as well as producing results that may be of interest in their own right. Here we discuss more generally the types of settings in which our approach might be applicable.

Two main requirements need to be met to use our approach sensibly. First, it has to be feasible to estimate the demand and cost curves credibly. This requires data on insurance prices, quantities, and insurer's costs, as well as identifying variation in prices. The required data elements of insurance options and choices and subsequent risk realization are not particularly stringent; researchers have already demonstrated considerable success in a wide range of insurance markets in obtaining such data.[14] Indeed, a nice feature of welfare analysis in insurance markets is that cost data are much easier to obtain than in many other markets, because they involve information on accident occurrences or insurance claims, rather than insight into the underlying production function of the firm.

13. The exercise we have just described would provide an estimate of the moral-hazard effect from the insurer's perspective. One might be interested in other measures of moral hazard, such as the effect of insurance on total spending rather than on insurer costs. The test of moral hazard can be applied in the same manner using other definitions of $c(\zeta_i)$. The same statement of course applies to our "cost curve" selection analysis; for the purpose of analyzing equilibrium and market efficiency, we have estimated selection from the insurer perspective, but again the approach could be used to measure selection on any other outcome of interest.

14. Examples include auto and homeowner's insurance (Cohen and Einav 2007; Barseghyan, Prince, and Teitelbaum 2010; Sydnor 2010), annuities (Finkelstein and Poterba 2004), long term–care insurance (Finkelstein and McGarry 2006), health insurance (Eichner, McLellan, and Wise 1998), and many others.

Identifying variation in prices is a considerably stronger empirical hurdle, although the near-ubiquitous regulation of insurance markets provide numerous potential opportunities. Although our application below assumes that prices are set exogenously to unobservable demand (and cost) characteristics, alternative research designs that isolate credible identifying variation, such as an instrumental-variable approach, would do. For example, state regulation of private insurance markets has created variation in the prices charged to different individuals at a point in time as well as over time (Blackmon and Zeckhauser 1991; Buchmueller and DiNardo 2002; Bundorf and Simon 2006). Tax policy is another useful potential source of pricing variation. For example, a large literature has documented (and used) the substantial variation in the tax subsidy for employer-provided health insurance (see Gruber [2002] for a review). Beyond the opportunities provided by public policy, researchers have also found useful pricing variation stemming from field experiments (Karlan and Zinman 2009) and specific idiosyncrasies of firm pricing behavior.[15] More generally, common instruments used in demand analysis, such as changes in the competitive environment (Lustig 2009) or perhaps shifters in the administrative costs of handling claims, could serve as the requisite source for identifying price variation. The validity of this variation for identification is of course a key issue, which can and should be evaluated in specific applications. Indeed, we see the transparency of our approach in this regard as an important attraction.

The second key requirement for applying our proposed framework stems from its focus on inefficient pricing. Given that it is designed to estimate the welfare consequences of pricing of *existing* contracts, it is best suited to settings in which the market or public policy response to asymmetric information will primarily manifest itself through pricing of observed contracts rather than other aspects of contract design. We note that a pricing response also covers mandating a specific (observed) contract or the elimination of certain contracts, which is of course equivalent to pricing a subset of the contracts at their "virtual price," at which

15. Examples include firm experimentation with pricing policy (Cohen and Einav 2007), discrete pricing policy changes (Adams, Einav, and Levin 2009), idiosyncratic pricing decisions made by human resource managers (Cutler and Reber 1998), and the nonlinearities and discontinuities associated with rules that firms use to risk adjust individuals' premiums (Abbring, Chiappori, and Pinquet 2003; Israel 2004).

demand for these contracts is zero; of course, credible applications in such settings would require price variation around the virtual price, which may be more difficult to find. However, our approach cannot accommodate a market or policy response that leads to the introduction of new contracts, which were not previously observed.

How closely a given setting fits this bill needs to be evaluated case by case. Perhaps the ideal setting is one in which regulation (or some other constraint) explicitly prevents firms from redesigning contracts. Although rare, examples exist. One such case is the (limited) set of contracts that can be offered in the Medigap market, the private health insurance that supplements Medicare. Since 1992, these contracts have been set by national regulation: private firms may decide which of the specified contracts to offer and at what price, but they cannot design and introduce new contracts (see, e.g., Fox, Rice, and Alecxih [1995]). A related example is the application we discuss below in which company headquarters design the coverage options and print the brochures that describe them, whereas different subsidiaries are allowed (some) choice over the relative pricing of these options.

A likely more common setting that doesn't quite fit this ideal standard but may come sufficiently close is the practice in many markets of first settling on the contract design, and then adjusting only prices over time and across individuals. For example, the Medicare Part D market (for subsidized prescription drug coverage for the elderly) divides the country into thirty-four geographical markets. Providers that operate in multiple markets (and most of them do) have designed and advertised a single (national) set of coverage plans (in terms of formularies, deductible, cost sharing, etc.), and adjust only their prices by region (Keating 2007). Similarly, in annuity markets, companies offer identical sets of contracts (in terms of tilt of payments and guaranteed payment features), with only the annuity rates varying with the annuitant mortality profile (Finkelstein and Poterba 2002).

IV. Empirical Illustration: Employer-Provided Health Insurance

IV.A. *Data and Environment*

We illustrate the approach we have just outlined using individual-level data from 2004 on the U.S.-based employees (and their dependents) at Alcoa, Inc. The primary purpose of the application is to show how the theoretical framework can be mapped into empirical welfare estimates. We view the direct link between the theoretical framework and the empirical estimates—and the resulting transparency this provides for evaluating the strengths and weaknesses of the empirical results—as a key strength of our approach.

In 2004 Alcoa had approximately 45,000 active employees in the United States, working at about 300 different job sites in 39 different states. At that time, in an effort to control healthcare spending, Alcoa introduced a new set of health insurance options for virtually all its salaried employees and about one-half of its hourly employees. We analyze the choices and claims of employees offered the new set of options in 2004.[16]

The data contain the menu of health insurance options available to each employee, the employee premium associated with each option, the employee's coverage choice, and detailed claim-level information on all the employee (and any covered dependents') medical expenditures during the coverage period. [17] Crucially, as we discuss below, the data contain plausibly exogenous variation in the prices of the insurance contracts offered to otherwise similar employees within the company. Finally, the data contain rich demographic information, including the employee's age, race, gender, annual earnings, and job tenure at the company and the number and ages of other insured family members. We suspect that we observe virtually everything about the employee that the administrators setting insurance premiums can observe without direct personal contact, as well as some characteristics that the price setters might not be able to observe (such as detailed medical expenditure information from previous years; this information is administered by a third party). This is important because it allows us to examine whether the variation in prices across employees appears correlated with the employee characteristics that could potentially influence the price setters' decisions.

16. Over the subsequent several years, most of the remaining hourly employees were transitioned to the new health insurance options as their union contracts expired. The variation over time in the contracts offered is not well suited to the approach developed here, which relies on variation in the pricing of the same set of contract offerings. Busch et al. (2006) study the effect of the change in plan options between 2003 and 2004 on the use of preventive care.

17. Health insurance choices are made during the "open enrollment" period at the end of 2003 and apply for all of 2004. We also observe medical expenditure in 2003 if the employee worked at the company for all of 2003.

We restrict our baseline analysis to a subsample of employees for whom the pricing variation is cleaner and the setting follows the theoretical framework more closely. Our baseline sample consists of 3,779 salaried employees with family coverage who chose one of the two modal health insurance choices: a higher and a lower level of PPO coverage (we refer to these hereafter as contract H and contract L and provide more details about them in Section IV.C). The Online Appendix provides many more details about these sample restrictions, provides results for other coverage tiers, and addresses concerns of sample selection.

IV.B. Variation in Prices

Company Structure as the Source of Variation. An essential element in the analysis is that there is variation across employees in the relative price they face for contract H and that this variation is unrelated to the employees' willingness to pay for contract H and to their insurable costs. We believe that Alcoa's business structure provides a credible source of such pricing variation across different employees in the company.

In 2004, as part of the new benefit design, company headquarters offered a set of seven different possible pricing menus for employee benefits. The coverage options are the same across all the menus, but the prices (employee premiums) associated with these options vary. For our purposes, the key element of interest is the incremental (annual) premium the employee must pay for contract H relative to contract L, $p = p_H - p_L$. We refer to this incremental premium as the "price" in everything that follows. There were six different values of p in 2004 (as two of the seven menus were identical in this respect), ranging (for family coverage) from \$384 to \$659.[18]

Which price menu a given employee faces is determined by the president of his business unit. Alcoa is divided into approximately forty business units. Each business unit has essentially complete independence to run its business in the manner it sees fit, provided that it does so ethically and safely, and at or above the company's normal rate of return. Failure on any of these dimensions

18. The annual pretax employee premium for contract H was around \$1,500 for family coverage, although of course it ranged across the different menus. The incidence of being offered a menu with a lower average price level (across different options) may well be passed on to employees in the form of lower wages (Gruber 1994). This is one additional reason that it is preferable to focus the analysis on the difference in premiums for the different coverage options, rather than the level of premiums.

can result in the replacement of the unit's president. Business units are typically organized by functionality—such as primary metals, flexible packaging materials, rigid packaging materials, or home exteriors—and are independent of geography. There are often multiple business units in the same state. The number of active employees in a business unit ranges from the low teens (in "government affairs") to close to six thousand (in "primary metals"). The median business unit has about 500 active employees. The business unit president may choose different price menus for employees within his unit based on their location (job site) and their employment type (salaried or hourly employee and, if hourly, which union the employee is in, if any).

As a result of this business structure, employees doing the same job in the same location may face different prices for their health insurance benefits due to their business unit affiliations. *A priori*, it struck us as more plausible that the pricing variation across salaried employees in different business units was more likely to be useful for identification—reflecting idiosyncratic characteristics of the business unit presidents rather than differences in the demand or costs of salaried employees in the different business units—than the pricing variation across hourly employees. This is because many of the jobs that salaried employees do are quite similar across business units. Thus, for example, accountants, paralegals, administrative assistants, electrical engineers, or metallurgists working in the same state may face different prices because their benefits were chosen by the president of the "rigid packaging" business unit, rather than by the president of "primary metals." By comparison, the nature of the hourly employees' work (which often involves the operation of particular types of machinery) is more likely to differ across different units, and may depend on what the business unit is producing. For example, the work of the potroom operators stirring molten metal around in large vats in the "primary metals" business unit is likely to be different from the work of the furnace operators in the "rigid packaging" unit.

Examination of Assumption of Exogenous Pricing. The available data appear consistent with this basic intuition. Table I compares mean demographic characteristics of employees in our baseline sample (all of whom are salaried) who face different prices. In general, the results look quite balanced. There is no substantive or statistically significant difference across employees

TABLE I
ASSESSING THE EXOGENEITY OF THE PRICE VARIATION

	Faced lowest relative price (2,939 employees) (1)	Faced higher relative prices (840 employees) (2)	Difference (3)	Coefficient (4)	p-value (5)
Age (mean)	42.74	42.40	0.33	−0.245	.31
Tenure (mean)	13.02	11.63	1.39	−0.565	.08
Fraction male	0.862	0.852	0.009	1.268	.79
Fraction white	0.874	0.825	0.049	−6.998	.40
Log(annual salary) (mean)	11.16	11.05	0.11	−8.612	.17
Spouse age (mean)	41.37	41.05	0.32	−0.200	.41
Number of covered family members (mean)	4.14	4.07	0.07	−1.400	.36
Age of youngest covered child (mean)	9.81	9.41	0.40	−0.3	.26
2003 medical spending (in US$)[a]					
All (mean)	7,027	5,922	1,105	−0.0001	.09
In most common 2003 plan (mean)	6,938	5,967	971	−0.0001	.10

Notes. The table reports average differences in covariates (shown in the left column) across employees who face different relative prices for the higher-coverage option in the baseline sample. The employee characteristics in the left column represent contemporaneous 2004 characteristics (except where noted). Note that everyone with family coverage has a covered spouse and at least one covered child. Columns (1) and (2) present, respectively, average characteristics for the approximately three-fourths of employees who faced the lowest relative price ($384; see Table II) and the remaining one-fourth who face one of the five higher relative prices ($466 to $659; see Table II). Column (3) shows the difference between columns (1) and (2). Columns (4) and (5) report, respectively, the coefficient and p-value from a regression of the (continuous) relative price variable (in US$) on the characteristic given in the left column; we adjust the standard errors for an arbitrary variance covariance matrix within each state.

[a]In the bottom two rows we look at 2003 medical spending for all employees in the sample who were in the data in 2003 (2,600 and 658 employees in columns (1) and (2), respectively), and for all employees who were in the data in 2003 in the most common 2003 health insurance plan (2,282 and 523 employees in columns (1) and (2), respectively). The latter attempts to avoid potential differences in spending arising from moral hazard effects of different 2003 coverages.

who face different prices in average age, fraction male, fraction white, average (log) wages, average age of spouse, number of covered family members, or age of the youngest child. The two possible exceptions to this general pattern are average job tenure and average 2003 medical expenditures (which we show both for all of our sample who was working in 2003 and when restricted to employees in the most common plan in 2003, to avoid potential differences in spending arising from moral-hazard effects of different 2003 coverages).[19] A joint F-test of all of the coefficients leaves us unable to reject the null that they are jointly uncorrelated with price.[20] The inference is similar when we include state fixed effects or extend the sample to include all coverage tiers (rather than family coverage only) or all salaried employees (rather than just the two-thirds who choose the two modal coverage options).

Ancillary support for the quantitative evidence we have just described comes from our qualitative investigation into benefit selection at Alcoa in 2004. Importantly, this was the first year ever when business unit presidents had the opportunity to make decisions regarding the relative prices of insurance contracts for their employees. Therefore, although one might suspect that over time their price selection might become more sophisticated with respect to demand or expected costs (which would invalidate our identification assumption), in the first year the decision makers had relatively little information or experience to go by. Relatedly, the new benefit system represented the first time in the company's history that it was possible to charge employees a substantial incremental price for greater health insurance coverage. Our discussions with the company suggested that many business unit presidents were (at least initially) philosophically opposed to charging employees much for (generous) health insurance coverage, which may explain why (as seen in Table II), about three-fourths of the salaried employees ended up facing the lowest possible incremental price that the business unit presidents were allowed to choose. Perhaps because of this, after 2004, Alcoa headquarters no longer gave the business unit presidents a choice on benefit prices, and chose a

19. We should note, of course, that when ten different variables are tested the p-value should be adjusted upward to take account of the multiple hypothesis testing, so that the p-values we report are too small.

20. When we examine the eight contemporaneous characteristics we obtain an F-stat of 1.71 (p-value = .14). When we also include 2003 spending for those in the same plan as a ninth covariate (so that our sample size falls by about 25%) we obtain an F-stat of 0.95 (p-value = .50).

(uniform) pricing structure with a price higher than any of the options available in 2004.

Interestingly, the story looks very different for hourly employees. A similar analysis of covariates for hourly employees suggests statistically significant differences across employees who face different prices. As noted, this is not surprising given the institutional environment, and motivates our sample restriction to salaried employees. Indeed, the fact that prices for hourly employees are not uncorrelated with employee characteristics is somewhat reassuring; in a large for-profit company, it makes sense to expect clear differences in employee characteristics to be reflected in the prices chosen. It may be that when there was more at stake (in terms of cost differences across employees) the business unit presidents paid more attention to setting prices and less to their idiosyncratic philosophical views. It is also possible—although we have no direct evidence for this—that the business unit presidents had fundamentally different objectives in setting prices for hourly and for salaried employees.

Thus, although we would of course prefer to be able to isolate the precise source of our pricing variation, we are nonetheless reassured by both the quantitative and qualitative evidence that the prices faced by salaried employees appear uncorrelated with their predictors of demand or costs. Of course, we are able to examine only whether prices are correlated with observable differences across salaried employees. We cannot rule out potential unobservable differences, for example, in the "culture" of the business unit, which could potentially affect price setting and be correlated with either demand or costs.

IV.C. Empirical Strategy and Relationship with the Theoretical Framework

As before, we denote by $p_i = p_i^H - p_i^L$ the relative price that employee i faces, where p_i^j is employee i's annual premium if he chose coverage j. We define D_i to be equal to 1 if employee i chooses contract H and 0 if employee i chooses contract L. Finally, we let m_i be a vector representing total medical expenditures of employee i and any covered family members in 2004.

Coverage Characteristics and Construction of the Cost Variable. In our theoretical discussion in Section II we defined (for simplicity) contract H to be full coverage and contract L to be no coverage. As a result, we could refer to c_i as the *total* cost to the insurance company from covering employee i. When contract H is not full coverage and contract L provides some partial coverage, the relevant cost variable (denoted c_i) is defined as the *incremental* cost to the insurer from providing contract H relative to providing contract L, holding m_i fixed. Specifically, let $c(m_i; H)$ and $c(m_i; L)$ denote the cost to the insurance company from medical expenditures m_i under contracts H and L, respectively. The incremental cost is then given by $c_i \equiv c(m_i) = c(m_i; H) - c(m_i; L)$. The AC curve is computed by calculating the average c_i for all individuals who choose contract H at a given relative price p (see equation (2)) and estimating how this average c_i varies as the relative price varies. We can observe $c(m_i; H)$ directly in the data, but $c(m_i; L)$ must be computed counterfactually using the claims data and the plan rules of contract L . For consistency, we calculate both $c(m_i; H)$ and $c(m_i; L)$ from plan rules.

Construction of c_i requires detailed knowledge of each plan's benefits as well as individuals' realized medical claims. This allows us to construct the cost to the insurance company of insuring medical expenditures m_i under any particular plan j. The two contracts we focus on vary only in their consumer cost-sharing rules. Specifically, contract L coverage has higher deductibles and higher out-of-pocket maximums.[21] The data are quite detailed and the plan rules are fairly simple, allowing us to calculate $c(m_i; j)$ with a great deal of accuracy. For example, for individuals with contract H the correlation between their actual (observed) share of out-of-pocket spending (out of total expenditure) and our constructed share is over 0.97. The Online Appendix provides more detail on our calculation of c_i.

Figure III presents the major differences in consumer cost sharing between the two coverage options. Cost-sharing rules differ depending on whether spending is in-network or out-of-network. Figure IIIa shows the annual out-of-pocket spending (on the vertical axis) associated with a given level of total medical spending m (on the horizontal axis) for each coverage option, assuming the medical spending is in-network. In network, contract H has no deductible whereas contract L has a \$500 deductible. Both contracts have a 10% coinsurance rate, and the out-of-pocket

21. The plans are similar in all other features, such as the network definition and the benefits covered. As a result, we do not have to worry about differences between contarcts H and L in plan features that might differ in unobservable ways across employees (for example, differences in providers or relative network quality).

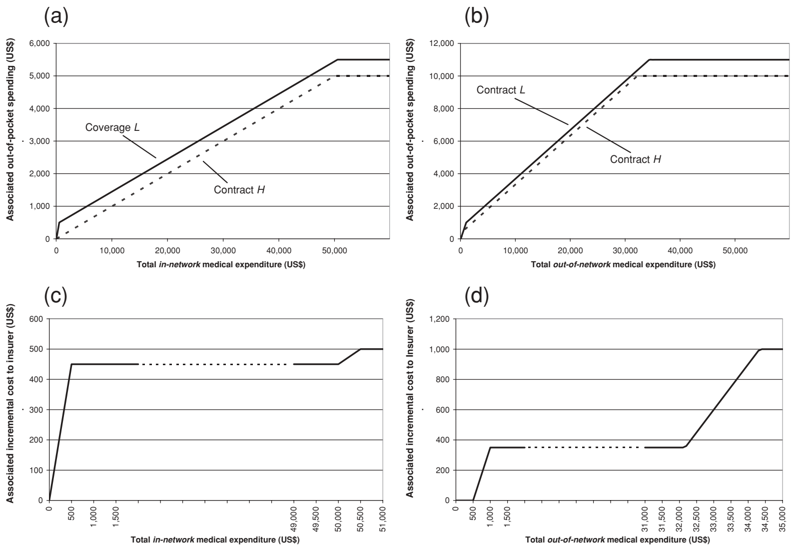

FIGURE III

Description of Contract H and Contract L

(a) and (b) present the main features of contract H (dashed) and contract L (solid) family coverages offered by the company, which are based on a deductible and an out-of-pocket maximum. (c) and (d) present the corresponding cost differences to the insurer from providing the contract H instead of contract L, for a given level of medical expenditure. In other words, (c) and (d) illustrate the in-network and out-of-network components of the constructed variable $c_i(m)$. (a) and (c) describe the rules for in-network medical spending (deductibles of \$0 and \$500, and out-of-pocket maxima of \$5,000 and \$5,500 for contracts H and L, respectively), and (b) and (d) describe the rules for out-of-network medical spending (deductibles of \$500 and \$1,000, and out-of-pocket maxima of \$10,000 and \$11,000 for contracts H and L, respectively). Coinsurance rates for both contracts are 10% (in network) and 30% (out of network). There is no interaction between the in-network and out-of-network coverage (i.e., each deductible and out-of-pocket maximum must be satisfied separately). The Online Appendix provides more details on the coverage rules and our construction of $c_i(m)$.

maximum is \$5,000 for contract H and \$5,500 for contract L. Figure IIIb presents the analogous graph for out-of-network spending, which has higher cost-sharing requirements under both plans. Although the vast majority of spending (96%) occurs in network, about 25% of the individuals in our baseline sample file at least one claim out of network, making the out-of-network coverage an important part of the analysis.[22]

22. There is no interaction between the in-network and out-of-network coverage. Each deductible and out-of-pocket maximum must be satisfied separately.

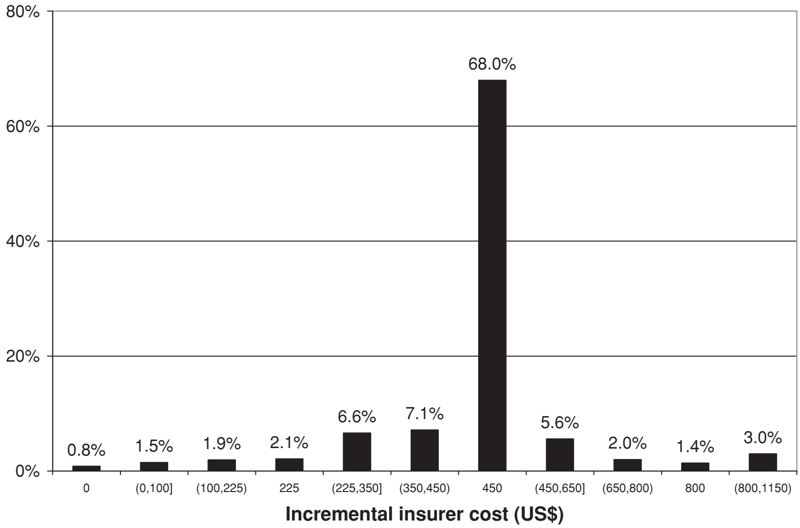

FIGURE IV

The Distribution of the Insurer's Incremental Costs (c_i)

This figure presents the distribution of the incremental insurer cost (c_i) for all 3,779 employees in our baseline sample. Note that the distribution has several mass points that are driven by the kinked formula of the coverages (Figure III). The largest mass point is at \$450, with about two-thirds of the sample. This point represents individuals who spent more than \$500 and less than \$50,000 in network, and less than \$500 out of network.

Figures IIIc and IIId show the implied difference in out-of-pocket spending between contracts H and L, for a given level of annual medical expenditure m_i. In other words, they illustrate the in-network and out-of-network (respectively) components of the constructed variable $c_i(m)$. Figure IV presents the empirical distribution of the constructed c_i variable. The distribution of c_i reflects the various kinks in the coverage plans presented in Figure III. The most visible example is that about two-thirds of the individuals in our baseline sample have $c_i = 450$. This represents individuals who had between \$500 and \$50,000 in-network (total) medical expenditures and less than \$500 out-of-network expenditures.[23]

23. Note also that, as emphasized by Figure IV, because our cost variable captures the incremental cost of increased coverage (rather than total medical expenditures) it is not heavily influenced by outliers (catastrophic expenditures). Indeed, as shown in Figure III, plan rules essentially cap incremental costs at \$1,500.

The nature of the plan differences is important for understanding the margin on which we may detect selection (or moral hazard). Empirically, because only a few people spend anywhere close to the out-of-pocket maximum of either contract, the difference in insurer's cost between the plans is primarily attributable to differences in the deductible. In terms of selection, this suggests that the differences in the plans could matter for the insurance choice of anyone with positive expected expenditures, and is increasing as expected expenditures increase. In terms of moral hazard, this suggests that if individuals are forward-looking and have perfect foresight, then differences in behavior for people covered by the different plans should be limited to the small percentage (9%) of employees who have total medical expenditures that are less than the contract L deductible.

Baseline Estimating Equations. For our baseline specification, we estimate the demand and average cost functions using OLS, assuming that the demand and cost curves are each linear in prices. That is, we estimate the two equations

$$D_i = \alpha + \beta p_i + \epsilon_i, \tag{11}$$

$$c_i = \gamma + \delta p_i + u_i, \tag{12}$$

where, as described earlier, D_i is a dummy variable that is equal to 1 if employee i chose contract H and equal to 0 if i chose contract L, c_i is the realized incremental cost to the insurer from covering individual i with contract H rather than contract L (see the Online Appendix for more details on the construction of c_i), and p_i is the incremental annual premium that employee i is required to pay to purchase contract H (rather than contract L). In all regressions, we adjust the standard errors to allow for an arbitrary variance-covariance matrix within each state. This is to allow for potential correlation in the residuals of the demand or cost equations across salaried employees in the same state. Following the theoretical framework, the demand equation is estimated on the entire sample, whereas the (average) cost equation is estimated on the sample of individuals who (endogenously) choose contract H.

Using the point estimates from the above regressions, we can construct our predicted demand and average cost curves and other estimates of interest. Following equation (10), the marginal cost curve is given by

$$\mathrm{MC}(p) = \frac{1}{\beta}\left(\frac{\partial\,(\alpha+\beta p)(\gamma+\delta p)}{\partial p}\right) = \frac{1}{\beta}\,(\alpha\delta+\gamma\beta+2\beta\delta p)$$

$$= \frac{\alpha\delta}{\beta}+\gamma+2\delta p. \tag{13}$$

With the demand curve, AC curve, and MC curve in hand, we can find where they intersect and compute any area of interest between them. In our baseline (linear) specification, the intersection points and areas of interest can be computed using simple geometry. The equilibrium price and quantity are given by equating $\mathrm{AC}(p) = D(p)$, resulting in $P_{\text{eq}} = \gamma/(1-\delta)$ and $Q_{\text{eq}} = \alpha + \beta(\gamma/(1-\delta))$. The efficient price and quantity are given by equating $\mathrm{MC}(p) = D(p)$, resulting in $P_{\text{eff}} = 1/(1-2\delta)(\frac{\alpha\delta}{\beta}+\gamma)$ and $Q_{\text{eff}} = \alpha + 1/(1-2\delta)(\alpha\delta+\beta\gamma)$. The efficiency cost associated with competitive pricing (measured by the area of triangle CDE in Figure I) is then given by

$$\Delta_{\text{CDE}} = \frac{1}{2}(Q_{\text{eff}}-Q_{\text{eq}})(P_{\text{eq}}-\mathrm{MC}(P_{\text{eq}})) = \frac{-\delta^2}{2(1-2\delta)\beta}\left(\alpha+\frac{\beta\gamma}{1-\delta}\right)^2. \tag{14}$$

In the Online Appendix we also report results from other, nonlinear specifications, in which we compute these price, quantity, and welfare estimates numerically.

IV.D. Baseline Results

Our baseline specification estimates the linear demand and cost curves shown in equations (11) and (12) on our baseline sample. This allows us to walk through the main conceptual points of interest in applying our proposed approach. In the Online Appendix we provide a more thorough and detailed discussion of empirical issues specific to our context, including alternative samples and specifications.

Table II shows the raw data for our key variables. The (relative) price ranges from \$384 to \$659, with about three-fourths of the sample facing the lowest price. Column (3) shows that the propensity to choose contract H is generally declining with price and ranges from 0.67 to 0.43. Column (4) shows that the average costs of the (endogenously selected) individuals who select contract H is generally increasing with price (or equivalently,

TABLE II
THE EFFECT OF PRICE ON DEMAND AND COSTS

(Relative) price ($) (1)	Number of employees (2)	Fraction chose contract H (3)	Average incremental cost ($) for those covered under: Contract H (4)	Contract L (5)
384	2,939	0.67	451.40	425.48
466	67	0.66	499.32	423.30
489	7	0.43	661.27	517.00
495	526	0.64	458.60	421.42
570	199	0.46	492.59	438.83
659	41	0.49	489.05	448.50

Notes. The table presents the raw data underlying our baseline estimates. All individuals face one of six different (relative) prices, each represented by a row in the table. Column (2) reports the number of employees facing each price, and column (3) reports the fraction of them who chose contract H. Columns (4) and (5) report (for individuals covered by contracts H and L, respectively) the average incremental costs to the insurer of covering these individuals with contract H rather than with contract L, taking the family's medical expenditures as given. The graphical analog to this table is presented by the circles shown in Figure V.

declining in quantity). This pattern of average costs indicates the existence of adverse selection (see Figure I). Column (5) shows the same for the individuals who (endogenously) select contract L. Recall that incremental cost is defined as the difference in costs to the insurer associated with a given employee's family's medical expenditures if those expenditures were insured under contract H rather than contract L. As shown in Figure III, this difference is a nonlinear function of expenditures.

In the spirit of the "positive correlation" test (Chiappori and Salanie 2000), a comparison of columns (5) and (4) reveals consistently higher average costs for those covered by contract H than for those covered by contract L. This indicates that either moral hazard or adverse selection is present. Detecting whether selection is present, and if so what its welfare consequences are, requires the use of our pricing variation, to which we now turn.

In column (1) of Table III we report OLS estimates of equation (11) with no additional controls. We obtain a downward-sloping demand curve, with a (statistically significant) slope coefficient β of -0.00070. This implies that a $100 increase in price reduces the probability that the employee chooses contract H by a statistically significant seven percentage points, or about 11%.

In column (2) of Table III we use OLS to separately estimate the average cost curve in equation (12). We obtain a (statistically

TABLE III
ESTIMATION RESULTS

Dependent variable (sample)	1 if chose High (both High and Low) (1)	Incremental cost (only High) (2)
Panel A: Estimation results		
Relative price of High (US$)	−0.00070	0.15524
	(0.00032)	(0.06388)
	[.034]	[.021]
Constant	0.940	391.690
	(0.123)	(26.789)
	[.000]	[.000]
Mean dependent variable	0.652	455.341
Number of observations	3,779	2,465
R^2	.008	.005
Panel B: Implied quantities of interest		
Competitive outcome (point C in Figure I)		$Q = 0.617$, $P = 463.5$
Efficient outcome (point E in Figure I)		$Q = 0.756$, $P = 263.9$
Efficiency cost from selection (triangle CDE)		9.55
Total surplus from efficient allocation (triangle ABE)		283.39
Efficiency cost from mandating contract H (triangle EGH)		29.46

Notes. The table reports the results from our baseline specification. Sample is limited to salaried employees with family coverage. Column (1) of Panel A reports the results from estimating the linear demand $D = \alpha + \beta p$ (equation (11)) on the sample of employees who choose contract H or contract L; D is an indicator variable for whether the employee chose contract H (as opposed to contract L). Column (2) reports the results from estimating the linear cost equation $c = \gamma + \delta p$ (equation (12)) on the sample of individuals who choose contract H; c is the incremental cost to the insurer of covering a given employee's (and covered dependents') medical expenditures with contract H rather than contract L. The price variable (p) is the incremental premium to the employee for contract H (as opposed to contract L). There are no other covariates in the regression besides those shown in the table. All estimates are generated by OLS. Standard errors (in parentheses) allow for an arbitrary variance covariance matrix within each state; p values are in square brackets. Results from alternative specifications are reported in the Online Appendix. Panel B reports the point estimates of several quantities of interest that are derived from the baseline specification and the estimates reported in Panel A.

significant) slope coefficient δ of 0.155. As noted, the slope of the cost curve represents a test for the existence and nature of selection, and the positive coefficient on price indicates the presence of adverse selection. That is, the average cost of individuals who purchased contract H is higher when the price is higher. In other words, when the price selects those who have, on average, higher willingness to pay for contract H, the average costs of this group are also higher. The average cost curve is therefore downward-sloping (in quantity, as in Figure I).

The point estimate from our baseline specification suggests that a dollar increase in the relative price of contract H is associated with an increase in the average cost of the (endogenous) sample selecting contract H at that price of about 16 cents. By itself, this estimate of the cost curve can only provide evidence of the existence of adverse selection. Without knowledge of the demand curve, it does not allow us to form even an approximate guess of the associated efficiency cost of adverse selection. A central theme of this paper is that we can combine the estimates from the demand curve and the cost curve to move beyond detecting selection to quantifying its efficiency cost and, relatedly, to calculating the welfare benefits from a set of public policy interventions.[24]

In this spirit, Figure V shows how to translate the baseline empirical estimates of the demand and cost curves into the theoretical welfare analysis. That is, Figure V presents the empirical analog to Figure I by plotting the estimated demand and average cost curves, as well as the marginal cost curve implied by them (see equation (13)). Based on these estimates, it is straightforward to calculate several quantities of interest (see Panel B of Table III), including the implied welfare cost of competitive pricing, that is, area CDE in Figure V (and Figure I). It should be readily apparent from the figure that, with the cost curve held constant, shifting and/or rotating the demand curve could generate very different welfare costs. This underscores the observation that merely estimating the slope of the cost curve is not by itself informative about the likely magnitude of the resultant inefficiency.

We estimate that the welfare cost associated with competitive pricing is \$9.55 per employee per year, with a 95% confidence interval ranging from \$1 to \$40 per employee.[25] Adverse selection raises the equilibrium price by almost \$200 above the efficient price (compare the estimated efficient price at point E to the estimated equilibrium price at point C), and correspondingly lowers the share of contract H by fourteen percentage points. The social benefit of providing contract H to the marginal employee who buys contract L in equilibrium (i.e., the vertical distance between points C and D in Figure V) is \$138.

24. As noted in Section II.D, when contract L involves partial coverage, welfare analysis will need to account for the (negative) externalities associated with any moral-hazard effects. Our analysis here does not account for such effects because, as we show and discuss in the Online Appendix, we are unable to reject the null of no moral hazard in our specific application.

25. We computed this confidence interval using nonparametric bootstrap. That is, we draw 1,000 bootstrapped samples and repeat our baseline analysis on each sample. The 95% confidence interval is given by the 2.5th and 97.5th percentiles in the distribution of welfare-cost estimates.

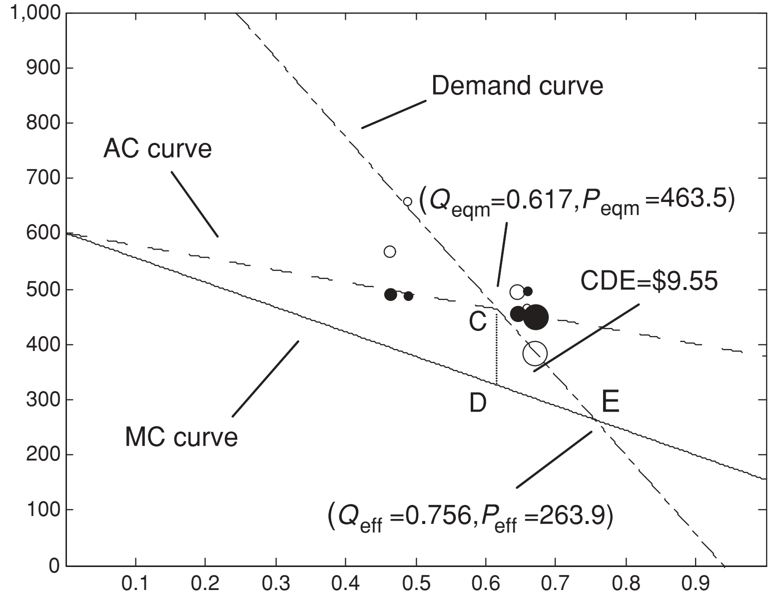

FIGURE V

Efficiency Cost of Adverse Selection—Empirical Analog

This figure is the empirical analog of the theoretical Figure I. The demand curve and AC curve are graphed using the point estimates of our baseline specification (see Table III). The MC curve is implied by the other two curves, as in equation (13). The circles represent the actual data points (see Table II, columns (3) and (4)) for demand (empty circles) and cost (filled circles). The size of each circle is proportional to the number of individuals associated with it. For readability we omit the one data point from Table II with only seven observations (although it is included in the estimation). We label points C, D, and E, which correspond to the theoretical analogs in Figure I, and report some important implied point estimates (of the equilibrium and efficient points, as well as the welfare cost of adverse selection).

Figure V also provides some useful information about the fit of our estimates, and where our pricing variation is relative to the key prices of interest for welfare analysis. The circles superimposed on the figure represent the actual data points (from Table II), with the size of each circle proportional to the number of individuals who faced that price. The fit of the cost curve appears quite good. The fit of the demand curve is also reasonable, although the scatter of data points led us to assess the sensitivity of the results to a concave demand curve, which is one of the exercises reported in the Online Appendix. The price range from \$384 to \$659 in our data brackets our estimate of the equilibrium price (point C) of \$463. The lowest (and modal) price in our sample of \$384 is about 45% higher than our estimate of the efficient price (point E) of $264. Thus, although in principle our approach does not require parametric assumptions, in practice the span of the pricing variation in our particular application requires that we impose some functional form assumptions to estimate the area of triangle CDE. In the Online Appendix we examine alternative functional forms.

IV.E. Welfare Analyses

We show how our framework can be used to produce a number of other welfare estimates. These may be of interest in their own right and also serve as useful comparisons for our baseline estimate of the welfare cost of inefficient pricing arising from adverse selection (triangle CDE).

Benchmarks for Our Welfare Cost Estimates. We can use the demand and cost curves shown in Figure V to calculate various benchmarks that provide some context for our estimate of the welfare cost of competitive pricing of $9.55 per employee. An important consideration in choosing a benchmark is how far out of sample we must take the demand and cost estimates in order to form it. Again, Figure V is informative on this point.

We calculate two useful denominators to scale our estimate of the welfare cost. One is a measure of how large this cost could have been before we started the analysis. Our thought experiment is to assume that we observe data (on price, quantity, and costs) from only one of the rows of Table II, so there is no price variation. We assume we observe the weighted average price of $414. Because individuals have the option to buy contract H at this price but choose not to do so, their welfare loss from not being covered by contract H cannot exceed $414. Our estimate of the efficiency cost of $9.55 is therefore 2.3% of this "maximum money at stake," as Einav, Finkelstein, and Schrimpf (2010) term this construct.

A second useful denominator is to scale the welfare cost from competitive pricing arising from adverse selection by the total surplus at stake from efficient pricing. We therefore calculate the ratio of triangle CDE (the welfare loss from competitive pricing) to triangle ABE (the total welfare from efficient pricing) in Figure I. To enhance readability, points A and B are not shown in Figure V, but are easily calculated from the parameter estimates. They are, however, fairly far out of sample relative to our data. For example, at point A we estimate the price to be about $1,350, which is more than twice the highest price we observe in the data. In our particular application therefore, this benchmark raises concerns about extrapolating too far out of sample, although we show in the Online Appendix that the result is relatively robust to alternative functional forms for that extrapolation. Using this benchmark as a denominator, we estimate that the welfare loss from adverse selection is about 3% of the surplus at stake from efficient pricing.

Welfare under Other Market Allocations. Although our welfare analysis has focused on the efficiency cost of competitive equilibrium pricing arising from adverse selection, the fact that we observe prices varying—and this is how we identify the demand and cost curves—underscores the point that to generate our pricing variation we observed a market that is not in equilibrium. Our analysis of "equilibrium" pricing, like our analysis of "efficient" pricing, is based on a counterfactual. By the same token, our analysis of the efficiency cost of such pricing is not an analysis of the realized efficiency cost for our sample but rather what this efficiency cost would be if, contrary to fact, these options were offered in a competitive market setting. Because our demand and cost curves are sufficient statistics for welfare analysis of the pricing of existing contracts, we can use them to compute the welfare cost of any other inefficient pricing. For example, we estimate that the weighted average of the welfare cost of adverse selection given the observed pricing in our sample (see Table II, columns (1) and (2)) is \$6.26 per employee per year.

Moreover, as we noted in Section II, we could also use the estimated demand and cost curves to estimate welfare under alternative assumptions about the market equilibrium, including monopoly or imperfect competition. For example, a monopolist facing our estimated demand and cost curves would set a (relative) price of \$907 for contract H. The resultant efficiency cost would be just below \$100 per employee, which is an order of magnitude higher than the estimated efficiency cost from competitive pricing.

Another interesting alternative is to compute what the welfare cost of competitive pricing would be if, contrary to what happens in the employment context, competitive prices were set based on some observable characteristics of the employees. To do so, we simply estimate the demand and cost curves separately for each "cell" of individuals who, based on their characteristics, would be offered the same price. As an example, we consider what would happen to our welfare estimate if prices were set differently based on whether the family coverage applied to three individuals, four individuals, or five or more individuals. About half of our baseline sample has four covered members, and the remaining sample is evenly split between the other two categories. We maintain the assumption that the equilibrium would involve average-cost pricing, although now the equilibrium is determined separately in each of the three market segments. We detect adverse selection in each segment separately, and estimate that the (weighted average) welfare cost of this selection would be \$12.92 if prices were set differently for each market segment, compared to our estimated welfare cost of \$9.55 when family size is not priced.

Welfare Consequences of Government Intervention. Adverse selection provides the textbook economic rationale for government intervention in insurance markets. We therefore show how we can use our framework to estimate the welfare cost of standard public policy interventions in insurance markets. We then compare this to our estimate of the welfare cost of competitive pricing. As mentioned, our approach allows us to analyze the welfare consequences of counterfactual public policies that change the price of existing contracts, such as price subsidies, coverage mandates, and regulation of the characteristics of individuals that can be used in pricing. This last potential policy was already discussed in the preceding section where we analyzed the welfare consequences of firms pricing on a characteristic (in our example, family size) that is not currently priced.

Our preferred policy analysis in our particular application is to compare the social welfare gain from efficient pricing (triangle CDE) to the social welfare cost of the price subsidy required to achieve this efficient price. An attraction of this calculation is that it does not require further out-of-sample extrapolation beyond what is needed to compute the area of triangle CDE itself. The social cost of such a subsidy is given by $\lambda(P_{\text{eqm}} - P_{\text{eff}})Q_{\text{eff}}$, where λ is the marginal cost of public funds. Given our estimates of the efficient and equilibrium outcomes (Figure V), and using 0.3 as the (standard estimate of) marginal cost of public funds (e.g., Poterba [1996]), we calculate the social cost of the price subsidy needed to achieve the efficient allocation to be \$45. That is, we estimate that the social cost of a price subsidy that achieves the efficient allocation is about five time larger than the social welfare (of \$9.55) it gains.

Of course, given a nonzero social cost of public funds, the welfare-maximizing subsidy would not attempt to achieve the efficient allocation. It is therefore also interesting to investigate whether there is any scope for welfare-improving government intervention in the form of a price subsidy to contract H. We do this by investigating whether, at the competitive allocation (point C), a marginal dollar of subsidy is welfare-enhancing. We calculate that in our application it is not, so that the welfare maximizing (additional) price subsidy by the government is therefore zero.[26]

We also compared welfare in the competitive equilibrium with adverse selection to welfare when everyone is mandated to be covered by contract H . Mandatory insurance is the canonical solution to the problem of adverse selection in insurance markets (Akerlof 1970), making the analysis of the mandate of considerable interest.[27] However, in our application, the welfare cost of mandating coverage by contract H (area EGH in Figure I) requires calculating points that are reasonably far out of sample. This suggests that in our particular application more caution is warranted with this analysis (although again we show in the Online Appendix that the estimate is reasonably robust). With this important caveat in mind, we estimate that the welfare cost from mandatory coverage by contract H is about three times higher than the welfare cost associated with competitive pricing.

V. Conclusions

This paper proposes a simple approach to quantifying and estimating the welfare cost caused by inefficient pricing in insurance markets with selection. We show how standard consumer and producer theory can be applied to welfare analysis of such markets, and we provide a graphical representation of the efficiency cost of competitive pricing. This graphical analysis not only provides helpful intuition but also suggests a straightforward empirical approach to welfare analysis. The key to estimation is the existence of identifying variation in the price of insurance. Applied welfare analysis usually requires pricing variation that allows

26. The marginal benefit from the first dollar of subsidy is \$137.4 (the distance between point C and point E) times the marginal number of newly covered individuals (0.0007 given our estimates of the demand curve). By contrast, the marginal cost of the dollar subsidy is the cost of public funds (0.3) times all of the inframarginal individuals at point C (i.e., 0.617).

27. Footnote 5 discussed some of the possible factors that may make it inefficient to allocate the H contract to the entire market.

the researcher to trace out a demand curve. The defining feature of selection markets is that costs vary endogenously as market participants respond to price. Welfare analysis in such markets therefore requires that we also trace out the (endogenous) cost curve. We show that this is straightforward to do using direct data on cost and the same price variation used to identify demand. In doing so, the slope of the estimated cost curve also provides a direct test of the existence and nature of selection.

We illustrated our framework by applying it in the context of employer-provided health insurance at a particular firm. We find evidence of adverse selection in the market, but we estimate that the welfare cost of the resultant inefficient pricing is quantitatively small. It is important to emphasize that our empirical estimates are specific to our particular setting and there is no reason to think that our welfare estimates are representative of other populations, other institutional environments, or other insurance markets. However, at a broad level, our findings illustrate that it is empirically possible to find markets in which adverse selection exists and impairs market efficiency, but where the efficiency cost of the pricing it produces may not be large, or obviously remediable using standard public policy tools. Whether the same is true in other markets, and in which, is an important area for future work.

We hope that such future work will apply our framework and strategy to other insurance settings (or, more generally, to other settings with hidden information, such as credit markets or regulated monopolists). The approach is relatively straightforward to implement and fairly general. As a result, comparisons of welfare estimates obtained by this approach across different settings may be informative. In any given application, we see the transparency of our approach as one of its key attractions. The direct mapping from the theoretical framework (Figure I) to its empirical analog (Figure V) facilitates an informed appraisal of the estimates, including such issues as in-sample fit, the extent of out-of-sample extrapolation needed for a particular welfare estimate, and the extent and validity of the pricing variation.

As we emphasize throughout, our approach is unable to shed light on the welfare consequences of any distortion in the contract space induced by selection, or of public policies that introduce contracts not observed in the data. Analysis of such questions would require a model of the primitives underlying the revealed demand and cost curves. We view such models as a useful and important complement to the empirical approach we have proposed here.

Stanford University and National Bureau of Economic Research
Massachusetts Institute of Technology and National Bureau of Economic Research
Stanford University

References

Abbring, Jaap, Pierre-Andre Chiappori, and Jean Pinquet, "Moral Hazard and Dynamic Insurance Data," *Journal of the European Economic Association*, 1 (2003), 767–820.

Adams, William, Liran Einav, and Jonathan Levin, "Liquidity Constraints and Imperfect Information in Subprime Lending," *American Economic Review*, 99 (2009), 49–84.

Akerlof, George, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84 (1970), 488–500.

Barseghyan, Levon, Jeffrey Prince, and Joshua Teitelbaum, "Are Risk Preferences Stable across Contexts? Evidence from Insurance Data," *American Economic Review*, forthcoming, 2010.

Blackmon, Glenn B., and Richard Zeckhauser, "Mispriced Equity: Regulated Rates for Automobile Insurance in Massachusetts," *American Economic Review Papers and Proceedings,* 81 (1991), 65–69.

Buchmueller, Thomas C., and John E. DiNardo, "Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut," *American Economic Review*, 91 (2002), 280–294.

Bundorf, Kate M., Jonathan Levin, and Neale Mahoney, " Pricing and Welfare in Health Plan Choice," Mimeo, Stanford University, 2008.

Bundorf, Kate M., and Kosali Simon, "The Impact of Rate Regulation on Access to Supplemental Health Insurance," *American Economic Review Papers and Proceedings*, 96 (2006), 67–71.

Busch, Susan H., Colleen L. Barry, Sally J. Vegso, Jody L. Sindelar, and Mark R. Cullen, "Effects of a Cost-Sharing Exemption on Use of Preventive Services at One Large Employer," *Health Affairs*, 25 (2006), 1529–1536.

Carlin, Caroline, and Robert J. Town, "Adverse Selection, Welfare and Optimal Pricing of Employer-Sponsored Health Plans," Mimeo, University of Minnesota, 2009.

Chetty, Raj, "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 116 (2008), 173–234.

———, "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced Form Estimates," *Annual Review of Economics*, 1 (2009), 451–488.

Chetty, Raj, and Emmanuel Saez, "Optimal Taxation and Social Insurance with Endogenous Private Insurance," *American Economic Journal: Economic Policy*, 2 (2010), 85–114.

Chiappori, Pierre-Andre, and Bernard Salanie, "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108 (2000), 56–78.

Cohen, Alma, and Liran Einav, "Estimating Risk Preferences from Deductible Choice," *American Economic Review* , 97 (2007), 745–788.

Cutler, David M., Amy Finkelstein, and Kathleen McGarry, " Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance," *American Economic Review Papers and Proceedings*, 98 (2008), 157–162.

Cutler, David M., and Sarah J. Reber, "Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection," *Quarterly Journal of Economics*, 113 (1998), 433–466.

Cutler, David M., and Richard J. Zeckhauser, "The Anatomy of Health Insurance," in *Handbook of Health Economics*, A. J. Culyer and J. P. Newhouse, eds., Volume 1, 563–643 (Amsterdam: Elsevier, 2000).

De Meza, David, and David C. Webb, "Advantageous Selection in Insurance Markets," *Rand Journal of Economics*, 32 (2001), 249–262.
Eichner, Matthew J., Mark B. McLellan, and David A. Wise, " Insurance or Self-Insurance? Variation, Persistence, and Individual Health Accounts," in *Inquiries in the Economics of Aging*, David Wise, ed. (Chicago: University of Chicago Press, 1998).
Einav, Liran, Amy Finkelstein, and Jonathan Levin, "Beyond Testing: Empirical Models of Insurance Markets," *Annual Review of Economics*, 2 (2010), 311–336.
Einav, Liran, Amy Finkelstein, and Paul Schrimpf, "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," *Econometrica*, 78 (2010), 1031–1092.
Finkelstein, Amy, and Kathleen McGarry, "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *American Economic Review*, 96 (2006), 938–958.
Finkelstein, Amy, and James Poterba, "Selection Effects in the Market for Individual Annuities: New Evidence from the United Kingdom," *Economic Journal*, 112 (2002), 28–50.
———, "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market," *Journal of Political Economy*, 112 (2004), 193–208.
Fox, Peter, Thomas Rice, and Lisa Alecxih, "Medigap Regulation: Lessons for Health Care Reform," *Journal of Health Politics, Policy and Law*, 20 (1995), 31–48.
Greenwald, Bruce C., and Joseph E. Stiglitz, "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics*, 101 (1986), 229–264.
Gruber, Jonathan, "The Incidence of Mandated Maternity Benefits," *American Economic Review*, 84 (1994), 622–641.
———, "Taxes and Health Insurance," in *Tax Policy and the Economy*, James Poterba, ed., Volume 16, 37–16 (Cambridge, MA: MIT Press, 2002).
Israel, Mark, Do We Drive More Safely When Accidents Are More Expensive? Identifying Moral Hazard from Experience Rating Schemes, Unpublished Mimeo (http://www.wcas.northwestern.edu/csio/Conferences/CSIO-IDEI-2004/moral_exp_paper1.pdf, 2004).
Karlan, Dean, and Jonathan Zinman, "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," *Econometrica*, 77 (2009), 1993–2008.
Keating, Bryan, "Switching in an Evolving Industry," Ph.D. Dissertation, Stanford University, Chapter 1, 2007.
Lustig, Joshua D., "Measuring Welfare Losses from Adverse Selection and Imperfect Competition in Privatized Medicare," Mimeo, Boston University, 2009.
Poterba, James, "Government Intervention in the Markets for Education and Health Care: How and Why?" in *Individual and Social Responsibility*, Victor Fuchs, ed. (Chicago: University of Chicago Press, 1996).
Rothschild, Michael, and Joseph E. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90 (1976), 630–649.
Stiglitz, Joseph E., "The Causes and Consequences of the Dependence of Quality on Price," *Journal of Economics Literature*, 25 (1987), 1–48.
Sydnor, Justin, "(Over) Insuring Modest Risks," *American Economic Journal: Applied Economics*, forthcoming, 2010.