Estimating welfare in insurance markets using variation in prices*

Liran Einav, Amy Finkelstein, and Mark R. Cullen†

October 2008

Abstract. We show how standard consumer and producer theory can be used to estimate welfare in insurance markets with selection. The key observation is that the same price variation needed to identify the demand curve also identifies how costs vary as market participants endogenously respond to price. With estimates of both the demand and cost curves, welfare analysis is straightforward. We illustrate our approach by applying it to the employee health insurance choices at Alcoa, Inc. We detect adverse selection in this setting but estimate that its quantitative welfare implications are small, and not obviously remediable by standard public policy tools.

JEL classification numbers: C13, C51, D14, D60, D82, I11.

Keywords: Asymmetric information; adverse selection; health insurance; efficiency cost.

*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 Kate Bundorf, Raj Chetty, Peter Diamond, Hanming Fang, David Laibson, Jonathan Levin, Jim Poterba, Jonathan Skinner, and seminar participants at MIT, Stanford, Yale, the NBER Health Care Meeting, and SITE 2008 for helpful comments. The data were provided as part of an ongoing service and research agreement between Alcoa, Inc. and Yale, under which Yale faculty and staff perform jointly agreed-upon ongoing and ad-hoc research projects on workers' health, injury, disability and health care, and Mark Cullen serves as medical director for Alcoa, Inc. We gratefully acknowledge support from the National Science Foundation grant #SES-0643037 (Einav), the Alfred P. Sloan Foundation (Finkelstein), and the John D. and Catherine T. MacArthur Foundation Network on Socioeconomic Status and Health, and Alcoa, Inc. (Cullen).

†Einav: Department of Economics, Stanford University, and NBER, leinav@stanford.edu; Finkelstein: Department of Economics, MIT, and NBER, afink@mit.edu; Cullen: Occupational and Environmental Medicine Program, Yale School of Medicine, mark.cullen@yale.edu.

1 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 alternative 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, 2007) and health insurance (Carlin and Town, 2007; Lustig, 2007; Bundorf, Levin, and Mahoney, 2008). 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 type and preferences. This allows for 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 non-trivial effects on the welfare estimates. Moreover, they are often specific to the particular market studied, making it difficult to meaningfully compare welfare estimates across different insurance markets. Technical estimation challenges further impairs the ability of researchers to readily adapt these approaches to other insurance market, or even to other data sets in the same market.

This paper therefore develops an alternative approach to empirical welfare analysis in insurance markets. We show how standard consumer and producer theory – familiar to any student of intermediate micro – can be applied to empirical 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 market costs are endogenous to price. Empirical welfare analysis therefore requires not only the usual estimation of how demand varies with price, but also estimation of how the costs of insuring the (endogenous) market participants vary with price.

This suggests a straightforward empirical approach to welfare analysis of selection in insurance markets. 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 trace out how costs vary as the set of market participants endogenously changes. With these two curves in hand, welfare analysis of the inefficiency caused by selection – or of the consequences of a range of alternative potential public policy interventions – is simple and familiar.

Our approach has several appealing features. First, it does not require the researcher to make assumptions about consumer preferences or the nature of their ex ante information about their ex post risk. 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 quite

similar in spirit to Chetty (2008) who shows how key ex-post behavioral elasticities are sufficient statistics for welfare analysis of the optimal level of unemployment insurance benefits.

Second, our approach is relatively straightforward to implement, and likely to be widely applicable. In particular, while cost data are often quite difficult to obtain in many product markets (so that estimation of the cost curve is often not feasible), this is less likely to be a problem in insurance markets. Cost data tend to be much easier to obtain in insurance markets since 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 of the exogenous 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. This suggests that it may be informative to compare estimates of the welfare cost of adverse selection obtained by this approach in different contexts, such as different populations, or different insurance markets.

The chief limitation of our approach is that counterfactual welfare analysis is limited to changes to the prices of existing contracts (for example, through mandates or price subsidies). Analysis of counterfactuals that would introduce different products than those observed in the data is not feasible. Such analysis requires estimation of the structural primitives underlying the demand and cost curves in the insurance market, as the recent papers mentioned above have done.

Given these trade-offs, we see our approach as highly complementary to – rather than competitive with – these earlier papers. The trade-off is a familiar one in economics. It is somewhat analogous to the trade-offs in demand estimation between product-space approaches (e.g. the Almost Ideal Demand System of Deaton and Muelbauer, 1980; see, e.g., Hausman (1997) for an application) and characteristic-space approaches (Lancaster, 1966; see, e.g., Berry, Levinsohn, and Pakes (1995) for an application). The latter can evaluate welfare from new goods, while the former can only do this after these goods have been introduced.

Finally, we note that beyond estimating welfare, an additional considerable appeal of our proposed approach is that the shape of the estimated cost curve provides a direct test of the existence and nature of selection. 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, while 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). This is quite important, since 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 important and widely used "bivariate probit" (a.k.a. "positive correlation") test of Chiappori and Salanie (2000) which jointly tests for the existence of either adverse selection or moral hazard (but not for each separately). The improvement comes at the cost of an additional data requirement, namely pricing variation that is exogenous to individual demand and insurer's costs.

The rest of the paper is divided into two main parts: framework and application. Section 2

describes our framework and provides some graphical intuition for the efficiency costs of selection in insurance markets. Section 3 shows how the framework translates naturally into a series of estimable equations, and discusses the data requirements.

Section 4 illustrates our approach by applying it to the market for employer-provided health insurance in the United States. This is a market of substantial interest in its own right. The workplace is the primary source of private health insurance in the United States, covering about 90 percent of the privately insured non-elderly, or about 160 million Americans (Fronstin, 2003). Government intervention in health insurance markets is widespread but also considerably varied in its choice of instrument, which includes both subsidies for private insurance purchases and mandatory coverage by a single public insurance contract. A standard economic rationale for these various programs is as a counterweight to adverse selection pressures in private health insurance markets.

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) is a notable exception. Like us, they analyze the welfare cost of adverse selection in the setting of employer-provided health insurance. A key distinction, however, is that while they estimate the demand curve, they do not estimate the cost curve, which is crucial for welfare analysis. We outline an approach for estimating the cost curve and implement it in our application.

We analyze individual-level data from Alcoa, Inc., a large multinational private producer of aluminum and related products. We observe the health insurance options, choices, and medical expenditures 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 face different employee premiums for purchasing more comprehensive relative to less comprehensive insurance. We verify that pricing appears random with respect to the characteristics of the employees that the managers setting employee premiums can likely observe.

Using this price variation, we estimate that marginal cost is increasing in price, and thus detect adverse selection in this market. We estimate that in a competitive market the annual efficiency cost of this selection would be about $10 per person, or about 3 percent of the total surplus at stake from efficient pricing. Our findings also suggest that there is limited scope for standard policy instruments to produce welfare gains over the equilibrium outcome. For example, we 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 important to emphasize that there is no general lesson in our empirical findings for the welfare consequences of government intervention in other insurance contexts. Our estimates are specific to our population and to the particular health insurance choices they face. Nonetheless, at a broad level, they highlight the importance of moving beyond detection of market failures to quantifying their welfare implications, and the welfare achievable under potential public policy

interventions. 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. We conclude the paper by discussing a wide range of settings in which the approach we propose could be possibly applied. We view this as a promising direction for further work.

2 Theoretical framework

2.1 Model

Setup and notation We consider a situation in which a given population of individuals is allowed to choose from exactly two available insurance contracts, one that offers high coverage (denoted by H) and one that offers less coverage (denoted by 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 illustrations. 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 we relax them 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. 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 need to specify the nature of ζ; it could describe multi-dimensional risk factors, consumers' ex ante information about their ex post risk, 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 coverages 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 $c(\zeta_i)$.[1]

Demand for insurance We assume that each individual makes a discrete choice of whether to buy insurance or not. Since 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 contingent on observed characteristics, one should think of our analysis as applied to a set of individuals that only vary in unobserved (or unpriced) characteristics. We assume that if individuals

[1] These costs may potentially depend on the coverage the individual chooses (i.e., there may be moral hazard effects). As we discuss in more detail in Section 2.4 below, this does not affect the analysis.

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 $v^H(\zeta_i, p) \geq v^L(\zeta_i)$. Define $\pi(\zeta_i) \equiv \max\left\{p : v^H(\zeta_i, p) \geq v^L(\zeta_i)\right\}$. That is, $\pi(\zeta_i)$ is the highest price of insurance at which individual i is willing to buy insurance. Aggregate demand for insurance is therefore given by

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

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 (a-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 supply side 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 $c(\zeta_i)$, although this assumption is straightforward to relax. The foregoing assumptions imply that the average (expected) cost curve in the market is given by

$$AC(p) = \frac{1}{D(p)} \int c(\zeta) 1\left(\pi(\zeta) \geq p\right) dG(\zeta) = E\left(c(\zeta) | \pi(\zeta) \geq p\right). \tag{2}$$

Note that the average cost curve is determined by the costs of the sample of individuals who endogenously choose H. The marginal (expected) cost curve[2] in the market is given by

$$MC(p) = E\left(c(\zeta) | \pi(\zeta) = p\right). \tag{3}$$

In order to straightforwardly characterize equilibrium, we make two further simplifying assumptions. First, we assume that there exists $\overline{p}$ such that $D(\overline{p}) > 0$ and $MC(p) < p$ 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.[3] Second, we assume that if there exists $\underline{p}$ such that $MC(\underline{p}) > \underline{p}$ then $MC(p) > p$ for all $p < \underline{p}$. That is, we assume that $MC(p)$ crosses the demand curve

[2] 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.

[3] 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.

at most once.[4] It is easy to verify that these assumptions guarantee the existence and uniqueness of equilibrium.[5] In particular, the equilibrium is characterized by the lowest price that implies zero profits, that is:

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

2.2 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. We perform our welfare analysis in partial equilibrium; we ignore any income effects associated with price changes.[6]

Denote by $ce^H(\zeta_i)$ and $ce^L(\zeta_i)$ the certainty equivalent of consumer i from 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 $\pi(\zeta_i) = ce^H(\zeta_i) - ce^L(\zeta_i) > 0$. We can write consumer welfare as

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

and producer welfare as

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

Total welfare will then be given by

$$TS = CS + PS = \int \left[\left(ce^H(\zeta) - c(\zeta) \right) 1\left(\pi(\zeta) \geq p\right) + ce^L(\zeta) 1\left(\pi(\zeta) < p\right) \right] dG(\zeta). \tag{7}$$

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 individual i.

[4] In the most basic economic framework of insurance the difference between $\pi(p)$ and $MC(p)$ is the risk premium and is non-negative if all individual are risk averse, implying that $MC(p)$ will never cross the demand curve. In practice, however, there are many reasons for such crossing. Those include, among others, 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.

[5] This is a similar result to the "buyers' equilibrium" in the (richer and more complex) setting analyzed by Wilson (1980).

[6] For standard consumer goods, this amounts to assuming that utility is quasi-linear in all other goods. In a textbook insurance context, the assumption that the income effects associated with changes in the premium do not change the willingness to pay for insurance amounts to assuming that the utility function exhibits constant absolute risk aversion (CARA), or that CARA is a reasonable approximation when the premium changes are small relative to the individual's income, as in the choice we study in our empirical application below.

In many contexts (including our application below), the only instrument available to affect the insurance allocation is through the price. 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 2).

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 insurance if and only if $\pi(\zeta_i)$ is at least as great as the expected social cost of providing the insurance to all individuals with willingness to pay $\pi(\zeta_i)$. That is, it is constrained efficient for individual i to purchase insurance if and only if

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

We use this constrained efficient benchmark throughout the paper, and hereafter refer to it simply as the efficient allocation.[7]

2.3 Graphical illustration

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

Adverse selection Figure 1 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 the H contract. Likewise, the average cost (AC) curve and marginal cost (MC) curve denote the average and marginal incremental costs to the insurer from coverage with the H contract relative to coverage with the L contract.

The key feature of adverse selection is that the individuals who value insurance the most (i.e., have the highest willingness to pay) are those who, on average, have the highest expected costs. This is equivalent in Figure 1 to a declining MC curve (i.e., that marginal cost is increasing in price and decreasing in quantity); as the price falls, individuals with lower willingness to pay select contract H, and bring down average costs. The essence of the private information problem is that firms cannot charge individuals based on their (privately known) marginal cost, but are instead restricted to charging a uniform price, which in equilibrium implies average cost pricing. Since average costs are always higher than marginal costs, adverse selection creates under-insurance, a familiar result first pointed out by Akerlof (1970). This under-insurance is illustrated in Figure 1. The equilibrium share of individuals who buy contract H is Q_{eqm} (where the AC curve intersects

[7] 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.

the demand curve), while the efficient number of insurance buyers (where the MC curve intersects the demand curve) is $Q_{eff} > Q_{eqm}$.

The welfare loss due to adverse selection is represented by the shaded region CDE in Figure 1; 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 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 L (normalized to zero), or the welfare effect of policies that subsidize (or tax) the equilibrium price. The relative welfare rankings of these alternatives is 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 under-insurance (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.

Our framework makes it easy to describe the nature and consequences of advantageous selection. Figure 2 provides an illustration. In contrast to adverse selection, with advantageous selection the individuals who value insurance the most are those who have the least 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 in equilibrium price is based on average cost. However, with advantageous selection the resultant market failure is one of over-insurance rather than under-insurance (i.e., $Q_{eff} < Q_{eqm}$ in Figure 2), as has been pointed out by de Meza and Webb (2001), among others. Intuitively, insurance providers have an additional incentive to reduce price, as the infra-marginal 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 2 including mandating insurance contract H (the area ABE minus the area EGH), mandating insurance 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 non-equilibrium pricing of the existing contracts. In other words, different underlying structures (i.e., vectors of preferences and private information as summarized by ζ) have the same welfare implications if they generate the same demand and cost curves.[8] This in turn is the essence of our empirical approach. We estimate the demand and cost curves, but remain agnostic about the underlying preferences that determine the demand curve and the underlying nature of the individuals' behavior that gives rise to the cost curve. As long as individuals' revealed choices can be used for welfare analysis, the precise *source* of the selection (i.e., the ζ) is not germane for analyzing the efficiency consequences of the resultant selection, or the welfare consequences of public policies that change the equilibrium price (e.g., by mandating or subsidizing a particular policy).[9] By the same token, the precise *source* of the cost curve – such as any effect that moral hazard plays in determining costs – is not germane for analyzing the efficiency consequences of selection that occurs as a result of the given cost curve.

Likewise, the demand and cost curves are also 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 assumption of perfect competition. It also 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 assumption of no additional costs beyond insurable claims.

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 can be used to analyze the effect of a wide variety of standard government interventions in insurance markets which 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, or 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.

2.4 Comment: moral hazard

Incorporating moral hazard Thus far we have not explicitly discussed any potential moral hazard effect of insurance. This is because moral hazard does not change the analysis, but slightly complicates the presentation. We illustrate this by returning to the original framework in which

[8] Note that we have placed no restrictions in Figures 1 or 2 on the nature of the underlying consumer characteristics 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 different unobserved characteristics.

[9] Needless to say, the source of selection – for example, whether selection is driven by unobserved preferences for insurance such as risk aversion or by heterogeneity among individuals as to how much they know about their risks – may be of independent interest; for example, it would be of interest for counterfactuals that stipulate changing the information structure.

we defined H to be full coverage and L to be no coverage, but an analogous extension applies to any of the variants discussed above.[10]

With moral hazard, the expected insurable cost for individual i is now a function of his insurance coverage because his insurance coverage may affect behavior. We therefore define two (rather than one) expected monetary costs for individual i; let $c^H(\zeta_i)$ and $c^L(\zeta_i)$ be individual i's expected insurable costs when he has full and no coverage, respectively. We assume throughout that $c^H(\zeta_i) \geq c^L(\zeta_i)$; if moral hazard exists this inequality will be strict, while without moral hazard $c^H(\zeta_i) = c^L(\zeta_i)$. 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 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 primary welfare analysis of selection therefore takes any moral hazard effect as given. We investigate the welfare cost of 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.

In order 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 choose H (see equation (2)); in the new notation their costs are given by $c^H(\zeta_i)$ since they are covered by the H contract (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. What uninsured individuals do has no implications to insurers. 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 (π) 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 to a part that arises from reduction in risk and a part that arises from a change in behavior.

Welfare analysis with behavior-contingent insurance contracts As a brief digression, it might be interesting to consider the welfare cost of adverse selection when moral hazard is not taken as given. In other words, how does the welfare cost of adverse selection change if the insurer (or the social planner) could provide insurance that is contingent on behavior? While largely irrelevant from a policy perspective – as noted the social planner is unlikely to have a comparative advantage in ameliorating moral hazard – this type of conceptual exercise may shed some insight on how the welfare cost of adverse selection might change as technological progress allows insurers to write

[10] We note that if the lower coverage contract includes some coverage, then the market equilibrium should be thought of as one in which firms offering H only compete on the incremental coverage.

contracts that are increasingly contingent on (previously unobservable) behavior.[11]

We consider the (counterfactual) case in which the insurer provides the H contract, but can force consumers to not change their behavior (that is, to behave as if they are under L). Our framework allows us to bound the demand curve for such a contract. To see this, denote by $\pi(\zeta_i, L)$ the willingness to pay for a High coverage contract which is contingent on L-like behavior. A simple revealed preference argument implies that

$$\pi(\zeta_i, L) \in \left[\pi(\zeta_i) - \left(c^H(\zeta_i) - c^L(\zeta_i)\right), \pi(\zeta_i)\right]. \tag{10}$$

The upper bound is clear: willingness to pay for a constrained contract cannot be higher than for an unconstrained contract. The lower bound is driven by the observation that if individuals change their behavior from $c^L(\zeta_i)$ to $c^H(\zeta_i)$ only in response to the change in coverage, then it must be that they value this change in behavior by less than the cost of the change. Together with the relevant cost curves for such a contract (AC^L and MC^L), these bounds on the demand curve can provide bounds on the welfare costs of adverse selection under such a contract. Figure 3 presents one special case of this exercise, where the moral hazard effect is homogeneous. In this case, triangle CMN provides an upper bound of the welfare cost of adverse selection, and the area of the lower bound is the same as that of triangle CDE, since all 3 curves are just shifted down in parallel.

3 Estimation

Applying our framework to estimating welfare in an insurance market requires data that allows estimation of the demand curve $D(p)$ and the average cost curve $AC(p)$. The marginal cost curve can be directly backed out from these two curves and does not require further estimation. To see this, note that

$$MC(p) = \frac{\partial TC(p)}{\partial D(p)} = \frac{\partial \left(AC(p) \cdot D(p)\right)}{\partial D(p)} = \left(\frac{\partial D(p)}{\partial p}\right)^{-1} \frac{\partial \left(AC(p) \cdot D(p)\right)}{\partial p}. \tag{11}$$

With these three curves – $D(p)$, $AC(p)$, and $MC(p)$ – in hand, we can straightforwardly compute welfare under various allocations.

As is standard, estimating the demand curve requires data on prices and quantities (i.e., insurance coverage), and price variation that is exogenous to demand which can be used to trace out the demand curve. To estimate the $AC(p)$ 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 very same variation in prices to trace out the $AC(p)$ 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.

[11] For example, there are new in-car devices that allow auto insurance companies to monitor driving behavior, so that in principle it is possible for contracts to now be written contingent on this behavior.

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 selection, then price variation that spans the range between the market equilibrium price (point C in Figures 1 and 2) and the efficient price (point E) allows us to estimate the welfare cost of selection (area CDE) non-parametrically (that is, without any functional form assumptions regarding 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 functional form assumptions.

It is also worthwhile to observe that – although this is not the focus of our paper – we could make some progress toward bounding the efficiency cost of selection with fewer data requirements. We use Figure 1 (adverse selection) for this discussion (it is easy to imagine an analogous discussion which uses Figure 2). Suppose we observe only the relative price of insurance. If we are willing to assume that the price we observe is the competitive equilibrium price P_{eqm}, we can obtain a (presumably not very tight) upper bound of the welfare cost of selection, given by $P_{eqm}Q_{max}$ (rectangle IJKO in Figure 1).[12] If we also observe the market share of contract L, denoted $(Q_{max} - Q_{eqm})$, this upper bound can be tightened to $P_{eqm}(Q_{max} - Q_{eqm})$ (rectangle CJKL in Figure 1). Finally, if we also have data on the average insurable costs of the individuals choosing contract L, denoted AC^L, we can further tighten up the upper bound to be $\left(P_{eqm} - AC^L\right)(Q_{max} - Q_{eqm})$ (equal to area CJGD in Figure 1).[13] Anything tighter will probably require price variation, which provides more information about the marginal cost and marginal willingness-to-pay for individuals currently not covered by H.

An example We illustrate the spirit of our empirical approach with a simple example. Consider a population of individuals making a binary choice of whether to fully insure or not to insure at all. Each individual is characterized by two parameters: his willingness to pay for insurance π and his expected costs to the insurer c.[14] Suppose individuals are uniformly drawn from a discrete distribution of three types, such that

$$(\pi, c) \in \{(2,1),(4,3),(6,5)\}. \tag{12}$$

Note that these types exhibit adverse selection in the sense that individuals who value insurance more (i.e., higher π) are expected to cost more to the insurance company (i.e., higher c). The competitive (i.e., zero profit) equilibrium price would be $p = 4$, at which price is equal to average costs. Because $\pi > c$ for all types, an efficient allocation requires that everyone purchases insurance, or that $p \leq 2$. Thus, we have the well-known result that adverse selection results in under-provision of insurance.

[12] This upper bound is what we used in Einav, Finkelstein, and Schrimpf (2007) to define the Maximum Money at Stake (MMS) concept, as a way to quantify the relevant size of an insurance market.

[13] To see this, note that $P_{eqm}(Q_{max} - Q_{eqm})$ is equal to area CJKL, while $AC^L(Q_{max} - Q_{eqm})$ is equal to area DGKL because AC^L is the average value of the MC curve between Q_{eqm} and Q_{max}.

[14] Characterizing individuals using these two dimensions of willingness-to-pay and expected costs is similar to the framework proposed by Feldman and Dowd (1982) and more recently used by Bundorf, Levin, and Mahoney (2008).

Of course, the econometrician does not directly observe an individual's willingness to pay π, or his individual-specific cost c. However, these can be recovered, and welfare analysis performed, if there exist data on the fraction insured and the average costs of the insured at different (exogenously generated) insurance prices.

For example, consider data on insurance coverage and costs for three different prices of $p = 2, 4, 6$. Given the assumptions above, the data available to the econometrician would consist of

$$(p, Q, AC) = \left\{ (2, 1, 3), \left(4, \frac{2}{3}, 4\right), \left(6, \frac{1}{3}, 5\right) \right\}, \tag{13}$$

where AC is the average costs of the insured. For example, the case of $p = 4$ will result in insurer share of $\frac{2}{3}$ (individuals with $\pi = 4$ or $\pi = 6$ will purchase, but individuals with $\pi = 2$ will not), and average costs of those who purchase insurance of $\frac{3+5}{2} = 4$. Similarly, the case of $p = 2$ will result in insurer's market share of 1 and average costs of 3, and the case of $p = 6$ will result in insurer share of $\frac{1}{3}$ and average costs of 5.

Using these three data points on the triplet (p, Q, AC), and in particular assuming that prices are exogenous with respect to both demand and insurable costs, we can immediately see that the competitive equilibrium price (i.e., where price is equal to average cost) is $p = 4$. We can also back out the cost c of the marginal individual whose allocation is affected when the price changes. For example, when the price is raised from $p = 2$ to $p = 4$ the marginal cost is given by $\frac{\Delta(AC*Q)}{\Delta Q} = \frac{AC(p=2)Q(p=2) - AC(p=4)Q(p=4)}{Q(p=2) - Q(p=4)} = \frac{3 \cdot 1 - 4 \cdot \frac{2}{3}}{1 - \frac{2}{3}} = 1$. Likewise, the willingness to pay π for the marginal individual is equal to the price, 2,[15] and the mass of such individuals is equal to the change in market share associated with this price change: $Q(p = 2) - Q(p = 4) = 1 - \frac{2}{3} = \frac{1}{3}$. Using such estimates of the expected cost and willingness to pay for insurance of the marginal individual, we can now compute total surplus for any given price. For example, we can conclude that it is inefficient for the marginal individual at $p = 2$ to not have insurance, that each such individual would gain a surplus (i.e., $\pi - c$) of $2 - 1 = 1$, and that there is a mass of $\frac{1}{3}$ such individuals in the market. Thus, the efficiency cost of adverse selection in such a market would be $\frac{1}{3}$ per market participant.

Extensions to the basic framework As mentioned, the basic framework we described in Section 2 made a number of simplifying assumptions for expositional purposes which do not limit the ability to apply this approach more broadly. It is straightforward to apply the approach to the case where the high coverage contract provides less than full coverage and/or where the low coverage contract provides some coverage; 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 H coverage rather than providing L coverage. We discuss a specific example of this in our application below.

[15] This is not completely precise. Given the example, all we would know is that the willingness to pay by the marginal guy is $2 \leq \pi < 4$. We would know that $\pi = 2$ with more continuous variation in price, or if we knew that the support of the willingness-to-pay distribution is 2, 4, and 6.

Likewise, while it was simpler to show the analysis graphically with only two coverage choices, estimation with more than two coverage choices is straightforward. 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 $AC(p)$, 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 $p = AC(p)$. From the estimated systems $D(p)$ and $AC(p)$ one can also back out the system of marginal costs $MC(p)$ which defines the marginal costs associated with each price vector. We can then solve $p = MC(p)$ for the efficient price vector and integrate $D(p) - MC(p)$ over the (multi-dimensional) difference between the competitive and the efficient price vectors to obtain the welfare cost of selection.

A direct test of selection Although the focus of our paper is on estimating the welfare cost of selection, a very nice feature of our proposed framework is that it simultaneously provides a direct test 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 (i.e., slope of zero) allows us to reject the null of no selection.[16] 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, while an upward sloping curve indicates advantageous selection.[17]

An appealing property of this selection test is that it allows a distinct test for selection that is not affected by the existence of moral hazard (or lack thereof). To see this, note that the AC curve is estimated (and hence the MC curve is derived) 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. Since coverage is held fixed, our estimate of the slope of the MC curve (our test of selection) is not affected by moral hazard (which determines how costs are affected as coverage changes).

By contrast, the influential and widely used "positive correlation" test (see, e.g., Cawley and Philipson, 1999; Chiappori and Salanie, 2000; and Finkelstein and Poterba, 2004), which compares realized risks of individuals with different insurance contracts, jointly tests for the existence of either

[16] Using the terminology we defined in Section 2.2, 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 is not first best.

[17] Conceptually, adverse selection refers to a monotonically declining marginal cost curve, and advantageous selection to a monotonically increasing marginal cost curve. In practice, most empirical tests of selection look globally at average costs under different insurance contracts rather than locally at the marginal costs for the marginal market participant (see, e.g., Finkelstein and Poterba (2004) for a case of adverse selection, or Fang, Keane, and Silverman (2008) for a case of advantageous selection). As long as the marginal cost curve is monotone, the inferences are valid.

selection or moral hazard (but not for each separately). Exogenous pricing variation – which is not required for the "positive correlation" test – is the key to a distinct test for selection. It allows us to analyze how the risk characteristics of the sample who selects a given insurance contract varies as we vary the price of that contract.

Estimating moral hazard As discussed in Section 2.4, our main analysis remains the same with or without moral hazard. Nonetheless, we note here that 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 he has H coverage – and $c^L(\zeta_i)$ – individual i's expected insurable cost when he has L coverage. 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 to the insurer from providing H that arises from the effect of coverage by H on the behavior of covered individuals. We already discussed above how price variation can be used to estimate what we previously referred to as the AC and MC curves, which are denoted by AC^H and MC^H when moral hazard is explicitly recognized. With data on the costs of the uninsured (or less insured, if L represents lower but not zero coverage), we can repeat the same exercise to obtain an estimate for AC^L and MC^L. That is, we can use the very same price variation to estimate demand for the L contract and to estimate the AC^L curve from the (endogenously selected) sample of individuals who chose L. We can then back out an MC^L curve analogously to the way we backed out the MC^H curve, using of course the demand curve for L rather than for H and AC^L rather than AC^H in translating average costs into marginal costs (see equation (11)). The (point-by-point) vertical difference between MC^H and MC^L curves provides an estimate of moral hazard (see Figure 3). 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.[18]

Of course, it is not a new observation that if we have an exogenous shifter of insurance coverage (which in our context comes from pricing) we can estimate the moral hazard effect of insurance. However, one attractive property of our proposed approach to estimating moral hazard (rather than, say, a more standard instrumental variable framework) is that (with sufficiently rich price variation) we can see how moral hazard varies across individuals with different willingness to pay $\pi(\zeta_i)$, or different expected costs to the insurer $c^H(\zeta_i)$.

[18] This would give 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 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 outcome of interest.

4 Empirical application: employer-provided health insurance

4.1 Data and environment

We implement and illustrate the approach we have just outlined using individual-level data from 2004 on the U.S.-based workers (and their dependents) at Alcoa, Inc. In 2004, Alcoa had approximately 45,000 active employees in the U.S. working at about 300 different job sites in 39 different states. At that time, in an effort to control health care spending, Alcoa introduced a new set of health insurance options to virtually all its salaried employees and about one-half of its hourly employees. We analyze the choices of employees offered the new set of options in 2004.[19]

The data contain the menu of health insurance options available to each employee, the employee premium associated with each option, the employee's health insurance choice from the menu, and detailed information on his (and any covered dependents') medical expenditures for the year.[20] Crucially, the data also 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, 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 prices can observe without direct personal contact, as well as some characteristics that the price setters might not be able to observe (such as detailed health care utilization and expenditure information from previous years). This is important because it allows us to examine whether the variation in contract pricing across employees appears random with respect to the characteristics that could potentially influence the price setters' decisions.

We make a number of sample restrictions. First, we make a number of restrictions for purposes of data purity, which brings the original sample of about 45,000 active workers down to about 37,000 active workers.[21] Second, because the new set of health insurance options we study did not apply to many hourly workers and because (as we discuss below) the pricing variation is cleaner for the salaried workers, we further limit the analysis in this paper to salaried employees, who are

[19] 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. In Einav et al. (in progress) we use the staggered timing across hourly employees in the transition from one set of contract offerings to another to study the impact of consumer cost sharing on medical expenditures.

[20] Health insurance choices are made during the open enrollment period at the end of 2003 and apply for all of 2004. The medical expenditure data cover all of 2004. We also observe medical expenditure in 2003 if the employee worked at the company for all of 2003.

[21] The biggest reduction in sample size comes from excluding workers who are not at the company for the entire year (for whom we do not observe complete annual medical expenditures, which are necessary for estimating the cost curve). In addition we exclude employees who are outside the traditional benefit structure of the company (for example because they were working for a recently acquired company with a different (grandfathered) benefit structure); for such employees we do not have detailed information on their insurance options and choices. We also exclude a small number of employees because of missing data or data discrepancies.

approximately one third of the employees.

Third, to illustrate most easily how the theoretical framework maps to the empirical strategy, we limit the analysis to the two modal health insurance choices: a higher and a lower level of PPO coverage; we refer to these hereafter as the "High" coverage and "Low" coverage options. Approximately two-thirds of salaried workers chose one of these two PPO options. In Section 4.5 we describe the other options in more detail and show that the pricing of the two PPOs we focus on does not affect the probability of the employee choosing one of the other options we do not analyze. This helps to alleviate concerns about potential biases from our sample selection.

Finally, for simplicity, our baseline specification further limits our sample of salaried workers who choose the High or Low coverage options to the slightly over one half of employees who chooses the most common coverage tier, which is family coverage.[22] In Section 4.5 we show that our results are similar when we include employees in all coverage tiers. We assume throughout that the choice of coverage tier is unrelated to the pricing variation.[23]

Table 1 provides some descriptive statistics on the employees. Column (1) presents descriptive statistics for the sample of 37,000 active workers for whom we have complete data. Column (2) limits the sample to the approximately one third of the sample who are salaried workers. Column (3) makes the further (minimal) restriction to the salaried workers who face the new benefit design. Column (4) further limits the sample to workers who choose High or Low coverage and column (5) further limits the sample to those in family coverage. Column (5) represents our baseline sample that we use for most of the empirical analysis; Section 4.5 presents analyses using all coverage tiers (column (4)) and using all plan options (column (3)). For comparison, columns (6) through (8) of Table 1 present statistics from the 2005 March Current Population Survey (CPS) on characteristics of various types of full time employees in the United States.

4.2 Empirical strategy and relationship with the theoretical framework

Notation Following our earlier nomenclature, we will use the notation H and L to reflect the High and Low coverage PPO options, respectively. We denote the relative price (i.e., premium) employee i faces for the H contract relative to the L contract to be $p_i = p_i^H - p_i^L$, where p_i^j is employee i's annual premium if she chose coverage j. Crucially for our approach, p_i varies across individuals in a plausibly random way; we defer a detailed discussion of the nature and source of this variation to Section 4.3. We define D_i to be equal to 1 if employee i chooses coverage H and 0 if employee i chooses coverage L. Finally, m_i is a vector representing total medical expenditures

[22] Employees always have a choice of four different tiers for health insurance coverage: employee only, employee plus spouse, employee plus children, and family coverage.

[23] A priori, this seems a reasonable assumption given that coverage tier options are limited by the demographic composition of the family, and that the price multiplier across coverage tiers is the same for all employees. Specifically, for any health insurance coverage option, for all employees the family price is always triple the "employee only" price, 1.58 times the "employee plus children" price, and 1.43 times the "employee plus spouse" option. Consistent with our assumption, we find that the relative price of High coverage compared to Low coverage in the family coverage tier does not predict (either economically or statistically) which coverage tier the employee chooses.

of employee i and any covered family members in 2004.[24]

Construction of the cost data $\mathbf{c}_i$ In our theoretical discussion in Section 2 we defined (for simplicity) H to be full coverage and 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. In more general settings (such as our current application), when H may not be full coverage and L may provide some coverage, the generalization requires us to define c_i to be the *incremental* cost to the insurer from providing higher coverage relative to providing lower coverage, holding m_i constant. In particular, let $c(m_i; H)$ and $c(m_i; L)$ denote the cost to the insurance company from medical expenditures totalling m_i under coverages H and L, respectively. Thus, the incremental cost is 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 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 from the rules of the Low coverage plan. For consistency, we therefore 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 expenditures. This allows us to construct the cost to the insurance company of insuring medical expenditures m_i under any particular plan j (i.e., $c(m_i; j)$).[25] Our two plans vary only in their consumer cost sharing rules; specifically, the L coverage has higher deductibles and higher out of pocket maximums.[26] We believe we can calculate $c(m_i; j)$ with a great deal of accuracy. For example, for individuals with High coverage the correlation between their actual share of spending out of pocket and our calculated share of spending out of pocket is over 97 percent.

Figure 4 illustrates the major differences in consumer cost sharing between the two coverage options – and the construction of c_i – graphically. Cost sharing rules differ depending on whether spending is in network or out of network. Figure 4(a) 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, the H plan has no deductible while the L plan has a \$500 deductible; both have a 10 percent coinsurance rate, and the out of pocket maximum is \$5,000 for H and \$5,500 for L. Figure 4(b) 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, 25% of the individuals in our

[24] While it is easy to think of m_i as a scalar, the specific details of the coverage often make the analysis depend on how the overall expenditure is allocated among family members and between in-network and out-of-network care. See Figure 4 and footnote 28 for more details.

[25] For example, in our setting, coverage rules vary depending in whether the claim occurs in or out of network, therefore to construct $c(m_i; j)$ we must observe whether each claim occurs in or out of network. In the context of a choice between an HMO and a PPO, construction of $c(m_i; j)$ would require (among other things) information on plan j's coverage rules for different medical providers, as well as data on the medical provider for each claim.

[26] 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 H and L in plan features that might differ in unobservable ways across different workers (for example, differences in the relative network quality across different parts of the country).

baseline sample file at least one claim out of network, making the out-of-network coverage an important part of the analysis.[27]

Figures 4(c) and 4(d) show the implied difference in out-of-pocket spending between the Low and High coverage, for a given level of annual total medical spending m_i; by construction, this is equal to c_i.[28] Figure 5(a) presents the empirical distribution of the c_i variable. The distribution of c_i reflects the various kinks in the coverage plans illustrated in Figure 4. 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 (total) medical expenditures.

The nature of the plan differences is important for understanding the margin on which we may detect selection (or moral hazard). Empirically, because very few people spend anywhere close to the out of pocket maximum of even the H contract, the difference in insurer's cost between the plans occurs mainly because of 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 of employees (see footnote 48) who have total medical expenditures that are either less than the Low coverage deductible or fall exactly in the small range that makes only the out-of-pocket maximum of High coverage (but not of Low coverage) apply.

Finally, Figure 5(b) shows the distribution of c_i separately for those who choose High coverage and those who choose Low coverage. It shows important heterogeneity in the cost distribution between the two plans. In the spirit of the "positive correlation" test (Chiappori and Salanie, 2000), the higher average costs for those in H coverage compared to those in L coverage 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.

[27] There is no interaction between the in-network and out-of-network coverages; each deductible and out of pocket maximum must be satisfied separately.

[28] Figure 4 and the description of the plans in the figure notes and text above abstract from a few details, all of which are accounted for in our construction of c_i. First, both plans (identically) specify certain expenditures that are fully covered (such as various types of preventive care). Second, we have described the rules for any coverage tier that covers multiple individuals (i.e., family coverage, employee plus children coverage, and employee plus spouse coverage). Employees who choose the employee only coverage tier face deductibles and out of pocket maxima that are one half of what is described in Figure 4 and in the preceding text above. Third, as is typical of most health insurance plans, the cost sharing rules in a plan that covers more than two family member can vary depending on how the spending is distributed among family members. In particular, a given individual in a family can exhaust his deductible or reach his out-of-pocket maximum either by spending the requisite amount that is required by employee only coverage or by having the cumulative spending of other members of the family reach the family deductible or the family out-of-pocket maximum. We account for the composition of spending within the family in generating the predicted consumer cost sharing under the different plans. Fourth, in addition to the in-network and out-of- network options, a very small number of employees are eligible for a "network not available" status, which allows them to go out of network at a lower cost (but still higher than the in-network cost).

Baseline estimating equations For our baselines 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 following two equations

$$D_i = \alpha + \beta p_i + \epsilon_i \tag{14}$$

$$c_i = \gamma + \delta p_i + u_i \tag{15}$$

where D_i, c_i, and p_i are defined earlier. We adjust the standard errors to allow for an arbitrary variance-covariance matrix within each state. Following the theoretical framework, the demand equation is estimated on the entire sample, while the (average) cost equation is estimated on the sample of individuals who (endogenously) choose the High coverage contract.

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 (11), the marginal cost curve is given by

$$MC(p) = \frac{1}{\beta}\left(\frac{\partial\left(\alpha+\beta p\right)\left(\gamma+\delta p\right)}{\partial p}\right) = \frac{1}{\beta}\left(\alpha\delta+\gamma\beta+2\beta\delta p\right) = \frac{\alpha\delta}{\beta}+\gamma+2\delta p. \tag{16}$$

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 $AC(p) = D(p)$, resulting in $P_{eq} = \frac{\gamma}{1-\delta}$ and $Q_{eq} = \alpha + \beta\frac{\gamma}{1-\delta}$. The efficient price and quantity are given by equating $MC(p) = D(p)$, resulting in $P_{eff} = \frac{1}{1-2\delta}\left(\frac{\alpha\delta}{\beta}+\gamma\right)$ and $Q_{eff} = \alpha + \frac{1}{1-2\delta}\left(\alpha\delta+\beta\gamma\right)$. The efficiency cost of adverse selection (measured by the area of triangle CDE in Figure 1) is then given by

$$\Delta_{CDE} = \frac{1}{2}\left(Q_{eff}-Q_{eq}\right)\left(P_{eq}-MC(P_{eq})\right) = \frac{-\delta^2}{2\left(1-2\delta\right)\beta}\left(\alpha+\frac{\beta\gamma}{1-\delta}\right)^2. \tag{17}$$

When the model is not linear, we compute these estimates numerically based on the point estimates from the demand and cost curves.

4.3 Variation in prices

Company structure as the source of variation An essential element in the analysis is that there is variation across workers in the relative price they face for the High coverage option, and that this variation is unrelated to the workers' willingness to pay for High coverage or to his insurable costs. Alcoa's business structure provides a credible source of such pricing variation across different workers in the company.

In 2004, 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 (i.e., employee premiums) of the options can vary. For our purposes, the key element of interest is the incremental premium the employee must pay for the High coverage option relative to the Low coverage option, $p = p_H - p_L$; we refer to this relative price of High coverage 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.[29]

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 their business in the manner they see fit, provided that they do so ethically and safely, and at or above the company's normal rate of return; failure on any of these dimensions 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 exterior – 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 6,000 (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 (i.e., job site) and their employment type (salaried or hourly worker and, if hourly, which union if at all the worker is in).

As a result of this business structure, workers doing the same job in the same location may face different prices for their health insurance benefits due to their business unit affiliation. A priori, it struck us as more plausible that the pricing variation across salaried workers in different business units is more likely to be exogenous – reflecting idiosyncratic characteristics of the business unit presidents rather than differences in the salaried workers in the different business units – than the pricing variation across hourly workers. Many of the jobs that salaried workers 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 workers' 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 are consistent with this basic intuition. Table 2 compares mean demographic characteristics of workers in our baseline sample who face different relative prices. Importantly, we observe all of the characteristics of the employee that the business unit president (or his human resource director) is likely to observe. In general, the results look quite balanced. There is no substantive or statistically significant difference across employees who face different prices in average age, fraction male, fraction white,

[29] The price to the employee of the High coverage options 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 reason why it makes sense to focus the analysis on the difference in prices for the different coverage options, rather than the level of prices.

average (log) wages, average age of spouse, number of covered family members, age of the youngest child, or in the average (log) 2003 medical spending. The two possible exceptions to this general pattern are average job tenure and average (log) 2003 medical expenditures 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).[30] A joint $F-test$ of all of the coefficients leaves us unable to reject the null that they are jointly uncorrelated with price; when we examine the eight contemporaneous characteristics (age, job tenure, gender, race, wages, age of spouse, number of covered family members, age of youngest child), we estimate an $F-stat$ of 1.71 ($p-value = 0.14$).[31]

We also examined if covariates appear balanced when we condition on state fixed effects, to allow for the possibility that prices – which can vary by job site and therefore by state within a business unit – may vary non-randomly across states (for example, reflecting different health care utilization patterns or costs). With state fixed effects, the value of the $F-stat$ declines to 1.16 ($p-value = 0.35$). There is still substantial variation in prices across workers within state; in our baseline sample, the overall standard deviation in price is \$60 and the within-state standard deviation is \$48.[32] We present results below both with and without state fixed effects and show that they are quite similar. We also present results below when all four coverage tiers are pooled and again find that covariates appear balanced for salaried workers in this pooled sample ($F-stat$ of 1.46, $p-value = 0.21$).[33] Finally, we present results below in which we include all salaried workers, rather than just the two-thirds who chose H or L; the covariates still appear balanced in this expanded sample ($F-stat$ of 1.70, $p-value = 0.14$).

By contrast, similar analysis of covariates for hourly workers suggests statistically significant differences across employees who face different prices; the $p-value$ on the joint test of covariates is less than 0.01 for hourly workers with family coverage (with or without state fixed effects). As noted, this is not surprising given the institutional environment, and motivates the restriction of our analysis to salaried employees.

4.4 Baseline results

Our baseline specification estimates the linear demand and cost curves shown in equations (14) and (15) on our baseline sample of 3,779 salaried workers in the family coverage tier who choose

[30] The $p-value$ on each of these two variables is 0.08. We should note, of course, that when testing 10 different variables 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.

[31] 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 percent) we obtain an $F-stat$ of 0.98 ($p-value$ of 0.48).

[32] While there are instances of multiple business units within the same job site, so that in principle there could even be within-job site variation in prices across salaried workers, in practice in our data the multiple business units in the same job site always chose the same pricing menu for their salaried workers.

[33] As we discuss in more detail in Section 4.5, in all analyses of a sample with more than one coverage tier (including the analysis just described) we include tier dummies – since prices vary with coverage tier – and we double the price of the "employee only" coverage to account for the fact that its deductible and out-of-pocket maximum are half that of the other three coverage tiers.

either High or Low coverage. This allows us to walk through the main conceptual points of interest involved in applying our proposed approach. We then turn in Section 4.5 to a more thorough and detailed discussion of empirical issues specific to our context, including alternative samples and specifications.

Table 3 shows the raw data for our key variables. The relative price of the high coverage ranges from $384 to $659, with about three-quarters of the sample facing the lowest price. Column (3) shows that the market share of the High coverage option 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 High coverage is generally increasing with price (or equivalently, declining in quantity) as we would expect with adverse selection (Figure 1). Column (5) shows the same for the individuals who (endogenously) select Low coverage. Recall that cost is defined as the difference in costs to the insurer associated with a given employee's family's medical spending if those expenditures were insured under the High coverage option relative to the Low coverage option; as shown in Figure 4, this difference is a non-linear function of the family's medical spending.

Table 4 reports our baseline results from these data. Column 1 shows our estimate of the demand curve. We estimate a downward sloping demand curve, with a coefficient β on price of -0.00070 (standard error = 0.00032). This implies that a $100 increase in price reduces the probability that the employee chooses the higher coverage plan by a statistically significant 7 percentage points, or about 11 percent.[34]

Column (2) shows our estimate of the average cost curve. As noted, the estimated slope of the cost curve represents a new test for the existence and nature (adverse or advantageous) of selection. The coefficient δ on price is 0.155 (standard error = 0.064). The positive coefficient on price indicates the presence of adverse selection. That is, the average cost of individuals who purchased High coverage is (statistically significantly) higher when the price is higher. In other words, when the price selects those who have, on average, higher willingness to pay for High coverage, the average costs of this group are also higher. The average cost curve is therefore downward sloping (i.e., decreasing in market share; see Figure 1).

The point estimate from our baseline specification suggest that a dollar increase in the relative price of the High coverage is associated with an increase in the average cost of the (endogenous) sample selecting High coverage at that price of about 16 cents (Table 4, column (2)). 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

[34] This semi-elasticity has been estimated in several other health insurance contexts. Differences in the contract choices across these different settings mean that comparisons of these semi-elasticities across different contexts are not very meaningful. Nonetheless, we note for completeness and casual interest that our estimate of a semi-elasticity (with respect to a $100 increase in premium) around -11 is somewhat larger than the typical semi-elasticities estimated (which tend to be around -3 to -4) although by no means the highest in the literature; Bundorf, Levin, and Mahoney (2008) and Chernew, Gowrisankaran, and Scanlon (2008) provide summaries of the existing studies.

quantifying its efficiency cost and, relatedly, to calculating the welfare benefits from alternative public policy interventions in the market.

In this spirit, Figure 6 shows how to translate the baseline empirical estimates of the demand and cost curves in Table 4 into the theoretical welfare analysis shown in Figure 1. That is, Figure 6 is the empirical analog to Figure 1. It graphs the estimated demand curve and average cost curve shown in Table 4. It also shows the marginal cost curve implied by these estimates (see equation (16)). Based on these estimates, it is straightforward to calculate the implied welfare cost of adverse selection (i.e., the area of CDE in Figure 6 (and 1); see also equation (17)). It should be readily apparent from the figure that, holding the cost curve constant, shifting and/or rotating the demand curve could generate very different welfare costs (that is, areas of the triangle CDE). This underscores the observation that merely estimating the slope of the cost curve (i.e., detecting adverse selection) is not by itself informative about the likely magnitude of the resultant inefficiency.

We estimate that the welfare cost of adverse selection (i.e., area of CDE) is \$9.55 per employee per year. 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 market share of High coverage by 14 percentage points. The social benefit of providing H coverage to the marginal employee who buys L in equilibrium (i.e., the vertical distance between point C and point D in Figure 6) is \$138.

Figure 6 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 3), 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 (see Section 4.5).

The price range of \$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 percent higher than our estimate of the efficient price (point E) of \$264. Thus, while in principle our approach does not require parametric assumptions – with a wide enough span of prices we could estimate the demand and cost curves non parametrically – 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 robustness analysis below we examine alternative functional forms.

Benchmarks for the welfare cost of adverse selection We can also use the demand and cost curves shown in Figure 6 to calculate various benchmarks that provide some context for our estimate of the welfare cost of adverse selection (i.e., triangle CDE) of \$9.55 per employee. An important consideration in choosing a benchmark is how out of sample we must take the demand and cost estimates in order to form it; again, Figure 6 is informative on this point.

We calculate two useful denominators to scale our estimate of the welfare cost of adverse selection. One is a measure of how large this cost could have been before we started the analysis. Here

we follow the upper bound discussed in Section 3 which can be derived without having exogenous variation in price. While there we assumed that we observed only the equilibrium price, it is easy to show that the bound is valid as long at we are willing to assume that the observed price is at or above the efficient price, so that the inefficiency generated by selection is one of under-insurance (which is consistent with the adverse selection we detect in the data). The thought experiment is to assume that we observe data (on price, quantity, and costs) from only one of the rows of Table 3, so there is no price variation; we assume we observe the weighted average price of $414. Since individuals have the option to buy High coverage at this price but choose not to do so, their welfare loss from being inefficiently uncovered by this option cannot exceed $414. Our estimate of the efficiency cost of adverse selection of $9.55 is therefore 2.3 percent of this "maximum money at stake," as Einav, Finkelstein, and Schrimpf (2007) term this construct.

A second useful denominator is to scale the welfare loss from adverse selection by the total surplus at stake from efficient pricing or, in other words, the maximum welfare loss that could have occurred given our estimates of the demand and cost curves. We therefore calculate the ratio of triangle CDE (welfare loss from adverse selection) to triangle ABE (welfare from efficient pricing) in Figure 1. To enhance readability, points A and B are not shown in Figure 6, but of course are easily calculated from the demand and cost estimates in Table 4. They are, however, fairly out of sample relative to our data. For example, at point A we estimate 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 below that the result is relatively robust to alternative functional forms for that extrapolation). We estimate that the welfare loss from adverse selection (area CDE) is about 3 percent of the surplus at stake from efficient pricing (area ABE).

Welfare under other market allocations Although our welfare analysis has focused on the efficiency cost of adverse selection in a competitive equilibrium, the fact that we observe prices varying – and this is how we identify the demand and cost curves – underscores the point that to generate pricing variation we observe a market that is not in equilibrium. Our analysis of "equilibrium" pricing – like our analysis of "efficient" pricing – is therefore based on counterfactuals. By the same token, our analysis of the efficiency cost of adverse selection in this market 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.

Since 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 adverse selection under any other assumption about 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 3, columns (1) and (2)) is $6.26 per employee per year. Moreover, as we noted in Section 2, we could also use the estimated demand and cost curves to estimate welfare under alternative assumptions about the market equilibrium, including imperfect competition.

One interesting alternative is to compute what the welfare cost of adverse selection would be if –

contrary to what happens in the employment context – in a free market setting 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 the welfare cost of adverse selection if prices were set differently based on whether the family coverage applied to 3 individuals, 4 individuals, or 5 or more individuals; roughly half of our baseline sample has 4 covered members, and the remaining sample is roughly 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 sub-sample separately, and estimate that the (weighted average) welfare cost of this selection would be \$12.92 if prices were set differently by 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 estimates the welfare cost of standard public policy interventions in insurance markets with adverse selection and compare this to our estimate of the welfare cost of adverse selection. Our preferred policy analysis 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. Such a calculation provides a guide to whether there is scope for welfare improving government intervention in the form of price subsidies; moreover, it does not require further out of sample extrapolation beyond what is needed to compute the welfare cost of adverse selection itself. The social cost of the optimal subsidy is given by $\lambda(P_{eqm} - P_{eff})Q_{eff}$ where λ is the marginal cost of public funds. Given our estimates of the efficient and equilibrium outcomes (Figure 6), and using a standard estimate of the marginal cost of public funds of about 0.3 (e.g., Poterba, 1996), we calculate the social cost of the price subsidy needed to achieve the efficient allocation to be \$45. In other words, 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.

We also compared welfare in the competitive equilibrium with adverse selection to welfare when everyone is mandated to be covered by the High coverage plan. Mandatory insurance is the canonical solution to the problem of adverse selection in insurance markets (e.g., Akerlof, 1970). This makes the analysis of the mandate of considerable interest.[35] However, in our application, the welfare cost of mandating coverage by H (i.e., area EGH in Figure 1) requires calculating points which are reasonably far out of sample. This suggests that more caution is warranted with this analysis (although again we will show below that the estimate is reasonably robust). With this important caveat in mind, we estimate that the welfare cost from mandatory coverage by H is about three times higher than the equilibrium welfare cost of adverse selection.

[35] Footnote 4 discussed some of the possible factors that may make it inefficient to allocate the H contract to the entire market.

4.5 Robustness and extensions

In this section we explore the sensitivity of our welfare estimates to a number of alternative specifications, and consider several extensions to our analysis. Our overall finding is that the magnitude of the various welfare estimates discussed above – even those that involve extrapolation considerably out of sample – are quite robust. In particular, across various alternative specifications, the welfare gain from a price subsidy that achieves the efficient price is always substantially below the social cost of the required price subsidy, the welfare loss from adverse selection when choice over contracts is allowed is always lower than the welfare loss from mandatory coverage by the H contract, and the welfare cost of adverse selection is always less than 10 percent of the surplus that could be generated from efficient pricing.

Functional form and theoretical restrictions on the demand curve Table 5 summarizes some of the sensitivity analyses. Panel A summarizes the implied welfare implications of each specification. For completeness, Panel B shows the corresponding parameter estimates from each specification (from which the welfare estimates shown in Panel A are derived). In the interest of brevity we focus our discussion primarily on the robustness of the resultant welfare estimates (columns 6 through 8 of Panel A), which are our main interest. The first row of Table 5 presents the results from our baseline specification (Table 4). Subsequent rows reports results from a single, specified departure from this baseline.

Rows 2-5 in Table 5 explore the sensitivity of our results to our functional form assumptions. Row 2 shows the results from our baseline specification (row 1) are quite similar if we estimate a probit for the demand equation rather than a linear demand. Quadratic demand (not reported) behaves very badly out of sample and is therefore not shown (but in row 5 we report and discuss a restricted specification that includes a quadratic demand curve). As can be seen in Figure 6, the linear specification fits the cost data very well.[36]

We also experimented with imposing restrictions on the demand curve that are implied by basic price theory. Willingness to pay is (theoretically) bounded from above at \$1,500 (the maximum possible out-of-pocket savings from High coverage; see Figure 4) and (theoretically) bounded from below by 0 (any rational individual should always prefer more coverage to less if the former is offered for free). Our baseline demand estimate (Table 5, row 1) satisfies the first constraint (market share for High coverage becomes 0 at a price of \$1,350), but not the second; at a price of 0, market share for High coverage is only 0.94.[37] The results in row 3 show that constraining market share to be 1

[36] We explored alternative functional forms for the cost curve – such as a quadratic, log-log, and log-linear functions. Not surprisingly, the results (not shown) were very similar in sample. However, curvature (concavity in particular) in the estimated AC curve sometimes led to an implied non-monotone MC curve (out of sample), which led to strange out-of-sample predictions.

[37] One reason why we may estimate demand below 1 for a price of 0 is that our functional form assumption of linear demand is not appropriate for extrapolating this far out of sample. Another possible explanation may be that the L option was the default in 2004. We suspect that default may be less important in our setting than in others because 2004 was the first year in which the new benefits were offered. These new benefits came with much effort by the company to advertise and explain the new options to its employees, making it likely that most individuals were

when price is 0 does not noticeably affect our welfare estimates. Row 4 shows the results are also similar if we impose the constraint that willingness to pay is bounded at $800, which may be a more reasonable upper bound in practice than the theoretically possible $1,500.[38] Row 5 estimates a quadratic demand curve, imposing both the (1,$0) and the (0,$800) constraints on (Q,P), and again the welfare estimates are quite stable.

Tax treatment of employee premiums We also considered the sensitivity of our results to the tax treatment of employee contributions to health insurance and to out-of-pocket medical expenditures. Employee premium contributions are made pre-tax. Employees can pay their out-of-pocket medical spending pre-tax as well, by contributing to a Flexible Spending Account (FSA). If all out-of-pocket expenses were paid pre-tax, the tax treatment of employee premiums and employee medical spending would be symmetric, and ignoring the tax subsidy to employee premiums (as we do in our baseline specification) would be appropriate. However, in practice, less than a quarter of employees contribute to an FSA. It is of course unclear whether employees who do not take advantage of the tax subsidy to out-of-pocket medical spending offered by FSAs are cognizant of the tax subsidy to employee premiums. However, to investigate the sensitivity of our findings to the tax subsidy, we consider the effect on our estimates of assuming that all employees (including those who contribute to FSAs) make their health insurance choices based on the pre-tax price.[39] We calculate the average tax subsidy (i.e., one minus the average marginal tax rate) for our sample to be 65 percent.[40] In row 6, we therefore re-estimate the baseline specification with the price variable in both the demand and cost equations multiplied by 0.65. Once again the core welfare estimates are not noticeably affected, although naturally our estimates of the equilibrium and efficient allocations (see columns 1 through 4) shift considerably.

"active" choosers. Moreover, it is possible to have a model of defaults in which our welfare anlaysis is unaffected. We discuss this in a little more detail below.

[38] Out of pocket savings from H of $1,500 is only hit if the covered family members spend enough in-network *and* out-of-network to hit the (separate) out of pocket maximums; in practice, this never occurs (indeed, none of the employees in our sample hits the out of pocket maximum out-of-network and only about 1 percent hits the in-network out-of-pocket maximum). A potentially more reasonable constraint therefore is that willingness to pay for H should not exceed $800, which is the reduction in out-of-pocket expenditures associated with H if the family spends more than the deductible in-network and more than the deductible out-of-network but less than the amount that would cause them to hit the out-of-pocket maximum (see Figures 4(c) and 4(d)).

[39] We do not observe in the data which individuals participate in the FSA.

[40] The tax subsidy is given by $(1 - \tau_f - \tau_s - \tau_{ss} - \tau_{mcr})$ where τ_f is the federal marginal tax rate, τ_S is the state marginal tax rate, τ_{ss} is the marginal Social Security (FICA) payroll tax on the employee, and τ_{mcr} is the marginal Medicare payroll tax on the employee. We estimate these marginal tax rates using the NBER's TAXSIM model, which takes as inputs the major determinants of marginal tax rates and computes the various marginal rates just mentioned. Many of the required data elements (or reasonable proxies for them) are available in our company's data, including annual wage and salary income, state, marital status, number of dependents and ages of family members. We assume all employees with family coverage file jointly and do not itemize. We impute wage and salary income of spouse, property income, and dividend income based on the ratio of each of these variables to own income for the sample of full time, white collar manufacturing workers in the March CPS; we pool the 2004 - 2007 March CPS to increase sample size (Table 1, column 8 presents descriptive statistics for this sample in the March 2005 CPS). All other inputs required by TAXSIM are assumed to be zero. For more information on TAXSIM, see www.nber.org/taxsim.

Additional covariates and alternative samples Our baseline estimates of the demand and cost curves include no covariates in the analysis besides the relative price. Only variables that are priced should be controlled for in our analysis of selection and its welfare costs. The fact that, for example, individuals of, say, different incomes or different ages may have different expected medical costs – and that this may affect which plan they choose – is part of the endogenous selection we wish to study, rather than control for, since these characteristics are not priced. However, to allow for the possibility that the price menu may be selected differently across states in a non random fashion (e.g., reflecting differences in health care costs across states), in row 7 we include state fixed effects in the demand and cost estimates. Although our estimates become somewhat less precise (Panel B), the welfare implications remain quantitatively similar (Panel A). In row 8 we add all of the contemporary employee characteristics (Table 2) as covariates to the demand and cost curves (in addition to the state fixed effects).[41] Once again the results are very similar. The fact that the slope of the estimated demand curve remains similar is unsurprising given the evidence in Table 2 that pricing is orthogonal to these employee characteristics. The fact that the slope of the estimated cost curve remains similar suggests that the adverse selection we detect is not driven by the fact that in our setting the observable characteristics of employees are not priced.[42]

Finally, in row 9 we estimate our baseline specification using all four coverage tiers rather than just employees with family coverage. Since prices vary by coverage tier, we include (de-meaned) indicator variables for the coverage tier in both the demand and cost estimates.[43] The parameter estimates and welfare implications are quite similar to our baseline results.

Possible sample selection An important potential concern with all of the foregoing analyses is that we limit the sample to only those who choose High and Low coverage PPO options, and exclude the approximately one-third of salaried workers who chose one of the five other available options. These five other options are an HMO (about 7% of salaried workers choose this), opting out of any employer-provided coverage (about 8%), two even lower coverage PPO options (3% in the two of them combined), and a Health Reimbursement Account (HRA) PPO option, which combines a high deductible health insurance policy with tax preferred employer contributions that can be used to pay out-of-pocket expenses (approximately 17%).[44]

[41] In both rows 7 and 8 the covariates are demeaned so that the constant term is comparable across specifications.

[42] In this sense, the robustness test on the cost curve is one sided; had we found that the slope of the cost curve changed once we controlled for unpriced observables, this would not necessarily be a cause for concern. It could simply reflect the fact that much of the selection in our setting is driven by these unpriced observables.

[43] The price variable is defined for the chosen coverage tier. As noted in footnote 23, for all employees the prices of contracts in the other coverage tiers are always the same fixed multiple of the prices in the family coverage tier. To account for the fact that for "employee only" coverage the deductible and out-of-pocket maximum is half of what it is for the other three coverage tiers (see footnote 28), we multiply price (p_i) and cost (c_i) by two for the 16 percent of employees with "employee only" coverage.

[44] The in-network deductibles for the two lower coverage PPO options are \$1,000 and \$1,500. The high deductible HRA PPO has a \$3,000 in-network deductible, but the employee receives \$1,250 tax free from the company each year which can be spent on eligible medical expenses (including but not limited to the plan's cost sharing provisions). Unspent funds in the HRA can be rolled over the subsequent years, but any unspent balance is forfeited upon

In practice, however, our analysis suggests that our sample selection is unlikely to have important effects on our demand estimates (and, of course, it is irrelevant for the estimate of the cost curve which by design is run on the endogenously selected sample of individuals choosing High coverage). In particular, we found that the price of the High coverage option relative to the Low coverage option (our key right-hand-side variable) does not predict whether or not the employee "opts in" to one of High or Low coverage – as opposed to "opting out" into one of the remaining options. We suspect that this in part reflects the fact that many of the other options (in particular the three with non trivial market share, the HMO, opting out of insurance, and the HRA) are quite horizontally differentiated.

Table 6 presents some of these findings. The dependent variable in the reported linear regressions is an indicator variable that takes the value 1 if the employee chose one of the "outside goods" and 0 if he chose either High or Low coverage. The right-hand-side variable p is (as before) the relative price of High coverage compared to Low coverage. Column (1) reports the results for employees with family coverage. We find that a \$100 increase in the relative price of High coverage compared to Low coverage is associated with an economically and statistically insignificant decline (of 0.09 percentage points) in the probability of choosing one of the outside goods**.** Column (2) shows similar results when all coverage tiers are pooled. A complication with both of these analyses is that because coverage tier is not available for the 8 percent of the sample who opt out of coverage, these employees are excluded from the analysis. In column (3) therefore we include in the sample the employees who opt out of coverage; however, since coverage tier is not known for these employees we cannot control for coverage tier and, moreover, we can no longer define the price variable based on the coverage tier; we instead assign all employees the family prices regardless of what coverage tier they actually chose (if known).[45] Once again there is no evidence that the relative price of High coverage compared to Low coverage has an economically or statistically significant effect on the probability of choosing the outside good.

More than two coverage choices As noted in Section 2, it would be conceptually straightforward to extend our empirical analysis to consider more than two choices. However, we face practical obstacles to doing so in our setting. In particular, as is typical in data sets like ours, we do not observe medical expenditures for employees covered by an HMO or who opted out of employer-provided coverage; we therefore cannot estimate the cost curve for these options. It is also difficult to model the demand for these two options, since the prices are not known, nor is it entirely clear how to define the "good" being purchased.[46] We experimented with estimating demand and cost

separation from the company. The out-of-pocket maximums of all these options are also higher, but this is largely irrelevant empirically since it is extremely rare (less than 1%) for any employee to hit the out-of-pocket maximum of even the most generous option. Coinsurance rates are the same across all PPOs.

[45] Since, as noted, the prices of other coverage tiers are proportional to the family tier, this is not an unreasonable approach.

[46] The price of the HMO is literally not known, and likely varies across geographic areas. Employees receive a \$1,000 "credit" if they opt out of any coverage. However, without knowing what price they face for purchasing insurance outside of the company it is not clear what the true price is. Relatedly, in contrast to the PPO options,

systems for the remaining five PPO options. However, the relatively small sample sizes on the other three PPO options combined with the relatively high multi-collinearity in relative prices among the different PPO options resulted in fairly imprecise (and therefore relatively uninformative) estimates of the demand and cost systems for more than the two PPOs we focus on.

Moral hazard As we discussed in Section 2.4, our framework also allows us to easily test for and quantify moral hazard, which is defined by the vertical distance between MC^H and MC^L (Figure 3). With two partial coverage contracts, c^H is defined as the incremental cost to the insurer of covering the employee with H rather than with L assuming he behaves as if he is covered by H. Analogously, c^L is the incremental cost to the insurer of covering the employee with H rather than with L assuming he behaves as if he is covered by L. Our foregoing estimates of AC – which were estimated on the sample of individuals who chose H – therefore gives us AC^H. And our estimate of MC – using our estimate of AC^H and our estimate of the demand curve for H (equation (14)) – similarly gives us MC^H. To estimate AC^L we estimate the same cost equation (equation (15)) but now estimate it on the sample of individuals who chose L. To back out MC^L from AC^L we use the demand curve for L, i.e. equation (14) estimated with D_i replaced by $1 - D_i$.

In practice, in our setting, we were unable to reject the null of no moral hazard (i.e. $H_0 : MC^L = MC^H$). Our estimates were quite imprecise, suggesting that we may lack sufficient power in our setting to detect moral hazard.[47] This may not be surprising given that the design of the insurance contracts in our setting (see Figure 4) should make moral hazard only affect those people who expect to spend less than the Low coverage deductible or expect to fall exactly in the range that makes only the High coverage (but not the Low coverage) out-of-pocket maximum binding. In practice, this is likely to be a small fraction of our data.[48] As a different way to make this point, we applied the widely used moral hazard estimate of Manning et al. (1987) from the RAND Health Insurance Experiment to the total spending of each employee with High coverage based on the realized change in the marginal cost-sharing he would face with L coverage compared to H coverage, holding his behavior (i.e., m_i) fixed. This back-of-the-envelope calculation led to an average change in insurer's cost of 3%, driven by the fact that three quarters of the employees did not experience any change in marginal cost sharing. In light of this, we find it unsurprising that it is hard to detect moral hazard in this setting.

---

the characteristics of the HMO option and any coverage offered outside the firm are not known.

[47] The variation required to be able to detect moral hazard is greater than that required to be able to estimate the demand curve or the cost curve. A useful analogy is to an instrumental variable setup in which the demand analysis (how demand varies with price) serves as the first stage in the moral hazard estimate (how costs vary with demand). The $F-statistic$ on price in the demand equation in our baseline specification is only 4.8 (Table 4, column 1) suggesting that we do not have enough power for a second stage analysis of the impact of demand on costs. By contrast, our test of adverse selection (the analysis of the slope of the cost curve) is more in the spirit of our demand analysis – analyzing how a characteristic varies with price. Loosely, with this kind of variation, detecting adverse selection requires less power than detecting moral hazard.

[48] Considering in-network spending, there are 9% of the employees in our baseline sample who spend less than the Low coverage (in-network) deductible of $500 and none for whom only the High coverage out-of-pocket maximum is binding. Out-of-network spending would increase this share (but not by much).

Departures from revealed preference As we noted at the outset, our approach to welfare analysis has relied on revealed preferences. It is possible to use our framework for welfare analysis when we are not willing to assume revealed preferences, although this would require specification of the precise alternative choice model and how it maps to welfare. Some "behavioral" models are easily translated to our approach. Consider, for example, the possible role of defaults; the default option in our setting is the L contract. If one believes that there is a (constant) fraction α of the population who always chooses the default option of L, then it is possible to implement our approach, and perform welfare analysis on the remaining $1-\alpha$ population (of "active" choosers).

5 Conclusions

This paper proposed a simple approach to estimating welfare in insurance markets. As indicated by the title, the key to the approach is the existence of variation in prices across individuals that is unrelated to their demand for coverage or to their expected insurable costs. Applied welfare analysis usually relies on pricing variation that allows the researcher to trace out a demand curve. The defining feature of selection markets is that costs vary endogenously as market participants respond to the price of insurance. Welfare analysis of selection markets therefore requires that we also trace out the (endogenous) cost curve. We show how we can estimate the cost curve as well as the demand curve, and then proceed with welfare analysis in a familiar and straightforward fashion. In addition, the slope of the estimated cost curve provides a direct test of the existence and nature of selection.

We illustrated our framework by applying it in the particular context of the employer-provided health insurance at Alcoa, Inc., a large private employer in the United States. Alcoa's business structure generates plausibly exogenous variation in the price of more comprehensive relative to less comprehensive health insurance for otherwise similar workers. Using this variation, we find evidence of adverse selection in the market. However, we estimate that its efficiency implications are quantitatively small in both absolute and relative terms. Our central estimate implies a welfare cost of adverse selection of about \$10 per employee or about 3 percent of the total surplus at stake from efficient pricing. Moreover, standard policy instruments do not appear able to improve upon the adverse selection equilibrium. For example, we estimate that the social cost of the price subsidy required to achieve the efficient allocation is about five times higher than the social welfare gain from doing so.

It is important to emphasize that our empirical estimates are specific to a particular subsample of workers in a particular firm choosing between two specific coverage options. There is no reason to think our results are representative of the welfare cost of selection in other populations, other institutional environments, or other insurance markets. Our findings illustrate, however, that it is empirically possible to find markets in which adverse selection exists and impairs market efficiency, but where these efficiency costs 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 future work will apply the framework and strategy we have developed

here to welfare analysis in other insurance settings.

We believe that the approach we propose in this paper is likely to be broadly applicable (and that this is one of its attractions). It requires three essential data elements: insurance options and choices, subsequent risk realization, and variation in prices that is exogenous to demand and to insurable costs. These data requirements are not particularly stringent. Researchers have already demonstrated considerable success in obtaining insurance company data that contain the first two essential elements in a wide range of different insurance markets.[49] Indeed, a nice feature of welfare analysis in insurance markets is that cost data are much easier to obtain than in many other product markets, since they involve information on accident occurrences or insurance claims, rather than insight into the underlying production function of the firm. Further reducing the data hurdles, our approach can be implemented using aggregate (as opposed to individual) level data, as long as the coverage rules are simple enough that the cost data can be constructed from aggregate data on risk realization.[50]

The near-ubiquitous regulation of insurance markets provide numerous instances of the third essential data element: the exogenous pricing variation that is essential for estimating the demand and cost curves. Changes in state regulations of private insurance markets create variation in the prices charged to different individuals at a point in time as well as over time.[51] Tax policy is another useful source of pricing variation; for example, a large literature has documented (and used) the substantial variation across space and time in the tax subsidy for employer-provided health insurance and hence the price that individual employees face for this benefit.[52] Beyond the myriad opportunities provided by public policy, researchers have also found useful pricing variation stemming from the idiosyncrasies of firm pricing behavior.[53] More generally, any of the standard instruments used in demand analysis - such as "cost shifters" to the production function for insurance or exogenous changes in market competition – can serve as the requisite pricing variation for welfare analysis in insurance markets.

A key issue is the validity of the pricing instruments. This can be evaluated in specific applications. Indeed, we see the transparency of the approach as one of its key attractions. It is also relatively straightforward to implement, and fairly general. As a result, comparisons of welfare estimates obtained by this approach across different settings may be quite informative. For example,

[49] Examples in property-casualty insurance markets include Chiappori et al. (2006) and Cohen and Einav (2007) for automobile insurance and Sydnor (2006) for home owner's insurance. In health- and life-related insurance markets, examples include Finkelstein and Poterba (2004) for annuities, Cutler and Reber (1998) and Eichner, McLellan, and Wise (1998) for health insurance, and Finkelstein and McGarry (2006) for long-term care insurance.

[50] This would be the case, for example, if one were analyzing a full coverage compared to a no coverage option, or if in all the contracts the insurer costs were a linear function of risk realization.

[51] See Buchmueller and DiNardo (2002) or Bundorf and Simon (2006) for examples in health insurance markets, and Blackmon and Zeckhauser (1991) for an example from automobile insurance.

[52] See Gruber (2002) for a review of this large literature.

[53] Examples include firm experimentation with their pricing policy (Cohen and Einav, 2007), discrete pricing policy changes (Adams, Einav, and Levin, forthcoming), out of equilibrium pricing decisions made by human resource managers for employee insurance benefits (Cutler and Reber, 1998), and the rules firms use to adjust individuals' prices in response to their prior claims experience (Abbring, Chiappori, and Pinquet, 2003; Israel, 2004).

they can be used to try to assess in which settings – e.g., particular insurance markets, products, or populations – the welfare gains from government intervention are more likely to be substantial (relative to the costs).

A final feature of our empirical approach to welfare analysis in insurance markets that deserves emphasis is that it does not require the researcher to make assumptions about the underlying nature of individuals' information or preferences that gives rise to the estimated demand and cost curves. As long as we are willing to use the individuals' revealed choices for welfare analysis, the precise source of the selection (for example, the role of unobserved preferences for insurance or private information about risk type) is not germane for analyzing the efficiency cost of the resultant selection, or the welfare consequences of public policies that change the price (for example, through mandating or subsidizing a particular policy). Since such modeling assumptions are often ad hoc, and may have non trivial effects on the welfare estimates, we view the ability to avoid them as a key feature of our proposed approach. It is not, however, without cost; it restricts our ability to analyze the welfare consequences of counterfactual policies to those that change the prices of existing products. Our approach is unable to shed light on the welfare consequences of introducing products that are not observed in the data. Analysis of such questions would require that we model the primitives underlying the revealed demand and cost curves; this is an important and useful complement to the empirical approach outlined in this paper.

6 References

Abbring, Jaap, Pierre-Andre Chiappori, and Jean Pinquet (2003), "Moral hazard and Dynamic Insurance Data." *Journal of the European Economic Association* 1(4): 767-820.

Adams, William, Liran Einav, and Jonathan Levin (forthcoming), "Liquidity Constraints and Imperfect Information in Subprime Lending," *American Economic Review*, forthcoming, available at http://www.aeaweb.org/aer/contents/accepted.html.

Akerlof, George (1970), "The market for 'lemons': quality uncertainty and the market mechanism," *Quarterly Journal of Economics* 84(3): 488-500.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Price in Market Equilibrium," *Econometrica* 63: 841-890.

Blackmon, Glenn B. and Richard Zeckhauser (1991), "Mispriced Equity: Regulated Rates for Automobile Insurance in Masscahusetts" *American Economic Review Papers and Proceedings* 81(2): 65-69.

Buchmueller, Thomas C., and John E. DiNardo (2002), "Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut," *American Economic Review* 91(1): 280-294.

Bundorf, Kate M., and Kosali Simon (2006), "The Impact of Rate Regulation on Access to Supplemental Health Insurance," *American Economic Review Papers and Proceedings* 96(2):.67-71.

Bundorf, Kate M., Jonathan Levin, and Neale Mahoney (2008), "Pricing, Matching and Efficiency in Health Plan Choice," mimeo, Stanford University.

Busch, Susan H., Colleen L. Barry, Sally J. Vegso, Jody L. Sindelar and Mark R. Cullen (2006), "Effects of a Cost-Sharing Exemption on Use of Preventive Services At One Large Employer," *Health Affairs* 25(6): 1529-1536.

Carlin, Caroline, and Robert J. Town (2007), "Adverse Selection, Welfare and Optimal Pricing of Employer-Sponsored Health Plans," mimeo, University of Minnesota.

Cawley, John, and Tomas Philipson (1999), "An Empirical Examination of Information Barriers to Trade in Insurance," *American Economic Review* 89(4): 827-846.

Chernew, Michael, Gautam Gowrisankaran, and Dennis P. Scanlon (2008), "Learning and the Value of Information: Evidence from Health Plan Reports Cards," *Journal of Econometrics* 144(1): 156-174.

Chetty, Raj (2008), "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy* 116(2): 173-234.

Chiappori, Pierre-Andre, Bruno Jullien, Bernard Salanie, and Francois Salanie (2006), "Asymmetric Information in Insurance: General Testable Implications," *Rand Journal of Economics* 37(4): 783-798.

Chiappori, Pierre-Andre, and Bernard Salanie (2000), "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy* 108(1): 56-78.

Cohen, Alma, and Liran Einav (2007), "Estimating risk preferences from deductible choice," *American Economic Review* 97(3): 745-788.

Cutler, David M., and Sarah J. Reber (1998), "Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection," *Quarterly Journal of Economics* 113(2): 433-466.

Cutler, David M., Amy Finkelstein, and Kathleen McGarry (2008), "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance," *American Economic Review papers and proceedings* 98(2): 157-162.

Cutler, David M., and Richard J. Zeckhauser (2000), "The anatomy of health insurance," in A. J. Culyer and J. P. Newhouse (eds.), *Handbook of Health Economics*, edition 1, volume 1, chapter 11, 563-643, Elsevier.

De Meza, David, and David C. Webb (2001), "Advantageous selection in insurance markets," *Rand Journal of Economics* 32(2): 249-262.

Deaton, Agnus, and John Muellbauer (1980), "An Almost Ideal Demand System," *American Economic Review* 70: 312-326.

Eichner, Matthew J., Mark B. McLellan, and David A. Wise (1998), "Insurance or Self-Insurance?: Variation, Persistence, and Individual Health Accounts," in David Wise (ed.), *Inquiries in the Economics of Aging*, University of Chicago Press.

Einav, Liran, Amy Finkelstein, Robin McKnight and Mark R. Cullen (in progress), "Estimating the Effect of Cost-Sharing on Medical Spending," manuscript in progress.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf (2007), "The Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," NBER Working Paper 13228.

Fang, Hanming, Michael Keane, and Dan Silverman (2008), "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy* 116(2): 303-350.

Feldman, Roger D., and Bryan E. Dowd (1982), "Simulation of a Health Insurance Market with Adverse Selection," *Operation Research* 30(6): 1027-1042.

Finkelstein, Amy, and Kathleen McGarry (2006), "Multiple dimensions of private information: evidence from the long-term care insurance market," *American Economic Review* 96(4): 938-958.

Finkelstein, Amy, and James Poterba (2004), "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market," *Journal of Political Economy* 112(1): 193-208.

Fronstin, Paul (2003), "Sources of Health Insurance and Characteristics of the Uninsured: Analysis of the March 2003 Current Population Survey," Employee Benefit Research Institute Issue Brief 264, available at http://www.ebri.org/pdf/briefspdf/EBRI_IB_10a-20061.pdf.

Greenwald, Bruce C., and Joseph E. Stiglitz (1986), "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics* 101(2), 229-264.

Gruber, Jonathan (1994), "The Incidence of Mandated Maternity Benefits," *American Economic Review* 84(3): 622-641.

Gruber, Jonathan (2002), "Taxes and Health Insurance," James Poterba (ed.), *Tax Policy and the Economy,* Volume 16: 37-16. MIT Press, Cambridge, MA.

Hausman, Jerry (1997), "Valuation of New Goods Under Perfect and Imperfect Competition," in Bresnahan and Gordon (eds.), *The Economics of New Goods*, NBER Studies in Income and Wealth 58: 209-237.

Israel, Mark (2004), "Do We Drive More Safely When Accidents Are More Expensive? Identifying Moral Hazard from Experience Rating Schemes," unpublished mimeo, available at http://www.wcas.northwestern.edu/csio/Conferences/CSIO-IDEI-2004/moral_exp_paper1.pdf.

Lancaster, Kelvin J.(1966), "A New Approach to Consumer Theory," *Journal of Political Economy* 74(2), 132-157.

Lustig, Joshua D. (2007), "The Welfare Effects of Adverse Selection in Privatized Medicare," mimeo, Yale University.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, Arleen Leibowitz, and M. Susan Marquis (1987), "Health insurance and the demand for medical care: Evidence from a randomized experiment," *American Economic Review* 77(3): 251-277.

Poterba, James (1996), "Government Intervention in the Markets for Education and Health Care: How and Why?" in Victor Fuchs (ed.), *Individual and Social Responsibility*, University of Chicago Press.

Rothschild, Michael, and Joseph E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics* 90: 630-649.

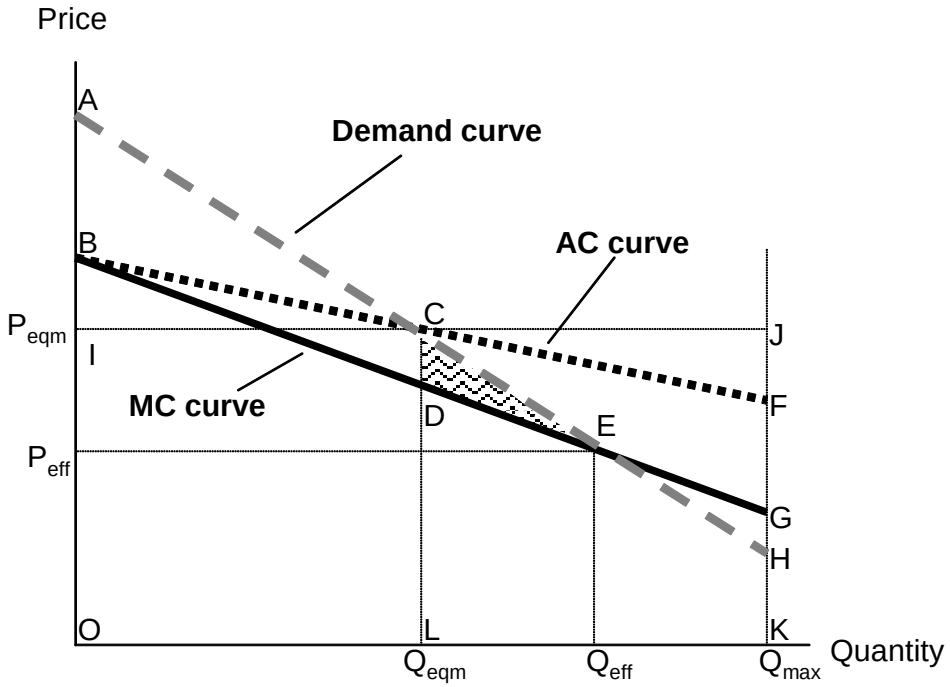
Stiglitz, Joseph E. (1987), "The Causes and Consequences of the Dependence of Quality on Price," *Journal of Economics Literature* 25(1): 1-48.

Sydnor, Justin (2006), "Sweating the Small Stuff: The Demand for Low Deductibles in Homeowners Insurance," mimeo, Case Western University, available at http://www.kellogg.northwestern.edu/

finance/faculty/seminars/Sydnor.pdf.

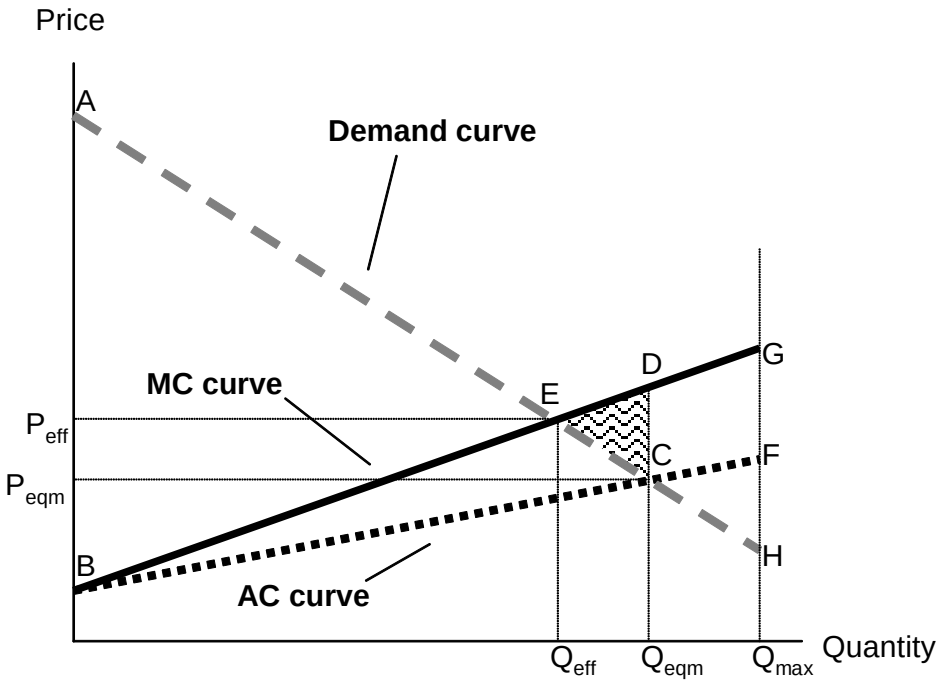
Wilson, Charles (1980), "The Nature of Equilibrium in Markets with Adverse Selection," *Bell Journal of Economics* 11(1): 108-130.

Figure 1: Efficiency cost of adverse selection



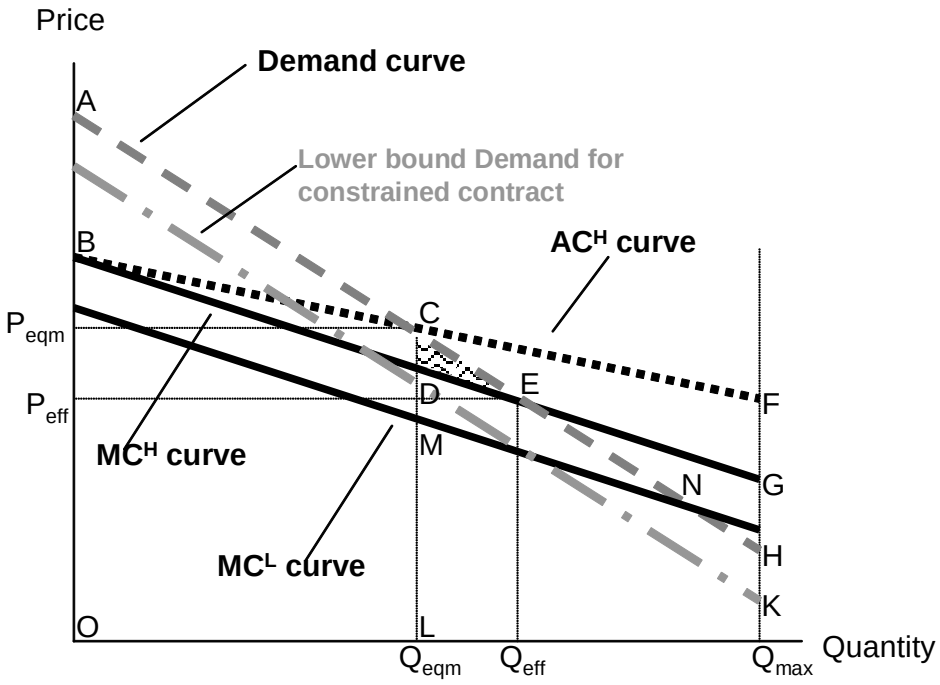
This figure illustrates the theoretical efficiency cost of adverse selection. It depicts a situation of adverse selection because the marginal cost curve is downward sloping (i.e. increasing in price, decreasing in quantity), indicating that the people who have the highest willingness to pay also have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

Figure 2: Efficiency cost of advantageous selection



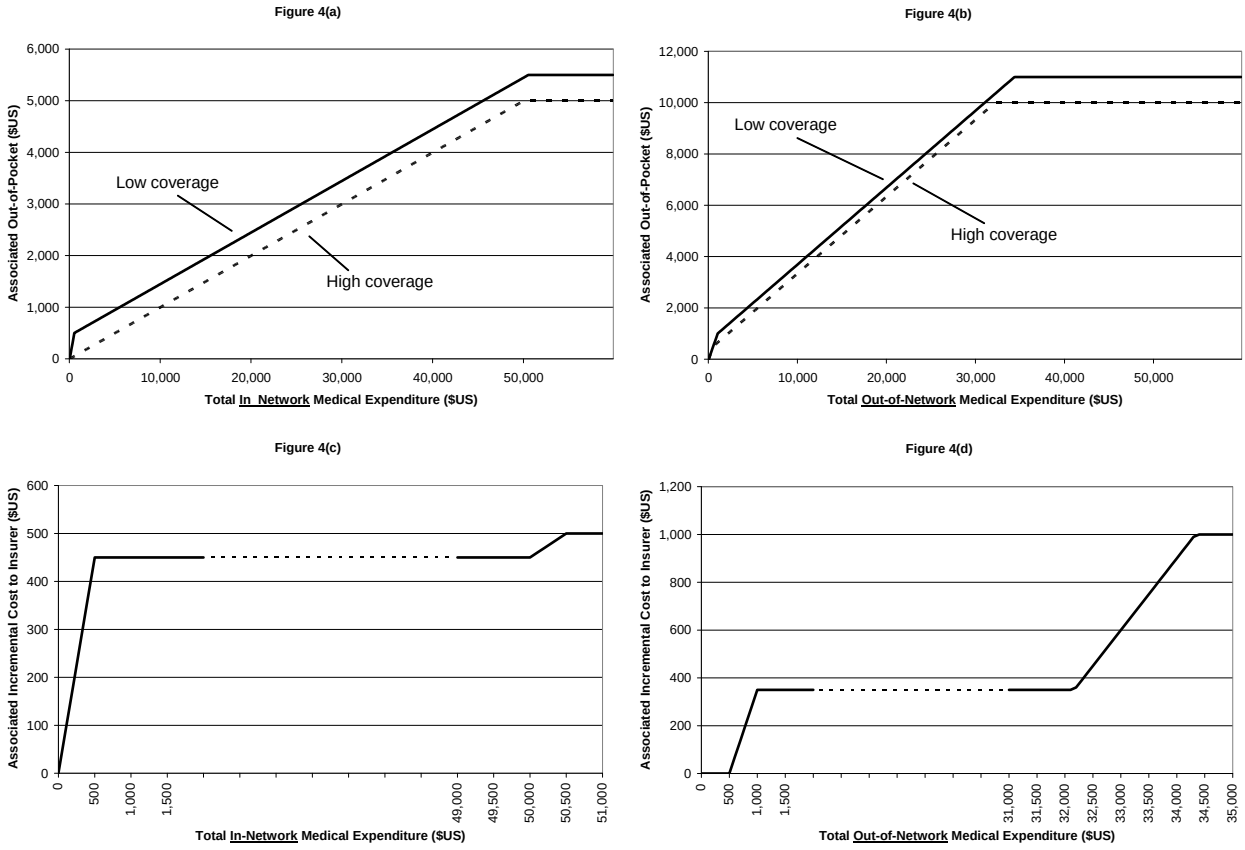
This figure illustrates the theoretical efficiency cost of advantageous selection. It depicts a situation of advantageous selection because the marginal cost curve is upward sloping, indicating that the people who have the highest willingness to pay have the lowest expected cost to the insurer. Competitive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

Figure 3: Moral hazard



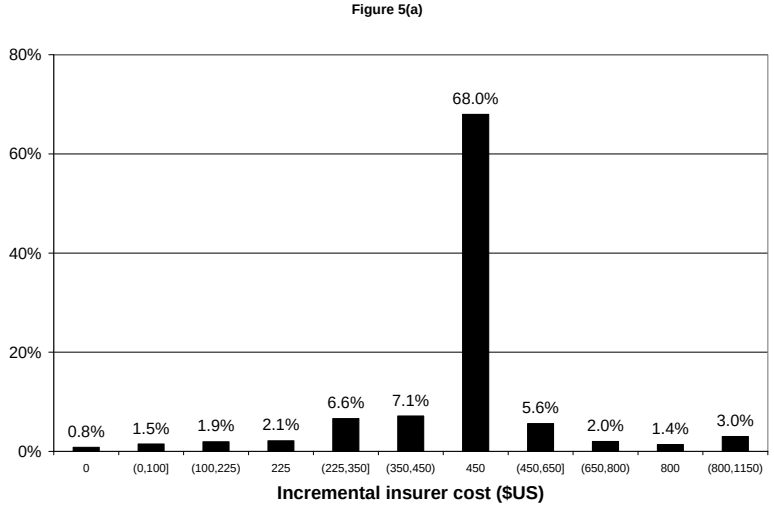
This figure extends Figure 1 to illustrate moral hazard; it shows a special case of moral hazard with parallel MC^L and MC^H curves. In the presence of moral hazard, the MC^L curve lies strictly below the MC^H curve. The (point-by-point) vertical distance between MC^H and MC^L provides a measure of moral hazard. The efficiency cost of adverse selection still remains the area CDE (as in Figure 1).

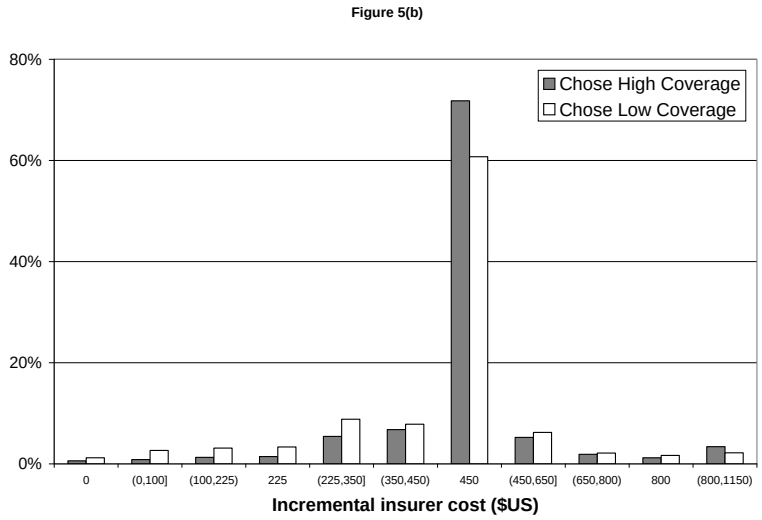
Figure 4: Description of the High and Low coverages



Figures 4(a) and 4(b) present the main features of the High (dashed) and Low (solid) coverages offered by the company, which is based on a deductible and an out-of-pocket maximum. Figures 4(c) and 4(d) present the corresponding cost difference to the insurer by providing the High coverage instead of the Low coverage, for a given level of medical expenditure; in other words, Figures 4(c) and 4(d) illustrate $c_i(m)$. Figures 4(a) and 4(c) describe the rules for in-network medical spending (deductibles of \$0 and \$500, and out-of-pocket maximums of \$5,000 and \$5,500 for High and Low coverage, respectively), and Figures 4(b) and 4(d) describe the rules for out-of-network medical spending (deductibles of \$500 and \$1000, and out-of-pocket maximums of \$10,000 and \$11,000 for High and Low coverage respectively). Coinsurance rates for both High and Low coverage are 10 percent (in network) and 30 percent (out of network). There is no interaction between the in-network and out-of-network coverages (i.e. each deductible and out of pocket maximum must be satisfied separately). The rules graphed in the figures and described in the foregoing notes are for any coverage that covers multiple individuals (i.e. family coverage, employee plus children coverage, or employee plus spouse coverage). Employees who purchase "employee only" coverage face deductibles and out of pocket maxima that are one-half of what is graphed in the figure (and described in the forgoing notes).

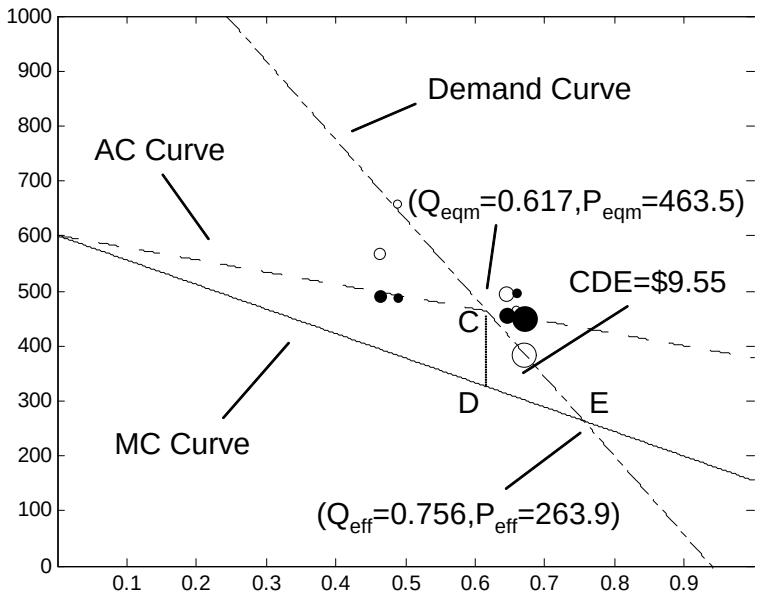
Figure 5: The distribution of the insurer's incremental costs (c_i)





This figure presents the distribution of the incremental insurer cost (c_i). Figure 5(a) does so for all 3,779 employees in our baseline sample (Table 1, column 5), while Figure 5(b) presents this separately for employees in High and Low coverages (2,465 and 1,314 employees, respectively). Note that the distribution has several mass points which are driven by the kinked formula of the coverages (Figure 4). 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.

Figure 6: Efficiency cost of adverse selection - empirical analog



This figure is the empirical analog of the theoretical Figure 1. The demand curve and AC curve are graphed using the point estimates of our baseline specification (Table 4). The MC curve is implied by the other two curves, as in equation (16). The circles represent the actual data points (Table 3) 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 3 with only 7 observations (although it is included in the estimation). We label points C, D, and E, that correspond to the theoretical analog in Figure 1, and report some important implied estimates (of the equilibrium and efficient points, as well as the welfare cost of adverse selection).

Table 1: Summary statistics

	2004 Company Data					March 2005 CPS		
	All employees	Only salaried workers	Only salaried workers with new benefit design	Col. (3) limited to only workers who chose High or Low	Col. (4) limited to workers with family coverage	All full time workers	Only in manufacturing	White collar workers in manufacturing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Individuals	36,814	11,964	11,325	7,263	3,779	83,118	11,178	4,688
Fraction Male	0.78	0.73	0.73	0.77	0.86	0.58	0.70	0.64
Fraction White	0.77	0.87	0.86	0.86	0.86	0.82	0.82	0.86
Fraction unionized	0.33	0.00	0.00	0.00	0.00	0.13	0.14	0.04
Age								
Mean	44.24	44.51	44.50	45.17	42.66	41.39	42.13	42.87
Std. Deviation	9.86	9.22	9.21	9.12	7.22	12.33	11.45	10.88
Median	45	45	45	46	43	41	42	43
Tenure with company (years)								
Mean	13.23	13.26	13.23	13.69	12.70	n/a	n/a	n/a
Std. Deviation	10.28	9.95	9.96	10.01	8.93	n/a	n/a	n/a
Median	11	12	12	13	12	n/a	n/a	n/a
Annual Salary (current $US)								
Mean	53,103	71,622	72,821	74,017	80,999	41,869	46,195	63,157
Std. Deviation	47,642	77,936	79,373	91,530	112,790	47,955	45,435	58,072
Median	47,283	60,484	61,433	61,822	66,335	32,000	35,000	50,000

Columns 1 to 5 present summary statistics for different cuts of the 2004 employees at the company. Column 1 presents statistics for all active employees in our sample, column 2 for salaried workers only. Column 3 looks at a slightly smaller group of salaried employees who faced the new benefit design, and column 4 further restricts attention to salaried employees who chose Low or High coverage (who are the primary focus of our analysis). Column 5 further limits the analysis to those who chose family coverage; this sample is used to generate our baseline estimates. For comparison, columns 6 to 8 present summary statistics for workers employed full time (defined as those who on average worked 35 or more hours per week in the previous year) in the March 2005 CPS. Column 6 shows all full time workers, column 7 shows all full time workers in manufacturing industries, and column 8 shows all full time white collar workers (defined based on occupation codes) in manufacturing industries; in these three columns we use CPS sampling weights ("earning weights" for the union variable, and "person weights" for all others).

Table 2: Assessing the exogeneity of the price variation

	Faced lowest relative price (2,939 workers) (1)	Faced higher relative prices (840 workers) (2)	Difference (3)	Coefficient (4)	p-value (5)
Age (Mean)	42.74	42.40	0.33	-0.245	0.31
Tenure (Mean)	13.02	11.63	1.39	-0.565	0.08
Fraction Male	0.862	0.852	0.009	1.268	0.79
Fraction White	0.874	0.825	0.049	-6.998	0.40
Log(Annual Salary) (Mean)	11.16	11.05	0.11	-8.612	0.17
Spouse Age (Mean)	41.37	41.05	0.32	-0.200	0.41
Number of covered family members (Mean)	4.14	4.07	0.07	-1.400	0.36
Age of youngest covered child (Mean)	9.81	9.41	0.40	-0.3	0.26
Log(2003 Medical Spending + 1)[a]					
All	8.13	7.79	0.32	-2.100	0.15
In most common 2003 plan	8.21	8.08	0.13	-1.700	0.08

The table reports average differences in covariates (shown in the left column) across workers who face different relative prices for the higher coverage option. Sample is limited to the 3,779 salaried workers with family coverage who choose High or Low coverage (Table 1, column 5). The worker 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-quarters of employees who faced the lowest relative price ($384; see Table 3) and the remaining one quarter who face one of the five higher relative prices ($466 to $659; see Table 3). 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 the workers in the sample who were in the data in 2003 (2,602 and 658 workers in columns 1 and 2, respectively), and for all the workers who were in the data in 2003 in the most common 2003 health insurance plan (2,284 and 523 workers in columns 1 and 2, respectively). The latter attempts to avoid potential differences in spending arising from moral hazard effects of different 2003 coverages.

Table 3: The effect of price on demand and costs

Relative Price	Number of Obs.	Fraction chose High Coverage	Average Incremental Cost	
			High Coverage	Low Coverage
(1)	(2)	(3)	(4)	(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

The table presents the raw data underlying our estimates of the demand and cost curves in the baseline specification. Sample is limited to our baseline sample of salaried workers with family coverage who chose High or Low coverage (Table 1, column 5). All individuals face one of six different relative prices, each represented by a row in the table. Column 2 gives the number of employees facing each price, and column 3 reports the fraction of them who chose High coverage. Columns 4 and 5 report (for High coverage and Low coverage individuals, respectively) the average incremental costs to the insurer of covering the individuals with High coverage rather than with Low coverage, given the family's medical expenditures. The graphical analog to this table is presented by the circles shown in Figure 6.

Table 4: Baseline results

Dependent Variable (Sample)	1 if chose High (both High and Low) (1)	Incremental Cost (only High) (2)
Relative Price of High ($US)	-0.00070	0.15524
	(0.00032)	(0.06388)
	[0.034]	[0.021]
Constant	0.940	391.690
	(0.123)	(26.789)
	[0.000]	[0.000]
Mean Dependent Variable	0.652	455.341
Number of Observations	3,779	2,465
R-Squared	0.008	0.005

The table reports the results from our baseline specification. Sample is limited to salaried workers with family coverage. Column 1 reports the results from estimating the linear demand $D = \alpha + \beta p$ (equation (14)) on the sample of employees who choose High or Low coverage; D is an indicator variable for whether the employee chose High coverage (as opposed to Low coverage). Column 2 reports the results from estimating the linear cost equation $c = \gamma + \delta p$ (equation (15)) on the sample of individuals who choose High coverage; c is the incremental costs to the insurer of covering a given employee's (and covered dependents') medical expenditures with High coverage rather than Low coverage. The price variable (p) is the incremental premium to the employee for High (as opposed to Low) coverage. 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 Table 5.

Table 5: Robustness

Panel A: Welfare estimates from different specifications

	Competitive Equilibrium		Efficient Allocation		Welfare cost of Adverse Selection			
	Q	P	Q	P	In dollar terms (per market participant)[a]	Relative to social cost of efficient subsidy[b]	Relative to welfare cost of High coverage mandate[c]	Relative to total achievable welfare[d]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 Baseline (family coverage, no state fixed-effects)	0.617	463.51	0.756	263.94	9.55	21.1%	32.4%	3.4%
Robustness to demand estimates								
2 Probit demand	0.619	463.59	0.790	187.85	11.32	17.3%	31.6%	3.5%
3 Linear demand, constrained to go through (Q,P)=(1,$0)	0.612	463.56	0.750	299.04	7.81	21.1%	30.2%	3.4%
4 Linear demand, constrained to go through (Q,P)=(0,$800)	0.562	463.59	0.688	387.90	3.30	21.1%	16.4%	3.4%
5 Quadratic demand, constrained to go through (1,$0) and (0,$800)	0.587	463.58	0.738	343.51	5.00	18.8%	45.6%	4.5%
Robustness to tax subsidy								
6 Baseline specification, but accouting for pre-tax premiums	0.389	514.49	0.567	348.53	7.71	27.3%	16.8%	9.8%
Robustness to sample and source of variation								
7 State fixed-effects included (in both demand and cost regressions)	0.622	460.16	0.699	341.40	3.65	14.6%	6.5%	1.2%
8 State fixed-effects and demographics included (in both regressions)	0.641	440.00	0.724	306.67	4.42	15.3%	9.2%	1.3%
9 All coverage tiers, no state fixed-effects[e]	0.593	434.20	0.704	244.83	7.67	19.2%	14.2%	2.5%

Panel B: Parameter estimates from different specifications

	Demand Equation				Average Cost Equation			
	alpha		beta		gamma		delta	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
1 Baseline (family coverage, no state fixed-effects)	0.940	(0.123)	-0.00070	(0.00032)	391.7	(26.8)	0.155	(0.064)
Robustness to demand estimates								
2 Probit demand	1.149	(0.316)	-0.00183	(0.00080)	------ same as Baseline ------			
3 Linear demand, constrained to go through (Q,P)=(1,$0)	1.000	(imposed)	-0.00084	(0.00005)	------ same as Baseline ------			
4 Linear demand, constrained to go through (Q,P)=(0,$800)	1.333	(imposed)	-0.00167	(0.00005)	------ same as Baseline ------			
5 Quadratic demand, constrained to go through (1,$0) and (0,$800)[f]	1.000	(imposed)	-0.00039 -1.1E-06	(imposed) (1.2E-07)	------ same as Baseline ------			
Robustness to tax subsidy								
6 Baseline specification, but accouting for pre-tax premiums	0.940	(0.123)	-0.00107	(0.00048)	391.7	(26.8)	0.239	(0.098)
Robustness to sample and source of variation								
7 State fixed-effects included (in both regressions)	0.919	(0.167)	-0.00065	(0.00040)	414.8	(37.0)	0.099	(0.090)
8 State fixed-effects and demographics included (in both regressions)	0.917	(0.170)	-0.00063	(0.00040)	394.5	(36.9)	0.104	(0.091)
9 All coverage tiers, no state fixed-effects[e]	0.848	(0.109)	-0.00059	(0.00032)	374.8	(22.8)	0.137	(0.062)

Table reports results from alternative specifications. Panel B reports parameter estimates, and Panel A reports the (corresponding) implications for welfare analysis. Row 1 reports the baseline specification (Table 4), rows 2-5 report specifications that change the functional form of demand. Row 6 re-estimates the baseline specification with the price in both the demand and cost equation multiplied by 0.65 (one minus the average marginal tax rate in the sample). Row 7 includes state fixed effects in both the demand and cost equations, and row 8 also controls for employee characteristics (see Table 2). Row 9 increases the sample to include employees in all four coverage tiers. Standard errors (in parentheses) allow for an arbitrary variance-covariance matrix within each state.

[a] Graphically, this is the area of triangle CDE (see Figure 1).

[b] This is triangle CDE divided by $0.3Q_{eff}(P_{eq}-P_{eff})$.

[c] Graphically, this is the area of triangle CDE divided by the area of triangle EGH (see Figure 1).

[d] Graphically, this is the area of triangle CDE divided by the area of triangle ABE (see Figure 1).

[e] N=7,263 for demand analysis, 4,622 for cost analysis; mean dependent variables are 0.64 (D) and \$424 (c), respectively. We include (de-meaned) indicator variables for the coverage tier in both the demand and cost equations (not shown); we multiply p and c by two for employees in the "employee only" coverage tier.

[f] In the quadratic demand specification, the top reported coefficient of beta is the coefficient on the linear term, while the second is the coefficient on the quadratic term.

Table 6: Potential sample selection

Dependent variable:	1 if "outside good" was chosen, 0 otherwise		
	"Outside Good" does not include "opt out"		"Outside good" does include "opt out"
	Family coverage tier only	All coverage tiers	All coverage tiers
	(1)	(2)	(3)
Relative price	-0.0000093	-0.000021	0.000002
	(0.00035)	(0.00040)	(0.000003)
	[0.98]	[0.96]	[0.66]
Constant	0.287	0.292	0.296
	(0.1580)	(0.1150)	(0.1580)
	[0.08]	[0.02]	[0.07]
Mean dependent variable	0.283	0.300	0.359
Number of obs.	5,271	10,386	11,325

The table reports results of estimating a variant of the demand equation shown in equation (14). The dependent variable is an indicator variable that takes the value of 1 if the employee chose any of the "outside options" and 0 if the employee chose either the High or Low coverage PPO. The "relative price" variable is, as in Table 4, the relative price of the High coverage compared to the Low coverage. In columns 1 and 2 the "outside good" includes two lower coverage PPOs, a Health Reimbursement Account PPO, and an HMO. The sample in column 1 is limited to family coverage. The sample in column 2 includes all coverage tiers; we therefore include (de-meaned) indicator variables for the coverage tier (not shown) and multiply the price variable p by two for employees in the "employee only" coverage tier. In column 3 the "outside good" definition is expanded to also include employees who opt out of coverage; since coverage tier is not known for these employees, we include all employees regardless of coverage tier and do not include indicator variables for coverage tier. We define the price variable as the relative price of High coverage compared to Low coverage if the employee chose the family coverage tier (regardless of the actual tier chosen, if known). 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].