

NBER WORKING PAPER SERIES

SOCIAL INSURANCE:
CONNECTING THEORY TO DATA

Raj Chetty
Amy Finkelstein

Working Paper 18433
<http://www.nber.org/papers/w18433>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2012

We are grateful to Alan Auerbach, Liran Einav, Hilary Hoynes, Emmanuel Saez, and participants at the Handbook of Public Economics conference held at UC-Berkeley for helpful comments. Shelby Lin, Heather Sarsons, and Michael Stepner provided excellent research assistance. We gratefully acknowledge support from the NIA R01 AG032449 (Finkelstein) and the National Science Foundation SES 0645396 (Chetty). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2012 by Raj Chetty and Amy Finkelstein. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Social Insurance: Connecting Theory to Data
Raj Chetty and Amy Finkelstein
NBER Working Paper No. 18433
October 2012, Revised January 2013, Revised July 2014
JEL No. H5

ABSTRACT

We survey the literature on social insurance, focusing on recent work that has connected theory to evidence to make quantitative statements about welfare and optimal policy. Our review contains two parts. We first discuss motives for government intervention in private insurance markets, focusing primarily on selection. We review the original theoretical arguments for government intervention in the presence of adverse selection, and describe how recent work has refined and challenged the conclusions drawn from early theoretical models. We then describe empirical work that tests for selection in insurance markets, documents the welfare costs of this selection, and analyzes the welfare consequences of potential public policy interventions. In the second part of the paper, we review work on optimal social insurance policies. We discuss formulas for the optimal level of insurance benefits in terms of empirically estimable parameters. We then consider the consequences of relaxing the key assumptions underlying these formulas, e.g., by allowing for fiscal externalities or behavioral biases. We also summarize recent work on other dimensions of optimal policy, including mandated savings accounts and the optimal path of benefits. Finally, we discuss the key challenges that remain in understanding the optimal design of social insurance policies.

Raj Chetty
Department of Economics
Harvard University
1805 Cambridge St.
Cambridge, MA 02138
and NBER
chetty@fas.harvard.edu

Amy Finkelstein
Department of Economics
MIT E52-383B
50 Memorial Drive
Cambridge, MA 02142
and NBER
afink@mit.edu

1 Introduction

Over the last century, social insurance – government intervention in providing insurance against adverse shocks to individuals – has emerged as one of the major functions of government in developed countries.¹ Social insurance programs began by providing limited coverage for risks such as injury at work and unemployment (Fishback and Kantor 1998, Baicker, Goldin, and Katz 1998). Today, governments provide substantial insurance for a broad range of risks, including health (Medicare and Medicaid in the U.S.), disability and retirement (the Old Age, Survivors, and Disability insurance program), work injury (Worker’s Compensation), and unemployment (Unemployment Insurance).² In the United States, expenditures on social insurance have risen from less than 10 percent of the federal government’s budget in the early 1950s to almost 60 percent today and continue to grow rapidly (Gruber 2009). Social insurance expenditures are now a defining characteristic of modern developed economies. The fraction of GDP devoted to social insurance increases sharply with GDP per capita (Figure 1).

Academic research on social insurance policies has grown alongside the expansion of these programs. Research on social insurance has addressed two broad questions. First, when should the government intervene in private insurance markets? The standard set of rationales includes private market failures, income redistribution, and paternalism. More recently, a growing empirical literature has sought to quantify the importance of these motives for government intervention. Much of this literature has focused on one particular market failure that can provide a rationale for social insurance: adverse selection due to asymmetric information. Second, if the government chooses to intervene, what is the optimal way to do so? The key issue here is that expanding social insurance creates moral hazard by distorting incentives. The literature on optimal policies seeks to identify the policies that maximize welfare, trading off the distortionary costs of social insurance programs with the benefits they provide in reducing exposure to risk. This literature has analyzed several dimensions of social insurance policies, ranging from the optimal level of benefits to whether the optimal tools are provision of liquidity (e.g., via loans) or state-contingent transfers.

Research on each of these two questions has traditionally been divided into two distinct methodological strands: a normative theoretical literature that focuses on welfare analysis and a positive empirical literature that documents the workings of private insurance markets or the impacts of social insurance programs. The limitation of this two-pronged approach is that the theoretical models do not by themselves offer quantitative answers to the key policy questions, while the descriptive empirical literature often has little to say about the welfare implications of its findings. For example, the rich theoretical literature on adverse selection has shown that private markets

¹We use the term “social insurance” to refer to government programs that transfer resources across states of nature after an individual is born rather than transfers of resources across individuals (e.g., through redistributive taxation). Transfers of resources across individuals – which effectively provide insurance behind the veil of ignorance – are discussed in the chapters on optimal taxation in this volume.

²See Social Security Administration (1997) for an excellent overview of modern social insurance programs in the United States. Krueger and Meyer (2002) also provide a description of many social insurance programs in the U.S., as well as a review of the empirical literature on their labor supply impacts.

may provide too little insurance in the presence of asymmetric information. A more recent empirical literature has documented that adverse selection does in fact exist in many private insurance markets. However, the empirical techniques developed to identify the existence of selection do not, by themselves, permit even qualitative comparisons of the welfare costs of selection across markets, let alone quantitative welfare statements. Similarly, a large theoretical literature has characterized the properties of optimal insurance contracts in the presence of moral hazard. In parallel, empirical work has documented the causal impacts of social insurance programs on a broad range of behaviors, ranging from job search to health expenditures. Again, however, the implications of estimates of parameters such as the elasticity of unemployment durations with respect to benefits for optimal policy were unclear from the initial empirical literature.

Over the past two decades, researchers have made considerable progress in connecting theoretical and empirical work on social insurance to make empirically grounded statements about welfare and optimal policy. For instance, recent work has shown how data on selection patterns in insurance markets can be used to quantify the welfare costs of adverse selection in models of asymmetric information. Similarly, researchers have developed new methods of mapping estimates of behavioral elasticities to statements about the optimal level of social insurance benefits.

In this survey, we provide an overview of some of the key advances in connecting theory to data in analyzing the welfare consequences of social insurance. In focusing on this goal, we deliberately do not provide a comprehensive survey of the literature on social insurance. We cover only a selected subset of the many theoretical studies that have advanced the literature. We also discuss only a small subset of the numerous empirical studies that have estimated relevant empirical parameters. Readers seeking a more detailed discussion of empirical evidence on behavioral responses to social insurance may refer to Krueger and Meyer (2002) and Cutler (2002).

We divide our review of the literature into two sections, corresponding to the two major questions discussed above. In Section 2, we discuss motives for government intervention in insurance markets. In Section 3, we discuss optimal policy design once the government has decided to intervene. This literature on optimal design of social insurance has proceeded mostly independently from the work on the economic motivations for social insurance. As a result, the two sections of the paper draw on fairly distinct literatures. Indeed, one limitation of existing work on optimal government policy is that it typically assumes away the existence of formal private insurance markets rather than considering optimal policy design in an environment with endogenous market failures. We conclude in Section 4 by discussing this as well as some of the other broad challenges that remain in going from the work we review here to statements about optimal policy design.

2 Motivations for Social Insurance

Research in public economics usually begins with the question of why the government might have a reason to intervene in a particular private market transaction. Only then can one move forward to consider potential forms of intervention and their consequences. Standard economic rationales for social insurance include redistribution, paternalism, and market failures (Diamond 1977). Within

this relatively broad remit, our focus here is quite narrow. Following much of the recent literature, we concentrate on the potential role for social insurance in ameliorating one particular type of private market failure, namely selection. We return at the end of this section to briefly comment on other potential rationales for social insurance and some of the existing empirical work on them.

Modern theoretical work on adverse selection in insurance markets dates to the seminal work of Akerlof (1970) and Rothschild and Stiglitz (1976), which introduced a key motivation for social insurance: the competitive private equilibrium may under-provide insurance, creating scope for welfare-improving government intervention. Relative to the rich theoretical literature, empirical work on adverse selection in insurance markets lagged decidedly behind for many years. Indeed, in awarding the 2001 Nobel Prize for the pioneering theoretical work on asymmetric information, the Nobel committee noted this paucity of empirical work (Bank of Sweden 2001).

Over the last decade or so, empirical research had made considerable progress in developing tools to identify whether asymmetric information exists in a given insurance market, as well as to begin to quantify the welfare costs of this asymmetric information and the welfare consequences of alternative public policy interventions. Some of the findings of this empirical work have suggested important refinements to the initial theory. In particular, a growing body of evidence suggests that in addition to heterogeneity in risk type, heterogeneity in preferences can be a quantitatively important determinant of demand for insurance. This is in contrast to the original theoretical literature on asymmetric information which focused on the potential for (unobserved) heterogeneity in risk type and assumed away the possibility of heterogeneity in preferences. Once one allows for heterogeneity in preferences in addition to risk type, the competitive equilibrium may look very different and the optimal policy intervention is no longer a priori obvious.

To summarize and discuss this empirical literature, we begin by presenting a highly stylized model and graphical framework that allow us to review the basic results of the standard theory and to describe their sensitivity to incorporating several “real world” features of insurance markets. The graphical framework provides a lens through which we discuss empirical work detecting whether selection exists and quantifying its welfare costs. Finally, we discuss some of the limitations of the work to date and some directions for further work.

2.1 Adverse Selection: Review of the basic theory

We structure our analysis using a simplified model of selection based on that presented in Einav, Finkelstein and Cullen (2010) and discussed further in Einav and Finkelstein (2011). We begin with the “textbook model” in which the qualitative results are unambiguous: adverse selection creates a welfare loss from under-provision of insurance, and public policy such as mandates can reach the efficient allocation and improve welfare. Even in this textbook case, however, the magnitudes of the welfare costs of adverse selection and the welfare gains from government intervention remain empirical questions. Moreover, these qualitative results can be reversed with the introduction of two important features of actual insurance markets: loads and preference heterogeneity. With loads, it is no longer necessarily efficient for all individuals to be insured in equilibrium, and

mandates can therefore reduce welfare in some cases. With preference heterogeneity, the market equilibrium may lead to over-insurance rather than under-insurance. Given the qualitative as well as quantitative uncertainty of the impact of selection and of government intervention, these naturally become empirical questions.

2.1.1 A stylized model

Setup and notation. A population of individuals chooses from two insurance contracts, one that offers high coverage (contract H) and one that offers less coverage (contract L). To further simplify the exposition, assume that contract L is no insurance and is available for free, and that contract H is full insurance. These are merely normalizations and it is straightforward to extend the analysis to partial coverage contracts or to more than two contracts (Einav, Finkelstein and Cullen 2010).

The key simplification we make is to fix the contract space, but allow the price of insurance to be determined endogenously. In other words, the set of contracts that insurance companies offer is determined exogenously, and the focus of the model is on how selection distorts the pricing of these existing contracts. The analysis is therefore in the spirit of Akerlof (1970) rather than Rothschild and Stiglitz (1976), who endogenize the level of coverage as well. This assumption greatly simplifies the analytical framework and makes it easier to both allow for multiple sources of heterogeneity across consumers and to illustrate some of the key insights and implications of selection models. However, it means that the analysis of the welfare consequences of selection or alternative possible government interventions is limited to the cost associated with inefficient pricing of a fixed set of contracts; it does not capture welfare loss that selection may create by distorting the set of contracts offered, which may be large in some settings. We return to this central issue below.

Define the population by a distribution $G(\zeta)$, where ζ is a vector of consumer characteristics. For our initial discussion of the “textbook case”, we will assume that these consumer characteristics ζ include only characteristics relating to their risk factors; later, we will relax this assumption and explore the implications of allowing for preference heterogeneity.

Denote the (relative) price of contract H by p , and denote by $v^H(\zeta_i, p)$ and $v^L(\zeta_i)$ consumer i 's (with characteristics ζ_i) utility from buying contracts H and L , respectively. Although not essential, it is natural to assume that $v^H(\zeta_i, p)$ is strictly decreasing in p and that $v^H(\zeta_i, p = 0) > v^L(\zeta_i)$. Finally, denote the expected monetary cost to the insurer associated with the insurable risk for individual i by $c(\zeta_i)$. For ease of exposition, we discuss the benchmark case in which there is no moral hazard; the cost c of insuring an individual does not depend on the contract chosen. Allowing for moral hazard does not fundamentally change the analysis, although it does complicate the presentation (Einav, Finkelstein and Cullen 2010). Of course, as we will discuss at length when we turn to the empirical work on selection in insurance markets, the potential presence of moral hazard as well as selection does pose important empirical challenges to the analysis of either one.

Demand for insurance. Assume that each individual makes a discrete choice of whether to buy insurance or not. Since there are only two available contracts and their associated coverages, demand is only a function of the (relative) price p . Assume that firms cannot offer different prices to

different individuals. To the extent that firms can make prices depend on observed characteristics, one should think of the foregoing analysis as applied to a set of individuals that only vary in unobserved (or unpriced) characteristics. Assume that if individuals choose to buy insurance they buy it at the lowest price offered, so it is sufficient to characterize demand for insurance as a function of the lowest price p .

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

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

and we assume that the underlying primitives imply that $D(p)$ is strictly decreasing 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). We further assume that when multiple firms set the same price, individuals who decide to purchase insurance at this price choose a firm randomly. In the “textbook case”, we assume that the only costs of providing contract H to individual i are the direct insurer claims $c(\zeta_i)$ that are paid out; later we will explore the implications of allowing for the possibility of loading factors, such as other administrative (production) costs of the insurance company.

The foregoing assumptions imply that the average (expected) cost curve in the market is given by

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

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

$$MC(p) = E(c(\zeta) | \pi(\zeta) = p). \quad (3)$$

In order to straightforwardly characterize equilibrium, we make two further simplifying assumptions. First, we assume that there exists a price \bar{p} such that $D(\bar{p}) > 0$ and $MC(p) < p$ for every $p > \bar{p}$. 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. 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 at most once. These assumptions guarantee the existence and uniqueness of an equilibrium. In particular, the equilibrium is characterized by the lowest break-even price, that is:

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

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. This is an attractive measure

of welfare because it is a money metric. Total surplus in the market is the sum of certainty equivalents for consumers and profits of firms. We ignore income effects associated with price changes. Note that price changes have no income effects if the utility function exhibits constant absolute risk aversion (CARA).

Denote by $e^H(\zeta_i)$ and $e^L(\zeta_i)$ 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 $\pi(\zeta_i) = e^H(\zeta_i) - e^L(\zeta_i) > 0$. We can write consumer welfare as

$$CS = \int [(e^H(\zeta) - p) 1(\pi(\zeta) \geq p) + e^L(\zeta) 1(\pi(\zeta) < p)] dG(\zeta) \quad (5)$$

and producer welfare as

$$PS = \int (p - c(\zeta)) 1(\pi(\zeta) \geq p) dG(\zeta). \quad (6)$$

Total welfare is

$$TS = CS + PS = \int [(e^H(\zeta) - c(\zeta)) 1(\pi(\zeta) \geq p) + e^L(\zeta) 1(\pi(\zeta) < p)] dG(\zeta). \quad (7)$$

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

$$\pi(\zeta_i) \geq c(\zeta_i). \quad (8)$$

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 to him the insurance.

2.1.2 The textbook case

Adverse selection equilibrium Figure 2 provides a graphical representation of the adverse selection insurance equilibrium for the “textbook case” we have just outlined. The relative price (or cost) of contract H is on the vertical axis. Quantity (i.e. share of individuals in the market with contract H) is on the horizontal axis; the maximum possible quantity is denoted by Q_{max} . The demand curve denotes the relative demand for contract H . Likewise, the average cost (AC) curve and marginal cost (MC) curve denote the average and marginal *incremental* costs to the insurer from coverage with contract H relative to contract L .

Because agents can only choose whether to purchase the contract or not, the market demand curve simply reflects the cumulative distribution of individuals’ willingness to pay for the contract. The difference between willingness to pay $\pi(\zeta)$ and $MC(\zeta)$ is the risk premium, and is positive for risk averse individuals.

Because of the “textbook” assumption that individuals are homogeneous in all features of their utility function – i.e. ζ_i includes only characteristics relating to one’s expected claims c_i – willingness to pay for insurance is increasing in risk type. This is the key feature of adverse selection: individuals who have the highest willingness to pay for insurance are those who, on average, have the highest expected costs. This is represented in Figure 2 by drawing a downward sloping MC curve. That is, marginal cost is increasing in price and decreasing in quantity. As the price falls, the marginal

individuals who select contract H have lower expected cost than infra-marginal individuals, leading to lower average costs.

The link between the demand and cost curve is arguably the most important distinction of insurance markets (or selection markets more generally) from traditional product markets. The shape of the cost curve is driven by the demand-side consumer selection. In most other contexts, the demand curve and the cost curve are independent objects; demand is determined by preferences and costs by the production technology. The distinguishing feature of selection markets is that the demand and cost curves are tightly linked since the individual's risk type not only affects demand but also directly determines cost.

As noted, the efficient allocation is to insure all individuals whose willingness to pay is at least as great as their expected cost of insuring them. In the textbook case, the risk premium is always positive, since by assumption all individuals are risk averse and there are no other market frictions. As a result, the demand curve is always above the MC curve and, as shown in Figure 2, it is therefore efficient for all individuals to be insured ($Q_{eff} = Q_{max}$). The welfare loss from not insuring a given individual is simply the risk premium of that individual, or the vertical difference between the demand and MC curves.

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 2. The equilibrium share of individuals who buy contract H is Q_{eqm} (where the AC curve intersects the demand curve), while the efficient number is $Q_{eff} > Q_{eqm}$; in general, the efficient allocation Q_{eff} is determined where the MC curve intersects the demand curve, which in the textbook case is never (unless there are people with risk probability of zero or who are risk neutral). The fundamental inefficiency created by adverse selection arises because the efficient allocation is determined by the relationship between *marginal* cost and demand, but the equilibrium allocation is determined by the relationship between *average* cost and demand.

The welfare loss due to adverse selection arises from the lost consumer surplus (the risk premium) of those individuals who remain inefficiently uninsured in the competitive equilibrium. In Figure 2, these are the individuals whose willingness to pay is less than the average cost of the insured population, P_{eqm} . Integrating over all these individuals' risk premia, the welfare loss from adverse selection is given by the area of the "dead-weight loss" trapezoid CDEF.

The amount of under-insurance generated by adverse selection, and its associated welfare loss, can vary greatly in this environment. As illustrated graphically in Einav and Finkelstein (2011), the efficient allocation can be achieved despite a downward sloping marginal cost curve if average costs always lie below demand. In contrast, if average costs always lie above demand, the private market will unravel completely, with no insurance in equilibrium.

Public policy in the textbook case One can use the graphical framework in Figure 2 to evaluate the welfare consequences of common public policy interventions in insurance markets that alter the insurance allocation. The comparative advantage of the public sector over the private sector is that it can directly manipulate either the equilibrium quantity of insurance (through mandates) or the equilibrium price of insurance (through either tax/subsidy policy or regulation of insurance company pricing). We briefly discuss each in turn.

Mandates. The canonical solution to the inefficiency created by adverse selection is to mandate that everyone purchase insurance, a solution emphasized as early as Akerlof (1970). In the textbook setting, mandates produce the efficient outcome in which everyone has insurance. However, the magnitude of the welfare benefit produced by an insurance purchase requirement varies depending on the specifics of the market since, as noted, the amount of under-insurance produced by adverse selection in equilibrium can itself vary greatly.

Tax subsidies. Another commonly discussed policy remedy for adverse selection is to subsidize insurance coverage. Indeed, adverse selection in private health insurance markets is often cited as an economic rationale for the tax subsidy to employer provided health insurance, which is the single largest federal tax expenditure. We can again use Figure 2 to illustrate. Consider, for example, a subsidy toward the price of coverage. This would shift demand out, leading to a higher equilibrium quantity and less under-insurance. The gross welfare loss would still be associated with the area between the original (pre-subsidy) demand curve and the MC curve, and would therefore unambiguously decline with any positive subsidy. A large enough subsidy (greater than the line segment GE in Figure 2) would lead to the efficient outcome, with everybody insured.

Of course, the net welfare gain from public insurance subsidies will be lower than the gross welfare gain due to the marginal cost of the public funds that must be raised to finance the subsidy; this may be quite large since the subsidy must be paid on all the inframarginal consumers as well as the marginal ones. Given a non-zero deadweight cost of public funds, the welfare maximizing subsidy would not attempt to achieve the efficient allocation. It is possible that the welfare maximizing subsidy could be zero. That is, starting from the competitive allocation (point C), a marginal dollar of subsidy may not be welfare enhancing. Although given the equilibrium distortion the welfare gain will be first order, the welfare cost of raising funds to cover the subsidy is first-order as well. Hence, the benefits of subsidies are again an empirical question.

Restrictions on characteristic-based pricing. A final common form of public policy intervention is regulation that imposes restrictions on the characteristics of consumers over which firms can price discriminate. Some regulations require “community rates” that are uniform across all individuals, while others prohibit insurance companies from making prices contingent on certain observable risk factors, such as race or gender. For concreteness, consider the case of a regulation that prohibits pricing on the basis of gender. Recall that Figure 2 can be interpreted as applying to a group of individuals who must be given the same price by the insurance company. When pricing based on gender is prohibited, males and females are pooled into the same market, with a variant of Figure 2 describing that market. When pricing on gender is allowed, there are now two distinct insurance markets – described by two distinct versions of Figure 2 – one for women and one for men, each of

which can be analyzed separately. A central issue for welfare analysis is whether, when insurance companies are allowed to price on gender, consumers still have residual private information about their expected costs. If they do not, then the insurance market within each gender-specific segment of the market will exhibit a constant (flat) MC curve, and the equilibrium in each market will be efficient. In this case, policies that restrict pricing on gender unambiguously reduce welfare because they create adverse selection where none existed before. However, in the more likely case that individuals have some residual private information about their risk that is not captured by their gender, each gender-specific market segment would look qualitatively the same as Figure 2 (with downward sloping MC and AC curves). In such cases, the welfare implications of restricting pricing on gender could go in either direction. Depending on the shape and position of the gender-specific demand and cost curves relative to the gender-pooled ones, the sum of the areas of the deadweight loss trapezoids in the gender-specific markets could be larger or smaller than the area of the single deadweight loss trapezoid in the gender-pooled market.³ See Einav and Finkelstein (2011) for a numerical illustrative example.

Comment: Pareto improvements. It is important to note that while various policies may be able to increase efficiency or even produce the efficient outcome – such as mandates – they are not, in this environment, Pareto improving. Consider for concreteness the case of mandates. The insurance provider (be it the government or the private market) must break even in equilibrium, and therefore the cost of providing the insurance must be recouped. The total cost is equal to the market size (Q_{\max}) times the average cost of insurance provision to Q_{\max} individuals, which is given by point G. Suppose the government uses average cost pricing, effectively issuing a lump sum tax on individuals equal to the average cost of insuring all individuals (given by the vertical distance at point G). While this policy achieves the efficient allocation, those whose willingness to pay is less than the price level at point G are made strictly worse off. Other financing mechanisms may generate welfare gains for a larger set of individuals, but assuming that the government does not observe the private information about individuals’ costs, the government – like the private sector – cannot price insurance to individuals based on their (privately known) marginal cost.

The inability for mandates to produce a Pareto-improvement are a direct consequence of the Akerloviaan modeling framework which has fixed the contract space. Some models that endogenize the contract offers generate Pareto-improving mandates (e.g., Wilson 1977) or Pareto-improving tax-transfer schemes (Rothschild and Stiglitz 1976). Crocker and Snow (1985) discuss the assumptions under which the decentralized equilibrium is constrained Pareto efficient in models with endogenous contracts.⁴

³This analysis focuses only on static welfare considerations and ignores the issue of insurance against reclassification risk (e.g., being a sick type, or – behind the veil of ignorance – being born a particular gender), which restrictions on characteristic-based pricing can provide. Bundorf et al. (forthcoming) investigate empirically the reclassification risk created by characteristic-based pricing of employer-provided health insurance. Hendel and Lizzeri (2003) examine issues of reclassification risk in the context of life insurance.

⁴All of the models discussed by Crocker and Snow (1985) assume that individuals differ only in their risk type. Allowing for preference heterogeneity as well presumably makes the potential for Pareto improvements more limited.

2.1.3 Departures from the textbook environment: loads and preference heterogeneity

The qualitative findings of the textbook model are unambiguous: private information about risk always produces under-insurance relative to the efficient outcome, and mandating insurance always improves welfare. We now discuss two empirically relevant departures from the textbook environment that change these qualitative findings.

Production costs (loads) Consider first the supply-side assumption we made above that the only costs of providing insurance to an individual are the direct insurer claims that are paid out. Many insurance markets show evidence of non-trivial loading factors, including long-term care insurance (Brown and Finkelstein, 2007), annuity markets (Friedman and Warshawsky, 1990; Mitchell et al., 1999; Finkelstein and Poterba, 2002), health insurance (Newhouse, 2002), and automobile insurance (Chiappori et al., 2006). While these papers lack the data to distinguish between loading factors arising from administrative costs to the insurance company and those arising from market power (insurance company profits), it seems a reasonable assumption that it is not costless to “produce” insurance and run an insurance company.

We therefore relax the textbook assumption to allow for a loading factor on insurance, for example in the form of administrative costs associated with selling and servicing insurance. In the presence of such loads, it is not necessarily efficient to allocate insurance coverage to all individuals. Even if all individuals are risk averse, the additional cost of providing an individual with insurance may be greater than the risk premium for certain individuals, making it socially efficient to leave such individuals uninsured. This case is illustrated in Figure 3, which is similar to Figure 2, except that the cost curves are shifted upward reflecting the additional cost of insurance provision.

In Figure 3, the MC curve crosses the demand curve at point E, which depicts the socially efficient insurance allocation. It is efficient to insure everyone to the left of point E (since demand exceeds marginal cost), but socially inefficient to insure anyone to the right of point E (since demand is less than marginal cost).

Implications for policy analysis The introduction of loads does not affect the basic analysis of adverse selection but it does have important implications for standard public policy remedies. The competitive equilibrium is still determined by the zero profit condition, or the intersection of the demand curve and the AC curve (point C in Figure 3), and in the presence of adverse selection (downward sloping MC curve) this leads to under-insurance relative to the social optimum ($Q_{eqm} < Q_{eff}$), and to a familiar dead-weight loss triangle CDE.

However, with insurance loads, the qualitative result in the textbook environment of an unambiguous welfare gain from mandatory coverage no longer obtains. As Figure 3 shows, while a mandate that everyone be insured recoups the welfare loss associated with under-insurance (triangle CDE), it also leads to over-insurance by covering individuals whom it is socially inefficient to insure (that is, whose expected costs are above their willingness to pay). This latter effect leads to a welfare loss given by the area EGH in Figure 3. Therefore whether a mandate improves welfare

over the competitive allocation depends on the relative sizes of triangles CDE and EGH. These areas in turn depend on the specific market’s demand and cost curves, making the welfare gain of a mandate an empirical question. It may also depend on factors outside of our model – such as the administrative costs of (publicly-provided) mandatory insurance relative to private sector competition. Naturally, if government-mandated or provided insurance has lower loads – e.g., because of less spending on marketing – then the welfare gains of a mandate could be larger.

Preference heterogeneity and advantageous selection Our “textbook environment” – like the original seminal papers of Akerlof (1970) and Rothschild and Stiglitz (1976) – assumed that individuals varied only in their risk type. In practice, however, consumers of course may also vary in their preferences. Thus the vector of consumer characteristics ζ_i that affects both willingness to pay $\pi(\zeta_i)$ and expected costs $c(\zeta_i)$ may include consumer preferences as well as risk factors.

Recent empirical work has documented not only the existence of substantial preference heterogeneity over various types of insurance, but the substantively important role of this preference heterogeneity in determining demand. Standard expected utility theory suggests that risk aversion will be important for insurance demand. And indeed, recent empirical evidence suggests that heterogeneity in risk aversion may be as or more important than heterogeneity in risk type in explaining patterns of insurance demand in automobile insurance (Cohen and Einav 2007) and in long-term care insurance (Finkelstein and McGarry 2006). In other markets, there is evidence of a role for other types of preferences. For example, in the Medigap market, heterogeneity in cognitive ability appears to be an important determinant of insurance demand (Fang, Keane and Silverman, 2008); in choosing annuity contracts, preferences for having wealth after death play an important role (Einav, Finkelstein and Schripf 2010); in annual health insurance markets, heterogeneity in switching costs can also play an important role in contract demand (Handel, 2011).

Such heterogeneity in preferences can have very important implications for analysis of selection markets. In particular, if preferences are sufficiently important determinants of demand for insurance and sufficiently negatively correlated with risk type, the market can exhibit what has come to be called “advantageous selection.”

Equilibrium and public policy with advantageous selection In our graphical framework, advantageous selection can be characterized by an *upward sloping* marginal cost curve, as shown in Figure 4. This is in contrast to adverse selection, which is defined by a *downward sloping* marginal cost curve.⁵ When selection is advantageous, as price is lowered and more individuals opt into the market, the marginal individual opting in has higher expected cost than infra-marginal individuals. Note that preference heterogeneity is essential for generating these upward sloping

⁵Allowing for preference heterogeneity can complicate the notion of efficiency since the the mapping from expected cost to willingness-to-pay need no longer be unique. In what follows, when we discuss the “efficient allocation” under preference heterogeneity we are referring to the constrained efficient allocation which is the one that maximizes social welfare subject to the constraint that price is the only instrument available for screening (see Einav, Finkelstein and Cullen 2010 for further discussion).

cost curves. Without it, willingness to pay must be higher for higher expected cost individuals. Marginal costs must be upward sloping because the individuals with the highest willingness to pay are highest cost.⁶

Since the MC curve is upward sloping, the AC curve lies everywhere below it. If there were no insurance loads (as in the textbook situation), advantageous selection would not lead to any inefficiency; the MC and AC curves would always lie below the demand curve, and in equilibrium all individuals in the market would be covered, which would be efficient.

With insurance loads, however, advantageous selection generates the mirror image of the adverse selection case; it also leads to inefficiency, but this time due to *over-insurance* rather than *under-insurance*. This can be seen in Figure 4. The efficient allocation calls for providing insurance to all individuals whose expected cost is lower than their willingness to pay—that is, all those who are to the left of point E (where the MC curve intersects the demand curve) in Figure 4. Competitive equilibrium, as before, is determined by the intersection of the AC curve and the demand curve (point C in Figure 4). But since the AC curve now lies below the MC curve, equilibrium implies that too many individuals are provided insurance, leading to over-insurance: there are $Q_{eqm} - Q_{eff}$ individuals who are inefficiently provided insurance in equilibrium. These individuals value the insurance at less than their expected costs, but competitive forces make firms reduce the price in order to attract these individuals, simultaneously attracting more profitable infra-marginal individuals. 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. As we discuss below, such advantageous selection is quite important empirically. Cutler, Finkelstein, and McGarry (2008) summarize some of the findings regarding the presence of adverse compared to advantageous selection in different insurance markets.

We can characterize the welfare loss from over-insurance due to advantageous selection as above. 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, the source of market inefficiency is that consumers vary in their marginal cost, but firms are restricted to uniform pricing.

From a public policy perspective, advantageous selection calls for the opposite solutions relative to the tools used to combat adverse selection. For example, given that advantageous selection produces “too much” insurance relative to the efficient outcome, public policies that tax existing insurance policies (and therefore raise P_{eqm} toward P_{eff}) or outlaw insurance coverage (mandate no coverage) could be welfare-improving. Although there are certainly taxes levied on insurance policies, to our knowledge advantageous selection has not yet been invoked as a rationale in public policy discourse, perhaps reflecting the relative newness of both the theoretical work and empirical evidence. To our knowledge, advantageous selection was first discussed by Hemenway (1990), who termed it “propitious” selection. De Meza and Webb (2001) provide a theoretical treatment of

⁶Once one allows for preference heterogeneity, the marginal cost curve need not be monotone. However for simplicity and clarity we focus on monotone cases here.

advantageous selection and its implications for insurance coverage and public policy.

Advantageous selection provides a nice example of the interplay between theory and empirical work in the selection literature. Motivated by the seminal theoretical papers on adverse selection, empirical researchers set about developing ways to test whether or not adverse selection exists. Some of this empirical work in turn turned up examples of advantageous selection, which the original theory had precluded. This in turn suggested the need for important extensions to the theory.

2.2 Empirical Evidence on Selection

Over the last decade, empirical work on selection in insurance markets has gained considerable momentum, and a fairly extensive and active empirical literature on the topic has emerged. We discuss this literature using the graphical framework described in the previous section. We begin with work designed to test whether or not selection exists in a particular insurance market. Existence of selection is a natural and necessary condition for investigation of its welfare consequences and, not surprisingly, where empirical work started first. We then discuss more recent work designed to empirically quantify the welfare consequences of adverse selection or public policy interventions.

2.2.1 Testing for selection

As is evident from our graphical framework, adverse selection is defined by a downward sloping marginal cost curve. Testing for adverse selection essentially requires testing whether the marginal cost curve is downward sloping. But making inferences about marginal individuals is difficult. Not surprisingly, initial empirical approaches focused on cases under which one could make inferences simply by comparing average rather than marginal individuals. We begin by discussing these “positive correlation tests”. We then move on to a “cost curve test,” which has the advantage of being able to make inferences about marginal individuals, but requires more data.

Positive correlation test for asymmetric information The graphical depiction of adverse selection in Figures 2 and 3 suggests one natural way to test for selection: compare the expected cost of those with insurance to the expected cost of those without. More generally, one can compare the costs of those with more insurance to those with less insurance. If adverse selection is present, the expected costs of those who select more insurance should be larger than the expected costs of those who select less insurance.

Figure 5 illustrates the basic intuition behind the test. Here we start with the adverse selection situation already depicted in Figure 3, denoting the AC curve shown in previous figures by AC^H to reflect the fact that it averages over those individuals with the higher coverage contract, H . We have also added one more line: the AC^L curve. The AC^L curve represents the average expected cost of those individuals who have the lower coverage contract L . That is, the AC^H curve is derived by averaging over the expected costs of those with H coverage (integrating from $Q = 0$ to a given

quantity Q) while the AC^L curve is produced by averaging over the expected costs of those with L coverage (integrating from the given quantity to $Q = Q_{\max}$).

A downward-sloping MC curve – i.e., the existence of adverse selection – implies that AC^H is always above AC^L . Thus, at any given insurance price, and in particular at the equilibrium price, adverse selection implies that the average cost of individuals with more insurance is higher than the average cost of those with less insurance. The difference in these averages is given by line segment CF in Figure 5 (the thick arrowed line in the figure). This basic insight underlies the widely used “positive correlation” test for asymmetric information. The positive correlation test amounts to testing if point C (average costs of those who in equilibrium are insured) is significantly above point F (average costs of those who in equilibrium are not insured).

Chiappori and Salanie (2000) formalized this intuition and emphasized that the basic approach requires some refinement because it does not clearly differentiate between individual characteristics that are observable and those that are not.⁷ In particular, one must stratify on the consumer characteristics that determine the contract menu offered to each individual. Implementing the positive correlation test requires that we examine whether, among a set of individuals who are offered the same coverage options at identical prices, those who buy more insurance have higher expected costs than those who do not. In the absence of such conditioning, it is impossible to know whether a correlation arises due to demand (different individuals self select into different contracts) or supply (different individuals are offered the contracts at different prices by the insurance company). Only the former is evidence of selection. As a result, some of the most convincing tests are those carried out using insurance company data, where the researcher knows the full set of characteristics that the insurance company uses for pricing. Absent data on individually-customized prices, which are sometimes difficult to obtain, one may instead try to flexibly control for all individual characteristics that affect pricing.

Chiappori and Salanie’s work has led to a large literature studying how average costs vary across different coverage options in several insurance markets, including health, life, automobile, and homeowners insurance. The widespread application of the test in part reflects its relatively minimal data requirements. The test requires that one observe the average expected costs of individuals (who are observationally identical to the firm) with different amounts of insurance coverage.

A central limitation in interpreting the results of the positive correlation test is that it is a joint test of the presence of either adverse selection or moral hazard. Even in the absence of selection (i.e. a flat marginal cost curve), moral hazard (loosely, an impact of the insurance contract on expected claims) can produce the same “positive correlation” property of those with more insurance having higher claims than those without. Intuitively, individuals with more generous insurance coverage may choose to utilize more services simply because their marginal out-of-pocket cost is lower. These

⁷Variants of this idea have been discussed in earlier work as well. For instance, Glied (2000) and Cutler & Zeckhauser (2000) summarize attempts to identify risk-based sorting in health insurance choice, where instead of directly comparing claims or “accidents” across individuals with different insurance coverage, the comparison is made over a particular individual characteristic thought to be associated with higher claims, such as age or chronic illness.

two very different forms of asymmetric information have very different public policy implications. In particular, in contrast to the selection case where government intervention could potentially raise welfare, the social planner generally has no comparative advantage over the private sector in ameliorating moral hazard (i.e. in encouraging individuals to choose socially optimal behavior). Thus distinguishing between adverse selection and moral hazard is crucial.⁸

Cost curve test of selection Faced with the challenge of how to interpret the results of the correlation test, researchers have taken a variety of approaches. One is to test for selection in insurance markets where moral hazard is arguably less of a concern, such as annuity markets. More generally, researchers have used experimental or quasi-experimental variation in prices that consumers face to try to separate selection from moral hazard (see e.g., Abbring et al. 2003a,b, Adams et al., 2009, Cutler and Reber 1998, Einav, Finkelstein and Cullen 2010, Karlan and Zinman 2009).

The intuition for how pricing variation that is exogenous to demand (and hence by definition to costs since demand depends on expected costs) allows one to separate selection from moral hazard is easily seen in our simple graphical framework. Consider an experiment that randomly varies the relative price at which the H contract is offered (relative to the L contract) to large pools of otherwise identical individuals. For each relative price, we observe the fraction of individuals who bought contract H and the average realized costs of the individuals who bought contract H .⁹ We thus can trace out the demand curve as well as the average cost curve in Figure 3. From these two curves, the marginal cost curve is easily derived. Total costs are the product of average costs and demand (quantity), and marginal costs are the derivative of total cost with respect to quantity. The features of the marginal cost curve then provide direct evidence on selection. Specifically, rejecting the null hypothesis of a constant marginal cost curve is equivalent to rejecting the null of no selection. Moreover, the sign of the slope of the estimated marginal cost curve informs

⁸If one finds convincing evidence of a negative correlation between insurance coverage and expected claims, this is arguably more informative, as it is consistent with advantageous selection, even in the presence of moral hazard. This is the approach taken by Fang, Keane and Silverman (2008) who document a negative correlation between insurance and claims in the Medigap market, thus pointing to the existence of *advantageous selection* in this market. When a positive correlation is found however – as is the case in many of the papers reviewed by Cohen and Siegleman (2010) – further work is needed to determine whether the results are driven by adverse selection, or by moral hazard (perhaps combined even with advantageous selection). Another vexing case occurs when one is unable to reject the null of no correlation between insurance coverage and expected claims in the market, as in Chiappori and Salanie (2000) or Finkelstein and McGarry (2006). Such a finding is consistent with symmetric information or with the presence of offsetting advantageous selection and moral hazard. Even if moral hazard is ruled out, the inability to reject the null of no correlation could reflect the presence of multiple forms of private information acting in offsetting directions, as Finkelstein and McGarry (2006) find to be the case in the US market for long-term care insurance.

⁹When the L contract is no coverage, the average realized costs of individuals who bought contract H are simply the average claims paid out for people who have contract H . When the L contract involves some (lower) coverage amount, then we measure the average *incremental* claims for those with policy H , or, in other words, the average additional claims that insurance policy H would have to pay out beyond what policy L would pay out, for the accident occurrences of those who have policy H .

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. Einav, Finkelstein and Cullen (2010) develop and discuss in more detail this “cost curve” test of selection.

Crucially, this “cost curve” test of selection is unaffected by moral hazard. Conceptually, variation in prices for a fixed contract allows us to distinguish selection from moral hazard. To see this, recall that the AC curve is estimated using the sample of individuals who choose to buy contract H at a given price. As prices change, the sample changes, but everyone always has the same coverage. Because the coverage of individuals in the sample is fixed, the estimate of the slope of the cost curve is not affected by moral hazard, which only affects costs when coverage changes.¹⁰

2.2.2 Evidence on selection

There is now a large body of empirical work testing for selection in many different insurance markets. The results of this empirical literature have been mixed. In some markets, researchers have found evidence consistent with adverse selection—that is, higher average costs for individuals with greater insurance coverage—while in others they have found evidence of advantageous selection—defined by a negative relationship between insurance coverage and average costs—or have been unable to reject the null of symmetric information, meaning no difference in average costs. Cohen and Siegelman (2010) provide a comprehensive review of this work. We focus instead on characterizing the literature as it pertains to markets with significant social insurance, such as old age assistance and health insurance. Our reading of the evidence for these two markets is that there is very clear evidence of selection in these markets.

Annuities In return for an up-front lump sum premium, annuities provide an individual with a survival-contingent income stream. They therefore offer a way for a retiree facing stochastic mortality to increase welfare by spreading an accumulated stock of resources over a retirement period of uncertain length (Yaari, 1965; Davidoff, Diamond, and Brown 2005). Yet private annuity markets remain quite small. As a result, annuities have attracted a great deal of interest in discussions involving the design and reform of public pensions. Many of these public pension systems, including the current U.S. Social Security System, provide benefits in the form of mandatory, publicly provided annuities. A major economic rationale for this form of benefit provision is the potential for adverse selection to undermine the functioning of private annuity markets, making it important to determine whether selection actually exists in these markets.

Several studies have implemented variants of the positive correlation test for selection in annuity markets. In the context of annuities, higher risk (i.e. higher expected claim) individuals are the

¹⁰Of course, it is possible that the moral hazard effect of insurance is greater for some individuals than others and that, anticipating this, individuals whose behavior is more responsive to insurance may be more likely to buy insurance. We should still view this as selection, however, in the sense that individuals are selecting insurance on the basis of their anticipated behavioral response to it. Einav et al. (2011) provide empirical evidence of such “selection on moral hazard.”

ones who are longer lived than expected; adverse selection therefore is expected to generate a positive correlation between annuitization and survival. Results from a number of studies all point to evidence of a positive correlation in annuity markets on both the extensive margin – individuals who purchase annuities tend to be longer lived than those who do not – and on the intensive (i.e., contract feature) margin – individuals who purchase annuity contracts with shorter guarantee periods tend to be longer lived than those who purchase less. These findings obtain conditional on the characteristics of individuals used to price annuities, namely age and gender. The positive correlation has been documented in several countries including the United States (Mitchell et al., 1999), the United Kingdom (Finkelstein and Poterba 2002, 2004, and 2006) and in Japan (McCarthy and Mitchell 2003).

In the case of annuities, it may be reasonable to assume that annuitization does not induce large behavioral effects. Indeed, work in this literature tends to assume that moral hazard – an impact of income in the form of an annuity on the length of life – is likely to be quantitatively negligible even though theoretically possible (see Philipson and Becker 1998). As a result, evidence of a positive correlation between annuitization and survival can be interpreted as clear evidence of adverse selection in this market (Finkelstein and Poterba 2004). Finkelstein and Poterba’s (2004) empirical findings also illustrate how selection may occur not only along the dimension of the amount of payment in the event the insured risk occurs, but also in the form of selection on different insurance instruments, such as the length of a guarantee period during which payments are not survival-contingent; see Sheshinski (2008) for a theoretical discussion of this point.

Health insurance Cutler and Zeckhauser (2000) review a large literature that tends to find a positive correlation between insurance coverage and risk occurrence in health insurance. Conditional on the menu of contracts offered to them, individuals with more health insurance tend to have higher medical spending than individuals with less insurance. This literature provides a nice example of the substantive importance of conditioning on the observable characteristics of the individual that determine the menu of contracts offered to him, as emphasized by Chiappori and Salanie (2000). In particular, since employer offering of health insurance is such a major determinant of coverage, some of the most convincing implementations of the positive correlation test compare health insurance coverage and medical spending among individuals within the same employer, who therefore all face the same option set. Indeed, without such conditioning, one can get the opposite result suggesting that the insured have lower medical spending than the uninsured, driven by the difference in who is offered health insurance.

In the case of health insurance, the potential moral hazard effects are non-negligible. There is compelling evidence – including results from randomized trials (Newhouse et al. 1996, Finkelstein et al. 2011) – that health insurance has a causal effect on medical spending. As a result, the large body of evidence on the “positive correlation” property in health insurance suggests that asymmetric information exists in health insurance, but does not itself provide direct evidence of selection.

The task of trying to separate selection from treatment effects is greatly aided by the avail-

ability of variation in the offered contracts that is uncorrelated with demand. Arguably the most compelling evidence of adverse selection in health insurance markets comes from observing “death spirals” brought about by changes in the contract set. For example, Cutler and Reber (1998) examine the response to a change in health insurance pricing at Harvard University that required employees to pay more on the margin for more comprehensive coverage. The introduction of this pricing change was staggered over time across different employees. They document a death spiral dynamic whereby the pricing change produced a decline in enrollment in the more comprehensive plan that was particularly concentrated among lower cost (specifically, younger) employees. This prompted a further price increase in the more comprehensive plan to prevent it from losing money, which in turn prompted further exit by disproportionately younger individuals. More recently, Einav, Finkelstein and Cullen (2010) implemented the “cost curve” test for selection using data from a large firm and variation across employees within the firm in the relative price they faced for more comprehensive coverage. They estimate a downward sloping marginal cost curve, providing direct evidence of selection distinct from moral hazard.

We conclude that in the market for acute medical insurance in the United States, there seems to be compelling evidence of the presence of adverse selection. The findings in two other health insurance markets – specifically Medigap and long-term care insurance – are more mixed. While there is compelling evidence of private information in these markets, the evidence suggests that the resulting selection is advantageous rather than adverse (Fang, Keane and Silverman 2008; Finkelstein and McGarry 2006, Oster et al., 2010).

Disability insurance, unemployment insurance, and worker’s compensation In contrast to the study of selection in annuity and health insurance markets there is, to our knowledge, a dearth of work on adverse selection in several settings where there are important social insurance programs including disability insurance, unemployment insurance, and worker’s compensation.¹¹ It would be interesting to test for selection in these markets, although the lack of a robust private market for these risks makes it much more challenging to implement the tests described above. To make progress, one would need to develop tests that – unlike the existing positive correlation or cost curve tests – do not require observing individual choices. In this respect, Hendren’s (2011) development of a method for characterizing the distribution of private information in a market where trade is not observed likely represents an important step forward for empirically-based estimates of private information in non-existent or virtually non-existent markets. He develops a method to infer agents’ private information based on subjective probability elicitation which he models as noisy measures of their beliefs. His results provide, among other things, the first evidence of private information about risk type in the disability insurance setting.

¹¹Hendren (2011) is a notable exception that tests for and detects evidence of private information in disability and unemployment insurance markets.

2.2.3 Welfare consequences

The tests for asymmetric information described above are relatively uninformative about the welfare impacts of interventions. Markets that appear to be “more adversely selected” by the positive correlation metric – i.e. in which there are larger differences between the expected costs of the insured and uninsured – are not necessarily ones in which there is a greater welfare cost imposed by that selection. Einav and Finkelstein (2011) provide a graphical illustration of this point. Intuitively, the degree of positive correlation is a statement about the shape of the cost curves in e.g. Figure 3 or 5. However, the welfare cost of adverse selection – i.e. the magnitude of the “deadweight loss triangle” CDE depends not only on the shape of the cost curve but also on that of the demand curve.

This problem has motivated recent empirical work that quantifies the welfare losses from asymmetric information and the potential impact of government policies such as mandates, pricing restrictions, and taxes. Conceptually, one must estimate both the demand and marginal cost curve to pin down the welfare cost of adverse selection. Once these have been estimated, one can identify the efficient allocation and compare it to alternative allocations induced by various government policies.

Abstractly, there are two approaches one can take to recovering the demand and marginal cost curves. The first is to estimate these curves directly without estimating the underlying primitives that generated these curves. It might be usefully called a “plan valuation” approach and is similar in approach to traditional discrete choice demand analysis. Einav, Finkelstein, and Cullen (2010) develop and implement such an approach to estimating the welfare cost of selection. They show that the demand and cost curves shown in the prior figures are sufficient statistics for welfare analysis of equilibrium and non-equilibrium pricing of existing contracts. That is, different underlying primitives (i.e. preferences and private information as summarized by ζ) have the same welfare implications if they generate the same demand and cost curves. As a result, the identifying variation used to trace out the demand and cost curves for the “cost curve” test of selection provides the estimates needed to estimate the welfare cost of adverse selection (triangle CDE in Figure 3).

Einav, Finkelstein, and Cullen’s approach to estimating welfare is attractive for its transparency and its reduced reliance on assumptions about consumer preferences or the nature of their ex ante information. Moreover, it is relatively straightforward to implement in terms of data requirements. Data on costs and quantities in insurance markets are relatively easy to obtain – as evidenced by the widespread application of the “positive correlation” test which requires both of these data elements. The key additional data requirement is exogenous price variation. While naturally more challenging to obtain, the near-ubiquitous regulation of insurance markets offers many potential opportunities to isolate such variation.

A major limitation of this approach, however, is that the analysis of the welfare cost of adverse selection is limited to the cost associated with inefficient pricing of a fixed (and observed) set of contracts. It does not capture the welfare loss that adverse selection may create by distorting the set of contracts offered, which in many settings could be the primary welfare cost of adverse

selection. Intuitively, in order to analyze the welfare effects of introducing contracts not observed in the data, one needs a model of the deeper primitives (ζ_i).

This limitation partly motivates the second approach that researchers have taken to estimating the welfare costs of selection, which is to directly estimate these primitives and then simulate the welfare cost of alternative policies. For example, Einav, Finkelstein and Schripf (2010) estimate a model of demand for annuities in which the utility from different annuity contracts depends on underlying consumer primitives (ζ_i). Specifically, they examine the semi-compulsory market for annuities in the United Kingdom in which individuals who have saved for retirement through certain tax preferred retirement vehicles are required to annuitize their savings but face a choice over some of the contract features. They focus on the choice of “gaurantee period,” the number of years in which the annuity is guaranteed to pay out even if one has already died. The demand for guarantees depends on both individuals’ unobserved risk type (i.e. survival probability) and unobserved preferences (i.e. for wealth when alive relative to wealth after death). All else equal, longer guarantee periods are more attractive both to individuals who believe they have high mortality and to individuals who have a greater value for wealth after death. Using the model, together with individual-level data on annuity choices and ex post survival length, they recover the joint distribution of survival types and preferences for wealth after death. Unlike the plan valuation approach, this realized utility approach allows recovery of the underlying consumer primitives (ζ_i).

Einav, Finkelstein and Levin (2010) provide more discussion of these two different approaches and their relative attractions. Broadly speaking, the choice between the realized utility approach and plan valuation approaches involves a standard trade-off. The realized utility approach requires stronger assumptions about how consumers derive value from insurance, but allows the researcher to use the resulting estimates to (at least in principle) examine counterfactual allocations that are much further from the observed data.¹² For instance, papers that model realized utility directly as a function of individual primitives such as risk aversion and beliefs about risk type are able in principle to analyze choice and welfare over contracts that vary over dimensions over which one observes no heterogeneity in the data. The papers by Cardon and Hendel (2001), Cohen and Einav (2007), Einav, Finkelstein, and Schripf (2010), and Einav, Finkelstein, Ryan, Schripf, and Cullen (2011) are examples in this vein.

The plan valuation approach requires weaker assumptions but commensurately limits the type of analysis one can do. At one extreme, the approach taken by Einav, Finkelstein and Cullen (2010) recovers the willingness to pay for one health plan over another, but provides no information on the characteristics of the plan determining that valuation. With this approach, inferences about contracts that are not observed in the data are not feasible. Other papers in this literature analyze valuation of contracts as a function of plan and individual characteristics, making it feasible to extrapolate to contracts not observed provided that the model’s assumptions are accurate outside the estimation sample. Examples in this vein include Carlin and Town (2010) and Lustig (2011).

¹²An additional attraction of this approach is that it recovers primitives that may be of inherent interest for other reasons, such as estimates of the distribution of risk aversion in the population.

Estimates of welfare costs of selection Relative to the literature discussed earlier that tests for the existence of selection, there has been substantially less empirical work attempting to estimate the welfare costs of selection. The work that has been done to date has focused on the welfare cost of selection in the health insurance market for acute medical expenses or the market for annuities. The empirical estimates of the welfare cost of selection have consistently tended to be a few percent of premia, bounding the potential welfare gains from policy interventions that aim to address selection at relatively low levels. This is true both in the insurance markets for acute medical expenses (see e.g., Cutler and Reber 1998, Carlin and Town 2010, Einav, Finkelstein and Cullen 2010, Einav, Finkelstein, Ryan, Schrimpf and Cullen 2011, Handel 2011, Lustig 2011) as well as annuity markets (Einav, Finkelstein and Schrimpf 2010). However, as emphasized above and as we return to below, virtually all of these papers have studied only the welfare cost of selection arising from inefficient pricing of a given set of contracts, and have not investigated the potentially much larger welfare losses arising from selection limiting the set of contracts offered or, in the extreme, causing a market to unravel completely.

Welfare consequences of public policy interventions Beyond estimating the welfare cost of selection, several papers have analyzed the welfare consequences of alternative public policy interventions. A recurring theme of this empirical work is that – as indicated by the stylized model we began with – the welfare consequences of “textbook” public policy interventions are not as straightforward as the simple theory suggests.

For example, recent work on mandatory insurance – perhaps the canonical public policy response to selection – has failed to find welfare improvements from the set of mandates considered (Einav, Finkelstein and Cullen 2010). Other work has concluded that while the optimal mandate would be welfare improving, it is difficult to determine which mandate would raise welfare with preference heterogeneity and some types of mandates may actually reduce welfare (Einav, Finkelstein and Schrimpf 2010).

Another focus of the literature has been on the welfare consequences of regulating the characteristics of consumers that can be used in pricing insurance contracts. Bundorf, Levin, and Mahoney (forthcoming) emphasize that in the presence of heterogeneity in preferences for coverage that is not perfectly correlated with risk, uniform pricing of contracts across consumers of different risk types cannot induce efficient consumer choice. This creates scope for welfare improvements through characteristic-based pricing, often known as “risk adjustment.” In their empirical application, which uses data on employer-provided health insurance at several firms, they estimate that the welfare gains from feasible risk adjustment turn out to be relatively modest. In a similar spirit, Geruso (2011) empirically explores the potential welfare gains from age-adjusted pricing in a different employer-provided health insurance context. Focusing instead on distributional impacts, Finkelstein, Poterba and Rothschild (2009) calibrate a stylized equilibrium screening model of annuities to investigate the impact of banning gender-based pricing in a compulsory annuity market; they find that the re-distribution inherent in requiring unisex pricing can be greatly undone by equilibrium adjustment of annuity contracts to the restricted pricing regime.

Other work, focusing once again on health insurance, has examined the implications of pricing restrictions for insurance coverage and government expenditures. Empirical examinations of restrictions on characteristic-based pricing, such as community rating in the small group and non-group health insurance markets have shown that such regulations can reduce coverage among lower risk individuals (e.g., Buchmueller and DiNardo 2002, Simon 2005, Bundorf and Simon 2006). Examining risk adjustment in Medicare Advantage plans, where the government is the insurer, Brown, Duggan, Kuziemko and Woolston (2011) conclude that more detailed pricing on consumer characteristics can exacerbate, rather than ameliorate, the selection problem as defined with respect to government expenditures. The intuition stems from the fact that the variance of medical spending is increasing with its mean. More detailed risk adjustment – which puts consumers into finer pricing “bins” on the basis of their expected health care costs – results in higher dispersion in the high expected spending bins than in the overall pool, increasing the incentives for an insurer to invest in cream skinning within the finer risk classifications.¹³

Welfare consequences with multiple imperfections Another interesting vein of this literature has investigated how adverse selection impacts welfare in the presence of other market imperfections. When selection is the only departure from the perfectly competitive neoclassical benchmark, eliminating private information is always welfare improving. This need not be the case when there are multiple frictions. For example, Starc (2011) discusses how, when firms have market power, moving from symmetric information to asymmetric information can improve consumer welfare. Intuitively, when firms have market power, prices are inefficiently high. Adverse selection reduces the incentive for firms to mark up prices because the marginal consumers they lose when they raise prices have lower costs (and thus are higher profits) than the inframarginal ones that they retain. In a similar spirit, Handel (2011) demonstrates how in the presence of adverse selection, switching costs that reduce consumer response to changes in plan pricing can be welfare increasing by blunting the selection pressures that would otherwise operate.

Measuring the welfare costs of selection is also more complicated in models with additional imperfections. Spinnewijn (2012) shows that calculations based on revealed preference – as in Einav, Finkelstein, and Cullen (2010) – may understate the welfare costs of adverse selection in the presence of frictions such as misperception of risks or adjustment costs. Intuitively, such frictions create a wedge between the revealed demand for insurance via the demand curve and the actual value of insurance. As a result, the demand curve overstates the surplus from insurance for those who buy insurance (because some who purchased do not truly value insurance above cost) and understates it for those who do not (because some who do not purchase should have purchased).

¹³Brown et al. (2011) provide a helpful example to illustrate this intuition: “pre-risk adjustment, Hispanics were roughly \$800 cheaper than their (non-risk-adjusted) capitation payments; after risk adjustment, Hispanics with a history of congestive heart failure (one of the most common conditions included in the risk formula) are \$4,000 cheaper than their (risk-adjusted) capitation payment” (page 3). As a result, the incentive of the insurer to try to recruit Hispanics into plans is much higher when plans are reimbursed differently on the basis of whether the individual has a history of congestive heart failure.”

More generally, though the graphical framework developed above assumes perfect competition among insurers, one can generalize the welfare analysis to any other *given* model of the insurance market as long as one can solve for the equilibrium allocation. Lustig (2011), for example, examines the welfare cost of adverse selection in Medicare Advantage, allowing for imperfect competition on the supply side.

2.2.4 Directions for future work

Most of the empirical papers to date on welfare in insurance markets have taken the relatively narrow (albeit practical) approach of focusing on the welfare costs associated with the pricing distortions selection induces in insurance markets. In general, these papers have concluded that, defined in this way, the welfare costs of selection are relatively small. One limitation to this work, discussed above, is that it analyzes adverse selection in the absence of other potential frictions, which can be important for both the sign and magnitude of the welfare costs of selection. In addition, in at least two important respects, the existing work may be missing important potential welfare consequences of selection or of government intervention. These omissions highlight both the challenges and opportunities for further empirical work.

First, most of the existing empirical welfare analysis has abstracted from a potentially more significant welfare cost of selection that could arise from distortions in the set of contracts offered. Selection may result in certain types of coverage not being available, as in the classic Akerlof (1970) unraveling of a market, and the welfare costs of the disappearance of certain contracts is potentially much larger than the welfare costs of pricing distortions of the contracts that do exist. The ability to make empirically-based estimates of the welfare cost of selection via selection's effect on the set of contracts offered remains a very important area for future work.

There are several challenges inherent in any such attempts. One is that although in principle estimates from realized utility models can use the recovered primitives to say something about the welfare consequences of offering contracts not observed in the data, researchers have been (reasonably) wary of using the estimates of such models to say much about contracts that are too far from the observed contracts. Another challenge stems from the supply side task of trying to characterize the counterfactual equilibrium for unobserved contracts; as discussed by Einav, Finkelstein and Levin (2010), this can be particularly challenging when allowing for realistic consumer heterogeneity as well as imperfect competition.

Even more challenging is empirical work in markets that have almost or completely unraveled, yet it may be that these markets are precisely where the welfare costs of selection are largest; in other words, the “lamp post problem” of empirical work gravitating to markets for which there are data and dimensions of coverage along which there is observed variation may be one reason that existing papers have found relatively small welfare losses.

A few recent papers have used calibration exercises to try to investigate the value of insurance in markets that are virtually non-existent; examples include the market for annuities (Hosseini 2010), long term care insurance (Brown and Finkelstein 2008), and high deductible health insurance (Ma-

honey 2012). Such exercises require that the researchers make assumptions about the population distribution of certain primitives such as risk aversion and risk type, which are often based on estimates made in other, thicker markets. As noted, Hendren (2011) makes important progress in empirically characterizing the distribution of private information in markets where trade does not occur. More work is needed in this area so that researchers may be equipped to examine the counterfactual functioning of private insurance markets that currently do not exist but where we have important social insurance programs such as unemployment insurance, worker’s compensation, and disability insurance.

Second, existing empirical work has focused on testing for the presence of selection and examining its welfare consequences *given the existing public policies*, such as tax subsidies to employer provided health insurance or publicly provided annuities through Social Security. This raises the question of whether selection would exist –and what its welfare consequences would be – in the absence of these public programs or under very different public programs than we currently have. Theoretically, it is not clear whether or when government intervention mitigates or exacerbates selection. For example, as discussed earlier, regulatory restrictions on the consumer characteristics insurance companies may use in setting pricing may potentially increase or decrease the welfare costs of selection in the private market. As another example, the impact of mandatory, partial social insurance (such as Medicare which covers some but not all medical expenses or Social Security which provides partial annuitization) on adverse selection in the residual private market for insurance is theoretically ambiguous. Under different assumptions regarding the ability to offer exclusive contracts, Abel (1986) finds that partial public annuities provided by Social Security exacerbates adverse selection pressures in the residual private market while Eckstein et al. (1985) document a potential welfare enhancing role for partial public annuities. Empirically, we know little about whether the existing partial public insurance programs such as Medicare and Social Security have exacerbated or ameliorated adverse selection problems in the residual private markets for the elderly for health insurance (Medigap) and annuities. Finkelstein (2004) attempts to try to begin to examine such questions empirically. The recent introduction of Medicare Part D, which covers some but not all prescription drug expenses, may provide a fruitful opportunity for empirical work on this question.

2.3 Other motivations

Following much of the recent literature, we have concentrated our discussion above on asymmetric information as a motive for social insurance.¹⁴ This recent focus should not be interpreted as a reflection of a conclusion that selection is the important rationale for social insurance. Here, we briefly summarize several other potential rationales for social insurance. Many of these are ripe for empirical work quantifying their importance.

Incomplete Contracts. Private insurance contracts can only insure risks which are realized after

¹⁴We focused on adverse selection and not moral hazard since, as previously noted, moral hazard is in general not an area where the public sector has a comparative advantage over the private sector in redressing market failure.

birth, as one cannot write contracts behind the Rawlsian veil of ignorance. Social insurance programs can address this problem by redistributing across individuals born with different endowments (e.g., of skills or health). The ability of social insurance to insure risks realized at or before birth might explain why most developed countries – including the United States in the near future – provide or mandate universal health insurance. Social insurance for risks behind the veil of ignorance is formally analogous to optimal taxation and hence we do not treat it further here; see the chapter by Piketty and Saez in this volume for a review of this literature.

Aggregate Risks. Some risks represent aggregate shocks for which the private insurance market’s ability to diversify the risk cross-sectionally may be impaired. By contrast, the government may be able to spread such risk inter-generationally. This may suggest a welfare-improving role for social insurance against such correlated risks as aggregate unemployment shocks, natural disasters, changes in population life expectancy, or technological change in medicine.

An interesting vein of this literature has investigated why such aggregate shocks cannot instead be effectively diversified – and thus insured – intertemporally through private capital markets. In the context of catastrophe risk (e.g., hurricanes and earthquakes), Froot (2001) discusses a variety of possible demand-side and supply-side explanations for why in practice the role of capital markets in reinsuring these risks appears to be limited and prices appear to be high. He also reviews the available evidence for each hypothesis. Demand side explanations include agency issues with insurance company managers who do not value protection for policyholders against extremely high losses if the protection does not avoid default by the firm, ex-post intervention by third parties such as the Federal Emergency Management Agency (FEMA) that substitute for insurance, and optimization failures. Supply side explanations include potential market failures such as adverse selection and moral hazard, firm market power, capital market imperfections that restrict the supply of reinsurance, and high transaction costs.

A related set of issues for private insurance concerns aggregate uncertainty. On the supply-side, it may be difficult for would-be providers to offer insurance against risks with considerable parameter uncertainty; terrorism risk insurance is an example of where this issue may be important (e.g., Brown, Kroszner and Jenn 2002). On the demand-side, aggregate uncertainty may reduce demand for long-term insurance contracts against future risk. For example, Brown and Finkelstein (2011) conjecture that aggregate uncertainty regarding future policy and the survivorship of private insurance companies may depress demand among prime age adults for long-term care insurance that would cover nursing and home health costs in old age.

Externalities. Externalities from insurance constitute another potential rationale for government intervention. One possibility is physical externalities, particularly in the case of health insurance, which subsidizes the treatment or prevention of infectious disease.¹⁵ Another is the possibility of fiscal externalities stemming from the Samaritan’s dilemma (Buchanan 1975). If an altruistic society will provide charitable assistance to those experiencing adverse events ex post, this can

¹⁵Physical externalities may be substantial in some cases; in a developing country context, recent work has documented the enormous social returns in both the short and longer run to the subsidized de-worming of children (Kremer and Miguel 2004, Baird et al., 2011).

reduce individuals' incentives to purchase insurance ex ante. Coate (1995) demonstrates that the optimal transfer policy therefore involves in-kind transfers of insurance to address the inefficient under-insurance that arises in response to unconditional transfers. Ex post unconditional public transfers are frequently observed in the context of health, natural disasters, and terrorism and thus may deter ex ante insurance purchases in these markets to some extent. In health insurance, researchers have documented the relationship between charity care and private insurance coverage (e.g., Herring 2005) and the role of bankruptcy protection in reducing demand for high deductible health insurance (Mahoney 2012). However, we have little evidence on the overall importance of the Samaritan's dilemma effects as a motive for social insurance.

Optimization Failures. Another class of motivations for social insurance is a paternalistic motive premised on imperfect optimization in individual insurance purchases. There is considerable evidence that individuals do not adhere to the lifecycle expected utility model underlying traditional models of risk and insurance. For instance, demand for insurance that covers very small risks with high loads – such as toaster warranties, flight insurance, or homeowners insurance deductibles (e.g., Sydnor 2010) – implies levels of risk aversion that are inconsistent with expected utility theory (Rabin 2000). Barseghyan et al. (2012) argue that this excess demand for low deductibles is explained by misperceptions of loss probabilities. The lack of demand for other types of insurance such as annuities is also difficult to explain in neoclassical models (e.g., Davidoff, Diamond, and Brown 2005).

In the context of Medicare Part D – the 2006 addition to the Medicare program that allows individuals to choose a subsidized prescription drug plan – there is also evidence that individuals make suboptimal choices in choosing not just the level but also the characteristics of insurance coverage. Abaluck and Gruber (2011) argue that many individuals choose sub-optimal drug insurance plans in the Medicare Part D program given the risks they face. Kling et al. (2012) implement a field experiment which suggests that these suboptimal choices in drug insurance plans are due to “comparison frictions” – difficulty for consumers in using available information about plan features to make choices. These are just some examples from a vast literature in psychology and economics that has documented that individuals are prone to various biases such as impatience, loss aversion, overconfidence, and inattention.

The need to account for such biases is especially evident in dynamic models of temporary shocks such as unemployment. As we discuss below, unemployment shocks are quite costly as judged by ex-post measures of consumption-smoothing or liquidity effects. Given these costs, rational agents would build up buffer stocks to help cushion temporary shocks (Deaton 1991, Carroll 1997). But in practice, most individuals build very limited buffer stocks. The median job loser has less than \$200 in liquid assets at the beginning of his unemployment spell (Chetty 2008).¹⁶

¹⁶One potential explanation for low asset holdings is that individuals would save much more in the absence of government-provided social insurance. However, empirical estimates of the impact of unemployment benefit levels on savings are relatively modest in magnitude (Engen and Gruber 2001) and increases in unemployment benefits appear to substantially relax liquidity constraints in practice (Card, Chetty, Weber 2007, Chetty 2008). These findings indicate that individuals would not accumulate substantial buffer stocks even if social insurance benefits

Thus, it is difficult to reconcile the ex-post costs of shocks with optimizing behavior even in an environment without any private insurance markets, because self insurance would be adequate to smooth most temporary shocks.¹⁷ Indeed, Lucas (1987) calculates that optimizing agents would pay less than 1% of lifetime consumption to entirely eliminate business cycle fluctuations even without any private insurance. Hence, the role for social insurance against temporary shocks is quite limited in optimizing models. This suggests that imperfect optimization must be an important motive for social insurance programs in practice. We briefly discuss recent work on optimal social insurance with agents who do not optimize perfectly in Section 3.3.5.

3 Design of Public Insurance Programs

A large literature has analyzed the optimal government response to the failures in private insurance markets discussed above. The goal of this literature is to find the optimal system in terms of trading off protection against risk with minimizing moral hazard. In practice, this problem has several dimensions. Conditional on deciding to insure a risk such as unemployment, there are several policy choices to be made. What level of benefits should be paid? Should benefits rise or fall over an unemployment spell? Should the insurance plan be financed by taxing firms or workers?

The traditional approach to answering such policy questions is to identify a model’s structural primitives and conduct welfare analysis by simulating alternative policies. In social insurance, Wolpin (1987), Hansen and İmrohorođlu (1992), Wang and Williamson (1996), and Hopenhayn and Nicolini (1997) are influential examples of such work. Lentz (2009) and Huggett and Parra (2010) provide recent state-of-the-art applications. While the structural approach is in principle the ideal method of analyzing policy, in practice it is difficult to fully identify all the primitives of complex dynamic models.

Because of this problem, recent studies have instead tackled the optimal policy problem using a “sufficient statistic” approach, which we focus on here. This approach seeks formulas for optimal policy that are a function of high-level empirically estimable elasticities and are relatively robust to changes in the underlying model of behavior. The advantage of this approach is that it offers results about optimal policy that do not rely on the strong assumptions made in structural studies for tractability and identification. The cost is that it can only be used to analyze marginal changes in policy, e.g., the impact of changing the level of benefits incrementally from its current observed level. See Chetty (2009) for a more detailed discussion of the advantages and disadvantages of the sufficient statistic approach.

Research on social insurance has focused primarily on identifying the optimal level of benefits. We organize our discussion of the optimal level of benefits into three subsections. First, we analyze a static model of insurance in which individuals live for a single period and face one risk. In this

were lower.

¹⁷The lack of such buffer stock savings is even more difficult to explain given that shocks such as unemployment generate long-lasting, possibly permanent earnings losses (von Wachter, Song, and Manchester 2009).

static model, it is straightforward to derive an intuitive condition for the optimal level of benefits that trades off the moral hazard costs with the benefits of smaller consumption fluctuations.

Second, we show that this condition for the optimal level of benefits can be written in terms of various empirically estimable parameters. We derive three representations of the formula discussed in recent work: consumption-smoothing benefits, moral hazard vs. liquidity effects, and changes in reservation wages. We discuss empirical evidence on each of these parameters and their implications for the optimal level of benefits.

Third, we analyze the implications of relaxing the assumptions made in the static model. Some of the assumptions are inconsequential. For instance, the formulas obtained from the static model carry over with minor modifications to more realistic dynamic models with endogenous savings, borrowing constraints, and persistent uncertainty. But other extensions to the model – in particular, introducing features such that total private surplus is not maximized by agents’ choices – have significant consequences. For example, if private markets provide insurance that also generates moral hazard, the simple formulas no longer apply because of “multiple dealing” externalities. Similarly, if agents do not maximize their expected utilities because of behavioral failures, the formulas again require modification.

After discussing the literature on the optimal level of benefits, we review recent work on three other aspects of optimal social insurance. First, we discuss work on using mandated savings accounts instead of tax-and-transfer systems to help agents smooth consumption when they face shocks. Next, we discuss imperfect takeup of social insurance programs and its implications for optimal policy. Finally, we review recent work on the optimal path of benefits in dynamic models. Unlike the work on the optimal level of benefits, this literature is primarily theoretical. We therefore present brief summaries of some of the key results in this literature and discuss ways in which theory could be connected to data to make further progress on these questions.

Most recent work on connecting theory to data in optimal social insurance has focused on the case of unemployment insurance. Formal models of unemployment translate readily to most other insurance programs such as worker’s compensation, disability, and catastrophic risks. One exception is the analysis of social security. The models we consider below focus on redistribution across states of nature for a given individual. Social security programs do insure against longevity risk by providing annuities, but also have important effects on the path of individuals’ consumption profiles over their lifecycle. As a result, models of optimal social security typically focus on other factors – such as discount rates and wealth accumulation for retirement – independent of risk reduction. We do not consider models of optimal social security design here; see the chapter by Feldstein and Liebman (2002) for a survey of these models. Gruber and Wise (1999) and Krueger and Meyer (2002) summarize the existing evidence on the impacts of social security programs on retirement behavior. Unlike other social insurance programs, the evidence on social security has not been integrated as tightly with theoretical models to make quantitative statements about welfare and optimal policy. We view this as a fertile area for future research and return to this issue in the concluding section.

3.1 Optimal Benefit Level in a Static Model

Setup. The simplest model of insurance is static and has two states, high (h) and low (l). These states could reflect the risk of job loss (unemployment insurance), injury at work (Worker’s Compensation), or natural disasters (home insurance). Let w_h denote the individual’s income in the high state and $w_l < w_h$ income in the low state. Individuals enter the model with exogenously determined assets A . The government pays a benefit b in the low state that is financed by an actuarially fair tax $\tau(b) = \frac{1-e}{e}b$ in the high state. Let $c_h = A + w_h - \tau(b)$ denote consumption in the high state and $c_l = A + w_l + b$ denote consumption in the low state. Let $v(c)$ denote the agent’s utility as a function of consumption in the high state and $u(c)$ utility in the low state. This allows for the possibility that utility is state-dependent, e.g., agents may value consumption more when healthy. We assume that both u and v are smooth and strictly concave.

A critical feature of the optimal social insurance problem is moral hazard. If individuals’ behaviors were not distorted by the provision of insurance, the planner would achieve the first best by setting b to perfectly smooth marginal utilities, $v'(c_h) = u'(c_l)$. We model moral hazard by assuming that the agent can control the probability of being in the bad state by exerting effort e at a cost $\psi(e)$. For instance, “effort” could reflect spending time to search for a job, taking precautions to avoid injury, or locating a house away from areas prone to natural disasters. We choose units of e so that the probability of being in the high state is given by $e \in [0, 1]$.

Agent’s Problem. The agent chooses effort e to maximize his expected utility:

$$\max_e V(e) = ev(c_h) + (1 - e)u(c_l) - \psi(e) \quad (9)$$

Importantly, we assume that the agent takes the tax and benefit levels offered by the government ($\tau(b)$, b) as fixed when solving this maximization problem. This assumption is a convenient analytical approximation to capture behavior in an economy with a large number of agents, in which the impact of any single agent’s choice of e on the tax rate $\tau(b)$ is negligible.¹⁸ The first order condition for the maximization problem in (9) is:

$$v(c_h) - u(c_l) = \psi'(e). \quad (10)$$

Intuitively, the level of e that maximizes $V(e)$ equates the marginal benefit of an extra unit of effort, given by the difference in utilities in the low and high states, with the marginal cost of exerting an extra unit of effort. Let $e(b)$ denote the agent’s optimal choice of effort given a benefit level b .

Planner’s Problem. The social planner’s objective is to choose the benefit level b that maximizes the agent’s expected utility, taking into account the agent’s endogenous choice of effort:

¹⁸Formally, consider an economy with $i = 1, \dots, N$ identical agents solving (9) and facing idiosyncratic risks. For each agent, the impact of changes in his own effort on $t(b)$ are proportional to $\frac{de}{db} \frac{1}{N}$. For the planner, the aggregate impact of changes in effort is $\sum_{i=1}^N \frac{1}{N} \frac{de}{db} = \frac{de}{db}$. As N grows large, the impact of agent i ’s effort on $t(b)$ approaches 0 and can therefore be ignored when solving the private optimization problem in (9). However, the impact of changes in effort on the planner’s problem in (11) are unaffected by N .

$$\begin{aligned} \max_b W(b) &= ev(A + w_h - \tau(b)) + (1 - e)u(A + w_l + b) - \psi(e) \\ \text{s.t. } e &= e(b) \end{aligned} \quad (11)$$

Differentiating (11) and using the first-order condition for e in (10) gives

$$\begin{aligned} \frac{dW(b)}{db} &= (1 - e)u'(c_l) - \frac{d\tau}{db}ev'(c_h) \\ &= (1 - e)\left\{u'(c_l) - \left(1 + \frac{\varepsilon_{1-e,b}}{e}\right)v'(c_h)\right\} \end{aligned} \quad (12)$$

where $\varepsilon_{1-e,b} = \frac{d(1-e)}{db} \frac{b}{1-e}$ denotes the elasticity of the probability of being in the bad state (which can be measured as the unemployment rate, rate of health insurance claims, etc.) with respect to the benefit level.¹⁹ Notice that in this expression, the behavioral response $\varepsilon_{1-e,b}$ enters only via its impact on the government budget constraint $t(b)$. The direct impact of changes in e on the agent's private welfare is second-order because the agent has already set e at the optimum that maximizes his private welfare ($\frac{\partial W}{\partial e}(e(b)) = 0$). This envelope condition plays a critical role in generalizing (12) to richer, more realistic models, as we discuss in Section 3.3.1.

Equation (12) does not have a cardinal interpretation because it is scaled in utils. One natural cardinal metric is to normalize the welfare gain from a \$1 (balanced budget) increase in the size of the government insurance program by the welfare gain from raising the wage bill in the high state by \$1:

$$\begin{aligned} M_W(b) &= \frac{\frac{dW}{db}(b)/(1-e)}{\frac{dW}{dw_h}(b)/e} \\ &= \frac{u'(c_l) - v'(c_h)}{v'(c_h)} - \frac{\varepsilon_{1-e,b}}{e} \end{aligned} \quad (13)$$

The first term in (13) measures the gap in marginal utilities between the high and low states, which quantifies the welfare gain from transferring an additional dollar from the high to low state. The second term measures the net cost to the government of transferring this \$1 across states due to behavioral responses. The second term arises because the agent does not internalize the fiscal externality that he imposes on the government budget when changing his level of effort. This creates a wedge between the private return to effort and the social return to effort, generating a welfare loss.

At the optimal benefit level b^* , $M_W(b) = 0$ and hence

$$\frac{u'(c_l) - v'(c_h)}{v'(c_h)} = \frac{\varepsilon_{1-e,b}}{e} \quad (14)$$

This expression is a simple variant of Baily's (1978) classic formula for the optimal level of social insurance. It captures a simple – and, as we show below, quite robust – intuition about optimal

¹⁹This elasticity measures the *total* effect of an increase in benefits on e , taking into account the tax increase needed to finance the higher level of benefits.

policy: the optimal benefit level equates the marginal gains from a smoother consumption path with the marginal cost, measured by the behavioral response in effort. Note that (14) is a condition that must hold at the optimal benefit level b^* but is *not* an explicit formula for the level of benefits because all the parameters in (14) are endogenous to b .

Another way to write (14) is in terms of the replacement rate, $r = b/w_h$:

$$\frac{r}{1-r} = -\frac{u'(c_l) - v'(c_h)}{v'(c_h)} \frac{e}{\varepsilon_{1-e, w_h-b}} \quad (15)$$

where $\varepsilon_{1-e, w_h-b} = -\frac{d(1-e)}{db} \frac{w_h-b}{1-e}$ denotes the elasticity of the probability of being in the bad state with respect to the net wage $w_h - b$. This formula bears a close resemblance to the inverse elasticity rules that are familiar from the literature on optimal commodity taxation (Auerbach 1985). Indeed, the optimal social insurance problem is formally identical to an optimal Ramsey taxation problem (Chetty and Saez 2010).

It is important to note that equation (15) is not an explicit formula for the optimal replacement rate r^* . The reason is again that the parameters on the right hand side are all functions of r . The very purpose of raising r is to reduce $\frac{u'(c_l) - v'(c_h)}{v'(c_h)}$. Moreover, the elasticity ε_{1-e, w_h-b} may also vary with r because of liquidity effects (Chetty 2008), as we explain below.

Even in this simple static model, calculating the marginal welfare gain (13) empirically requires some work. The challenge is estimating the gap in marginal utilities $\frac{u'(c_l) - v'(c_h)}{v'(c_h)}$, which requires knowledge of the utility function as well as assets A and wage rates, which may be unobserved by the econometrician. We now discuss recent approaches to tackling this problem.

3.2 Sufficient Statistics Implementation

The modern literature on social insurance has developed three approaches to recover the marginal utility gap in (13): studying consumption fluctuations (Gruber 1997), liquidity and substitution effects in effort (Chetty 2008), and reservation wages (Shimer and Werning 2007). We present simple derivations of each approach here and review empirical evidence in each context.

3.2.1 Consumption Smoothing

Gruber (1997) implements (13) under the assumption that utility is state independent, i.e. $v = u$. We first present Gruber's approach under this assumption and then show how it can be extended to allow for state-dependent utility. Taking a quadratic approximation to the utility function yields:

$$\frac{u'(c_l) - u'(c_h)}{u'(c_h)} = \gamma \frac{\Delta c}{c_h}(b) \quad (16)$$

where $\gamma = -\frac{u''(c_h)}{u'(c_h)} c_h$ is the coefficient of relative risk aversion evaluated at c_h and $\Delta c = c_h - c_l$. Plugging this expression into (13), one obtains the following expression for the marginal welfare gain of raising b :

$$M_W(b) = \gamma \frac{\Delta c}{c_h}(b) - \frac{\varepsilon_{1-e, b}}{e}.$$

This equation shows that risk aversion γ , the observed consumption drop from the high to low state $\frac{\Delta c}{c_h}$, and the elasticity $\varepsilon_{1-e,b}$ are together sufficient to calculate the marginal welfare consequences of changing benefits from the current level. It follows that estimating these statistics is adequate to determine whether the current benefit level is too high or low if the welfare function is concave. To go further and calculate the optimal level of benefits, Gruber estimates the relationship between the size of the consumption drop $\frac{\Delta c}{c_h}(b)$ and the level of benefits b . He posits that the effect of benefits on consumption is a linear function of the replacement rate $r = b/w_h$:

$$\frac{\Delta c}{c_h}(b) \simeq \Delta \log c = \alpha + \beta r \quad (17)$$

In this specification, α measures the drop in consumption that would occur absent government intervention while β measures the slope of the consumption function with respect to the benefit level. Putting this equation together with (16) and (13) yields the following expression for the marginal welfare gain from increasing the benefit level:

$$M_W(b) = (\alpha + \beta r)\gamma - \frac{\varepsilon_{1-e,b}}{e}. \quad (18)$$

Gruber solves for the level of r that sets (18) equal to zero to identify the optimal replacement rate.²⁰ Implementing this formula empirically requires estimates of how consumption fluctuates around shocks as a function of benefit levels (α, β) , the curvature of the utility function γ , and the elasticity that measures distortions in behavior $\varepsilon_{1-e,b}$. There are now several studies estimating each of these parameters for various social insurance programs; we briefly review some illustrative examples of quasi-experimental studies from this literature here.

Evidence on Consumption Smoothing. An early study by Hamermesh (1982) investigates the impacts of unemployment insurance on consumption using cross-sectional consumption data from the Consumer Expenditure Survey. Because Hamermesh does not have panel data, he cannot study changes in consumption around unemployment shocks. Instead, he compares individuals who are currently unemployed and receiving UI with those who are employed. He finds evidence that the marginal propensity to consume out of UI benefits is significantly higher than out of other sources of income, which he interprets as evidence supporting a consumption-smoothing role of UI.

Cochrane (1991) improves upon the analysis in Hamermesh (1982) by using panel data from the Panel Study of Income Dynamics (PSID). Using panel data, he studies how unemployment shocks affect within-household food consumption fluctuations. This is a significant advance over cross-household comparisons, which are likely to be plagued by omitted variable bias. Cochrane finds that unemployment shocks are imperfectly insured – i.e., $\alpha > 0$ in (17) – implying that there is a potential role for government intervention via unemployment insurance. However, Cochrane does not estimate the extent to which providing insurance through a UI system would affect consumption.²¹

²⁰This approach assumes that the other parameters in (18) – namely γ and $\varepsilon_{1-e,b}$ – do not vary with b . In practice, these parameters are likely to vary with b . For example, the liquidity effects documented e.g., in Chetty (2008) imply that $\varepsilon_{1-e,b}$ is likely to fall as b rises, as liquidity constraints bind more tightly when b is low. Hence, one should ideally estimate *all* the sufficient statistics in (18) as a function of b to calculate the optimal benefit level.

²¹Cochrane’s conclusion that insurance markets for unemployment are incomplete rests on the assumption that

Gruber (1997) exploits variation in UI benefit levels that is driven by state law changes to identify β using data on food consumption from the PSID. By controlling flexibly for cross-sectional determinants of the level of UI benefits (such as prior wage rates) and simulating UI benefits based on state laws, Gruber isolates variation in UI benefits that is plausibly orthogonal to other determinants of consumption. Gruber’s point estimates of (17) are $\alpha = 0.24$ and $\beta = -0.28$. These estimates imply that consumption drops on average by 10% given existing UI replacement rates, which are approximately 50% of wages. In the absence of UI, consumption would drop by 24%. Hence, UI plays a significant role in smoothing consumption. However, a 10 percent increase in UI replacement rates generates only a 2.8 percentage point reduction in the consumption drop. This implies that part of the increase in UI benefits is crowded out by other responses, such as reductions in savings (Engen and Gruber 2001) and changes in spousal labor supply (Cullen and Gruber 2000).²²

Gruber’s approach has since become the benchmark quasi-experimental strategy for analyzing how social insurance affects consumption.²³ For instance, Browning and Crossley (2001) implement a similar analysis using data on a broader set of consumption goods from Canada. They find that the average impact of increases in UI benefits on consumption is quite modest, but the impacts are especially large among a subset of households that are likely to be liquidity constrained. Gertler and Gruber (2002) show that severe health shocks have large effects on consumption using panel data from Indonesia and that buffering these shocks by reducing income fluctuations would significantly reduce consumption fluctuations. Bronchetti (2012) implements an approach analogous to Gruber (1997) to the Worker’s Compensation program in the U.S. and again finds evidence that increases in Worker’s Compensation benefits significantly increase consumption levels while individuals are out of work due to injury.

While the evidence that has been accumulated clearly demonstrates that insurance markets are incomplete – i.e., consumption does fall when individuals are hit with shocks – the consumption-smoothing role of social insurance programs is less clear. We can be confident given available evidence that $\alpha > 0$ for at least a subset of households, but we have very imprecise estimates of β . For instance, the estimates of β from Gruber (1997) have a confidence interval spanning $\beta = 0.08$ to 0.48. The imprecision and instability of estimates arise from the fact that consumption is very difficult to measure accurately due to noise and recall errors and is typically available for relatively small samples. Obtaining a more precise understanding of the consumption-smoothing benefits of

utility is not state-dependent. If utility is state-dependent, consumption may fall during unemployment spells simply because the marginal utility of consumption is lower when not working. Gruber’s (1997) approach provides more definitive evidence of incomplete insurance by using variation in UI benefit amounts rather than simply quantifying the size of consumption drops during unemployment.

²²If $c_h \simeq w_h$, as we would expect with low unemployment risk, then a 10 percentage point increase in the replacement rate would lead to a 10 percentage point reduction in the consumption drop absent crowd-out.

²³Another prominent approach is to analyze the impacts of income fluctuations on consumption using statistical decompositions of the income process into permanent and transitory components and examining the covariances of these components with consumption (see e.g., Blundell, Pistaferri, and Preston 2008). Jappelli and Pistaferri (2010) provide a comprehensive review of this work.

insurance will likely require administrative data on consumption, e.g., from credit-card databases, scanner data, or value-added tax registers.

Empirical studies of consumption smoothing have focused on the short-run drop in consumption from employment to unemployment. We show below that this short-run consumption drop is what matters for calculating optimal unemployment benefit levels using (18) irrespective of how consumption evolves after the individual finds a new job. However, it is important to recognize that long-term impacts of temporary shocks on consumption are also significant. von Wachter, Song, and Manchester (2009) show that unemployment shocks due to mass layoffs have large, permanent impacts on earnings. Given that consumption must converge to income in the long run for all workers except the few with substantial wealth before job loss, this result strongly suggests that even temporary unemployment shocks have long-lasting effects. If shocks have persistent impacts on consumption, the optimal insurance policy may not be just to provide benefits while agents are out of work, but rather a wage insurance system that insures long-lasting earnings losses, as proposed e.g., by Lalonde (2007).²⁴ An interesting direction for further work would be to apply the methods reviewed here to analyze optimal wage insurance policies.

Evidence on Distortions in Behavior. The literature on measuring behavioral responses to social insurance programs – the impacts of unemployment insurance on unemployment durations, health insurance on health expenditures, disability insurance on labor force participation rates – has a long tradition that predates the theoretical work on social insurance discussed here. We have much more evidence on the distortions created by insurance programs than their consumption-smoothing benefits because of data availability. For instance, administrative data on unemployment durations must be collected in order to make UI payments, making it much easier to study the impacts of UI on durations than on consumption.

There are many excellent surveys of the literature on how social insurance affects behavior; see e.g., Krueger and Meyer (2002) for a review of work on how UI, DI, and Worker’s Compensation affect labor supply and Cutler and Zeckhauser (2000) or Cutler (2002) for a review on how health insurance affects the demand for medical care. Here, we briefly discuss selected findings from the literature that have been used to inform theoretical calculations of optimal benefit levels using sufficient statistic formulas.

In the context of unemployment, most studies have focused on measuring the impacts of increases in UI benefits on the duration of unemployment. The probability of being laid off could also respond to the level of benefits. The literature has focused less on this issue because UI benefits are typically at least partially experience rated, meaning that firms bear the unemployment insurance cost of laying off workers. In a perfectly experience rated system, changes in the level of benefits do not distort incentives to lay off workers. However, with imperfect experience rating, changes in the level of UI benefits can also affect unemployment rates by distorting firms’ layoff decisions

²⁴To be clear, persistent wage shocks do not invalidate the use of temporary consumption drops to analyze optimal UI, because the observed consumption drop incorporates all future changes in income in an optimizing model (see Section 3.3.1). However, persistent wage shocks raise the possibility that the optimal insurance policy is not merely to provide benefits while the agent is out of work but also after he is re-employed.

(Feldstein 1978, Blanchard and Tirole 2008). While studies such as Topel (1983) and Andersen and Meyer (1993) have documented significant effects of experience rating on firm layoffs, there is relatively little recent work on this issue. Analyzing whether social insurance programs affect the rate at which firms hire and lay off workers using modern quasi-experimental designs is a very promising area for further research.

The modern literature estimating the impact of UI on durations has adopted the hazard model specifications used by Meyer (1990). Meyer estimates semi-parametric models for the hazard of exiting unemployment as a function of UI benefits and other variables using administrative data on the duration of UI claims. He exploits variation in UI benefits coming from differential changes in benefits over time across states, as in Gruber (1997). Meyer finds that higher UI benefits reduce the hazard of exiting unemployment significantly, with an implied elasticity above 0.8 in most specifications.

Subsequent studies have obtained qualitatively similar results using a variety of different data sources. For instance, Lalive, van Ours, and Zweimuller (2006) use a regression-discontinuity design in administrative data from Austria and find that UI benefit increases significantly raise unemployment durations, although to a lesser extent than suggest by Meyer's estimates. Chetty (2008) estimates elasticities of approximately 0.5 using survey data from the Survey of Income and Program Participation. Landais (2012) replicates Meyer's analysis using a regression-kink design and estimates smaller elasticities, around 0.3. In general, the literature has settled on a consensus estimate of $\varepsilon_{1-e,b}$ for UI and unemployment durations of about 0.5 (Krueger and Meyer 2002).

Meyer (1990) and Katz and Meyer (1990) document a spike in hazard rates when unemployment benefits expire. This is typically viewed as *prima facie* evidence that UI distorts search behavior, as it suggests that people time their unemployment exits to coincide with the expiry of social assistance. This spike in unemployment exit hazards in the weeks prior to benefit exhaustion is now a well established empirical regularity; see Card, Chetty, and Weber (2007a) for a review of this literature.

However, Card, Chetty, and Weber (2007a) use data from Austria to show that the spike in *job-finding* rates when UI benefits expire is far smaller than the spike in unemployment exit rates. In Austria, as in most other European countries, individuals can stay on the UI system to receive job finding assistance and other benefits even after their benefits expire, but the majority of individuals choose to drop out of the UI system when their benefits end. Most of these individuals, however, remain unemployed even after they leave the UI system. In the U.S., individuals may choose not to collect their last unemployment check because it is often a small leftover amount, which would create the appearance of a surge in hazard rates in the weeks before benefits expire. Because the margin relevant for calculating the efficiency costs of the UI system are time spent working rather than time spent on the UI system, this evidence suggests that the original sharp spikes documented in the literature likely overstate the degree of moral hazard created by UI. The more general lesson is that it is crucial to measure distortions in real economic choices rather than simply use measures that are well recorded in administrative databases.

Analogous behavioral responses have been documented for other social insurance programs

beyond unemployment insurance. Meyer, Viscusi, and Durbin (1995) use differential changes in worker's compensation benefits across states to show that higher benefit levels induce injured workers to stay out of work longer before returning to work. Gruber (2000) analyzes a disability insurance expansion in Canada that raised benefit levels for individuals in all provinces except Quebec. He finds that this benefit increase significantly reduced labor force participation rates for males ages 45-59, implying an elasticity of the non-participation rate with respect to DI benefits of 0.25. Maestas, Mullen, and Strand (forthcoming) use random variation in assignment to disability insurance examiners to estimate that eligibility for DI reduces labor force participation rates for the marginal entrant to DI by approximately 20 percentage points, with significantly smaller effects for those with more severe impairments. In the context of health insurance, the RAND health insurance experiment (Manning et al. 1987) and Oregon health insurance lotteries (Finkelstein et al. 2011) have demonstrated that increases in consumer cost-sharing significantly reduce health care expenditures.

Evidence on Risk Aversion. Economists have estimated risk aversion using a broad array of techniques. The most direct and widely used method of estimating risk aversion is to assess preferences over gambles. Using empirical estimates of the distribution of risk and an expected utility model with a specific functional form for utility such as constant relative risk aversion, one can back out the value of γ implied by individuals' choices over risky streams of income. Early work in asset pricing inferred risk aversion from portfolio choice and asset returns in standard asset pricing models (e.g., Mehra and Prescott 1985, Kocherlakota 1996). More recent work has used responses to hypothetical large-stake gambles (Barsky et al. 1997), automobile insurance choices (Cohen and Einav 2007), risk-taking in game shows (Metrick 1995), and home insurance deductible choices (Sydnor 2010) to infer risk aversion. There is little consensus on the value of γ from this literature: the estimates range from 1 to well above 10 in the case of deductible choices and asset prices. One explanation of this discrepancy in estimates is that they reflect the behavior of different subgroups of the population. Barseghyan, Prince and Teitelbaum (2011) and Einav et al. (forthcoming) test this explanation by examining the risk preferences of the same individuals in different domains of choice, such as health insurance deductibles and 401(k) portfolio allocations. While individuals' risk preferences are correlated across the domains, there is substantial heterogeneity in estimated risk aversion from each choice.

All of these estimates of risk aversion are based on ex-ante choices, which requires that individuals' subjective assessment of risks (e.g., the probability of a large fluctuation in stock prices) and other parameters are consistent with the model assumed by the researcher as well as the maintained assumptions of expected utility theory. Chetty (2006a) proposes a different method of estimating γ that does not rely on subjective probabilities. He shows that expected utility models imply a direct connection between the curvature of the utility function over consumption and the impacts of wage changes on labor supply. Intuitively, if the utility function is very curved, individuals should become satiated with goods as their income rises, and should choose to work *less* as their wages rise. The fact that uncompensated wage increases almost always raise labor supply in practice implies an upper bound on the coefficient of relative risk aversion of approximately 1 without any assumptions

about the structure of the utility function.²⁵ Because this method of estimating risk aversion uses the same types of ex-post data used to measure $\varepsilon_{1-e,b}$ and $\frac{\Delta c}{c}$, it offers a more direct estimate of the curvature of utility that matters for evaluating the welfare cost of shocks.²⁶

Unfortunately, even this ex-post measure of risk aversion varies significantly across contexts. Chetty and Szeidl (2007) develop a theoretical model of risk preferences in which individuals have “consumption commitments” – goods such as housing or fixed service contracts which can only be adjusted by paying fixed transaction costs. In this environment, individuals have amplified risk aversion over moderate-stake shocks because of their commitments. To understand the intuition, consider a two good model in which the agent spends half his income on housing (which can only be adjusted by paying a transaction cost) and half on food (which is freely adjustable). When facing a shock such as temporary job loss that forces them to reduce expenditure by say 10%, most individuals will rationally choose to bear the shock by cutting food consumption by 20% in order to avoid having to move out of their house. This concentrated reduction in food expenditures raises marginal utility sharply, amplifying risk aversion. Chetty and Szeidl confirm this prediction of the model in the PSID data used by Cochrane and Gruber: homeowners who become unemployed do not change housing consumption but cut back on food consumption significantly, while renters (who face lower adjustment costs) diversify the shocks more broadly by reducing consumption of both food and housing. Chetty and Szeidl’s analysis suggests that the value of γ relevant for shocks such as unemployment could be as high as $\gamma = 4$ because of fixed commitments. However, for large shocks such as permanent disability that induce households to abandon commitments, the relevant value of γ could be closer to 1.

Because of the tremendous uncertainty about the appropriate value of γ , researchers typically report welfare calculations for a range of values of γ . Gruber implements the formula in (18) using his own estimates of the consumption smoothing response and estimates of $\varepsilon_{1-e,b}$ from Meyer (1990). He finds that with a coefficient of relative risk aversion $\gamma < 2$, increasing the UI benefit level above the levels observed in his data (roughly 50% of the wage) would lead to substantial welfare losses. Extrapolating out-of-sample based on the assumption that $\varepsilon_{1-e,b}$ remains constant and the consumption function is given by (17), Gruber shows that it is difficult to justify having a positive level of UI benefits ($b^* > 0$) with risk aversion $\gamma < 2$ given his estimates of the consumption-smoothing benefit of UI. With $\gamma = 4$, however, the optimal benefit level could be as large as 50%. Bronchetti (2012) presents estimates of the optimal level of Worker’s Compensation benefits based on a range of values for γ using the formula in (18). She concludes that the optimal level of worker’s compensation benefits is likely to be below the current level of 68%, but her estimates of the optimal replacement rate range from 26% to 61% as risk aversion varies from $\gamma = 1$ to $\gamma = 4$. Bound et al. (2004) calculate the welfare gains from the current Disability Insurance program under varying

²⁵If utility is non-separable, one must bound the degree of complementarity between consumption and leisure in order to bound risk aversion using this method. Chetty uses the estimates of Cochrane (1991) and Gruber (1997) to place an upper bound on this complementarity parameter and shows that even at this upper bound, $\gamma < 1.25$.

²⁶Conversely, however, this approach is less likely to yield accurate predictions about ex-ante choices in risky environments.

degrees of risk aversion taking account of heterogeneity across individuals. Based on simulations of the benefits of DI, they conclude that the optimal level of benefits is likely somewhat lower than current levels, but the optimal level of benefits again is quite sensitive to assumptions about γ .

Estimating risk aversion accurately is particularly important because the size of the consumption drop $\frac{\Delta c}{c}$ is inversely related to γ , as shown by Chetty and Looney (2006). As a result, the welfare gains from insurance could be large even if consumption drops are small, as documented by Townsend (1994) and others in developing economies. Intuitively, highly risk averse households – e.g., those facing subsistence constraints – are likely to have very smooth consumption paths because they will go to any effort (e.g., by taking their children out of school in developing countries) in order to subsist. But these efforts to smooth consumption are very costly – i.e., $\psi(e)$ is highly convex. Chetty and Looney show that the marginal gains from insurance, given by $\gamma \frac{\Delta c}{c}$, could actually be larger in economies with smoother consumption paths if that smoothness is driven by greater risk aversion.

State-Dependent Utility. When utility is state-dependent, the consumption-smoothing approach requires estimation of an additional parameter that measures the degree to which marginal utilities vary across states. With state-dependent utility, the quadratic approximation used above yields

$$\frac{u'(c_l) - v'(c_h)}{v'(c_h)} = \gamma_v \frac{\Delta c}{c} + \theta$$

where $\gamma_v = -\frac{v''(c_h)c_h}{v'(c_h)}$ denotes the coefficient of relative risk aversion in the employed state and $\theta = \frac{u'(c_l) - v'(c_l)}{v'(c_h)}$ measures the degree of state-dependence in marginal utilities. The parameter θ answers the question, “starting from equal consumption in the low and high states, how much would the agent pay to reallocate \$1 of consumption from the high state to the low state?” If utility is not state-dependent, the answer to this question would be $\theta = 0$. If the marginal utility of consumption is higher when the agent is in the low state, the willingness to pay is $\theta > 0$.²⁷ The parameter θ directly enters the formula for $M_W(b)$ in (18) as an additive term. If $\theta > 0$, insurance has greater value because it is transferring resources to a state where money has more value at any given level of consumption.

Finkelstein, Luttmer and Notowigdo (2009) describe some approaches to estimating θ . Broadly speaking, θ can be estimated based on either choice data that reveals individuals’ demand for moving resources across states or based on observed utility changes as states change. For example, the extent to which agents voluntarily choose to have more or less consumption in high vs. low states reveals θ in an environment with perfect insurance. Unfortunately, most individuals are not perfectly insured in practice – if they were, there would be no reason for social insurance to begin with! – and it is rare to be able to observe consumption across state changes in data (e.g., health shocks or unemployment spells). Absent perfect insurance, one can try to focus on subgroups that are better insured. This requires assumptions about the degree of insurance one has. Moreover, since the life cycle budget constraint must be satisfied, inferring state dependence

²⁷Technically, one must measure the willingness to pay in the high state, as $v'(c_h)$ appears in the denominator of θ .

from the consumption fluctuations of individuals who experience different unexpected shocks (such as health events) requires strong assumptions about the nature of bequest motives. For example, Lillard and Weiss (1997) estimate a structural model of health shocks and use data on consumption trajectories to identify θ under the assumption that the marginal utility of bequests does not depend on health. They estimate $\theta > 0$, i.e. positive state dependence for health and disability.²⁸

An alternative approach to identify state-dependence that does not require choice data from environments with full insurance is to use data on subjective well-being. Intuitively, by estimating whether a cash grant has a larger impact on happiness in the low vs. high state, one can learn about θ . The challenge in implementing this approach is that subjective well being measures have no inherent cardinal interpretation, whereas θ is a cardinal parameter.²⁹ One must therefore choose a cardinal scale for happiness to estimate θ using data on subjective well-being. One approach is to scale happiness so that one obtains estimates of the coefficient of relative risk aversion γ that match estimates from choice data. Finkelstein, Luttmer, and Notowidigdo (2008) implement such an approach and conclude based on survey data that $\theta < 0$, i.e. the marginal utility of consumption is higher when individuals are healthy. In sum, there is currently little consensus in the literature on the sign of θ for shocks such as unemployment, disability, and sickness, let alone its magnitude in these contexts. We view this as an important but challenging area for future work.

Because optimal benefit calculations are highly sensitive to the assumed value of γ and θ , more recent studies have sought alternative techniques for recovering the gap in marginal utilities that do not require estimates of these parameters. We now turn to these alternative approaches.

3.2.2 Liquidity vs. Moral Hazard

Chetty (2008a) shows that the gap in marginal utilities in (13) can be inferred from the comparative statics of effort choice, yielding a formula for optimal benefits that does not require any data on consumption or risk preferences. Recall that the first order condition for effort in our static model is $\psi'(e) = v(c_h) - u(c_l)$. Now consider the effect of an exogenous cash grant (such as a severance payment to job losers) on effort, holding fixed the tax τ :

$$\partial e / \partial A = \{v'(c_h) - u'(c_l)\} / \psi''(e) \leq 0 \quad (19)$$

The effect of increasing the benefit level on effort (again holding τ fixed) is:

$$\partial e / \partial b = -u'(c_l) / \psi''(e) \quad (20)$$

Finally, the effect of increasing the wage in the high state on effort is:

$$\partial e / \partial w_h = v'(c_h) / \psi''(e) \quad (21)$$

²⁸In a different context, Browning and Crossley's (2001) finding that unemployment shocks have little impact on consumption for individuals with high levels of assets suggests that θ may be 0 for unemployment (i.e. no state dependence for leisure).

²⁹For instance, suppose giving an agent \$1,000 increases his reported happiness from 2 to 3 in the low state and 7 to 9 in the high state. One cannot identify θ without knowing whether an increase in happiness from 2 to 3 translates into a smaller or larger welfare gain than an increase in happiness from 7 to 9.

Combining (19) and (20), we see that the ratio of the “liquidity” effect ($\partial e/\partial A$) to the “substitution” effect ($\partial e/\partial w_h = \partial e/\partial A - \partial e/\partial b$) recovers the gap in marginal utilities:

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

Plugging this expression into (13) yields the following expression for the welfare gain from increasing the benefit level:

$$M_W(b) = \frac{-\partial e/\partial A}{\partial e/\partial A - \partial e/\partial b} - \frac{\varepsilon_{1-e,b}}{e} \quad (22)$$

In this formula, the degree to which marginal utilities fluctuate across states is identified from the relative size of liquidity and moral hazard effects in the impact of benefit levels on effort.³⁰ In a model with perfect consumption smoothing, the liquidity effect $\partial e/\partial A = 0$, because a cash grant raises $v(c_h)$ and $u(c_l)$ by the same amount. Note that unlike the consumption-smoothing method above, this approach does not require estimation of the degree of risk aversion or state-dependence in utility.³¹

One intuition for the liquidity vs. moral hazard formula comes from familiar results from price theory. The impact of an increase in b on e can be decomposed into two terms, analogous to a Slutsky decomposition: $\partial e/\partial b = \partial e/\partial A - \partial e/\partial w_h$. The first term is analogous to an income effect, and reflects the fact that higher benefits also raise agents’ cash-on-hand and thus reduce the supply of effort. The second term is a pure substitution (price) effect and arises from the distortion in marginal incentives created by the social insurance program. The substitution effect is efficiency reducing because it reflects second-best behavior arising from the wedge between private and social incentives. In contrast, the liquidity effect is efficiency *enhancing* because it allows the agent to choose a level of e that is closer to what he would choose with complete markets. The size of the liquidity effect measures the extent to which insurance markets are incomplete. The ratio of the liquidity effect to the distortionary substitution effect thus captures the marginal benefit of social insurance.

Another way to understand (22) is that it uses revealed preference to value the benefits of insurance. Consider an application to health insurance. The effect of a lump-sum cash grant on health care consumption reveals the extent to which health insurance permits the agent to attain a more socially desirable allocation. If the agent chooses to spend a lump-sum grant on buying a new car instead of purchasing more healthcare, we infer that the agent only spends more on health care when health insurance benefits are increased because of the price subsidy for doing so. In this case,

³⁰ A technical issue which arises in empirical implementation of (22) is that $\frac{\partial e}{\partial b}$ must be measured holding the tax τ fixed, whereas the elasticity $\varepsilon_{1-e,b}$ must be measured while permitting τ to vary. Instead of attempting to estimate both parameters, Chetty uses numerical simulations to show that the effect of a UI benefit increase on job finding rates is virtually identical whether or not UI taxes are held fixed. This is because the fraction of unemployed individuals is quite small, making UI tax rates very low.

³¹ If utility is not state dependent, one could infer the agent’s degree of risk aversion γ from the size of moral hazard vs. liquidity effects and the consumption drop: $\gamma = \frac{-\partial e/\partial A}{\partial e/\partial A - \partial e/\partial b} / (\Delta c/c)$. This restriction, which is closely related to the formula for risk aversion in Chetty (2006a), could be evaluated empirically in future work.

health insurance simply creates inefficiency by distorting the private cost of health care below the social cost, implying $\frac{dW}{db} < 0$. In contrast, if the agent raises his health expenditures substantially even when he receives a non-distortionary lump sum cash grant, we infer that insurance permits him to make a more (socially) optimal choice, i.e. the choice he would make if insurance market failures could be alleviated without distorting incentives. The liquidity vs. moral hazard approach in (22) thus identifies the policy that is best from the libertarian criterion of correcting market failures as revealed by individual choice.

It follows from equation (22) that larger elasticities $\varepsilon_{1-e,b}$ do not necessarily mean that social insurance is less desirable, a point emphasized by Nyman (2003) in the context of health insurance. It matters whether a higher value of $\varepsilon_{1-e,b}$ comes from a larger liquidity ($-\frac{\partial e}{\partial A}$) or moral hazard ($\frac{\partial e}{\partial w_h}$) component. To the extent that comes from a liquidity effect, insurance reduces the need for agents to make suboptimal choices driven by insufficient ability to smooth consumption. In contrast, if $\varepsilon_{1-e,b}$ is large primarily because of a moral hazard effect, insurance is distorting incentives. For instance, the reductions in labor force participation caused by disability insurance documented by Maestas, Mullen, and Strand (2011) may not be undesirable. If DI helps individuals whose marginal product is lower than their disutility of work, this behavioral response is welfare-improving.

Evidence on Liquidity vs. Moral Hazard. The advantage of this formula relative to (18) is that it can be implemented purely using data on e (e.g., unemployment durations or health expenditures). Chetty (2008) implements (22) using survey data from the U.S. He estimates $\frac{\partial e}{\partial b}$ using data from the Survey of Income and Program Participation, following the specifications in Meyer (1990). Using the SIPP data, Chetty shows that individuals who have low levels of assets prior to job loss exhibit much higher levels of $\frac{\partial e}{\partial b}$ than those with higher levels of assets. This suggests that a significant part of the impact of UI benefits on durations may be driven by a liquidity effect. Chetty then estimates $\frac{\partial e}{\partial A}$ by studying the effects of lump-sum severance payments on unemployment durations using data from a survey of UI exhaustees conducted by Mathematica in collaboration with the Department of Labor. He finds that individuals who receive lump-sum severance payments have significantly longer unemployment durations, especially if they have low levels of assets prior to job loss. Using his estimates, Chetty calculates the welfare gain of raising the UI benefit level from the current replacement rate of approximately 50% of wages. He finds that the welfare gains from raising b are small but positive, suggesting that the current benefit level is slightly below but near the optimum.

Chetty's (2008) analysis relies on cross-sectional variation across individuals in severance pay, and therefore rests on the strong identification assumption that severance recipients and non-recipients are comparable. More recent studies have documented similar results using research designs that make weaker assumptions. Card, Chetty, and Weber (2007b) use a regression-discontinuity design that exploits a universal cutoff for severance pay eligibility based on job tenure in Austria. Using administrative data for the universe of job losers in Austria, they show that individuals laid off just after the tenure cutoff – who receive 2 months of wages as a lump sum severance payment as a result – have unemployment durations that are about 10 days longer than

individuals laid off just before the cutoff. They compare these liquidity effects to the impacts of UI benefit extensions using a similar RD design and show that the size of liquidity effects is large relative to the impact of benefit extensions, implying that many unemployed individuals are liquidity constrained. Centeno and Novo (2009) present evidence that liquidity-constrained households are more sensitive to UI benefit changes using a regression-discontinuity design in Portugal. Lalumia (2011) shows that individuals who happen to be laid off after they receive a tax refund (and have more cash-on-hand) have longer unemployment durations relative to individuals laid off at other times of the year.

There is less evidence on liquidity effects in other insurance programs because of the relatively recent development of these formulas and because researchers have not yet formulated quasi-experimental designs to estimate liquidity effects $\frac{\partial e}{\partial A}$ for many programs. One exception is Nyman (2003), who presents considerable anecdotal evidence and theoretical arguments suggesting that liquidity effects play a very important role in health insurance. The benefits of health insurance are extremely difficult to quantify, making the liquidity-based revealed preference approach particularly attractive in that context. Finding research designs to identify liquidity effects in health and other insurance programs is thus a very promising area for further work.

3.2.3 Reservation Wages

Shimer and Werning (2007) show that the gap in marginal utilities can be recovered from the comparative statics of reservation wages in a standard model of job search. They analyze a model in which the probability of finding a job, e , is determined by the agent's decision to accept or reject a wage offer rather than by search effort. Wage offers are drawn from a distribution $F(w)$. If the agent rejects the job offer, he receives income of $w_l + b$ as in the baseline model above.³²

The agent rejects any net-of-tax wage offer $w - \tau$ below his outside option $w_l + b$, i.e. his reservation wage once searching for a job is $w_l + b + \tau$. Therefore, $e = 1 - F(w_l + b + \tau)$ and the agent's expected utility when searching for a job is

$$W(b) = e\mathbb{E}[v(A + w - \tau) | w - \tau > w_l + b] + (1 - e)u(A + w_l + b).$$

Now suppose we ask the agent what wage he would be willing to accept with certainty *before* the start of job search. Define the agent's reservation wage prior to job search as the wage \bar{w}_0 that would make the agent indifferent between accepting a job immediately vs. starting the process of job search, which yields expected utility of $W(b)$. This pre-job-search reservation wage \bar{w}_0 satisfies

$$v(A + \bar{w}_0 - \tau) = W(b)$$

The government's problem is to

$$\begin{aligned} \max W(b) &= \max v(A + \bar{w}_0 - \tau) \\ &\Rightarrow \max \bar{w}_0 - \tau \end{aligned} \tag{23}$$

³²Both formulas derived above continue to hold in this model with stochastic wages. Conversely, the Shimer and Werning formula also holds in a model with variable search intensity.

Differentiating (23) gives the following formula for the marginal welfare gain of raising b :³³

$$M_W(b) = \frac{d\bar{w}_0}{db} - \frac{d\tau}{db} = \frac{d\bar{w}_0}{db} - \frac{1-e}{e} \left(1 + \frac{1}{e} \varepsilon_{1-e,b}\right) \quad (24)$$

In this formula, $\frac{d\bar{w}_0}{db}$ encodes the marginal value of insurance because the agent's reservation wage directly measures his expected value when unemployed. Intuitively, if agents can smooth marginal utilities perfectly across states, their reservation wage will depend purely on their expected income, and the marginal value of raising b by \$1 (holding fixed τ) is simply $(1 - e)$. But this marginal gain is outweighed by the cost of financing the extra dollar of benefits, which exceeds $1 - e$ because of the behavioral response $\varepsilon_{1-e,b}$. When marginal utilities fluctuate across states, agents value an increase in b at more than its actuarial cost, increasing the marginal value of public insurance. The extent to which agents value insurance can be captured by asking them how their valuation of being in the unemployed state varies with b , which is the parameter $\frac{d\bar{w}_0}{db}$. Like the liquidity vs. moral hazard method, this approach also does not require identification of state-dependence in utility or risk aversion.

An interesting implication of this result is that a higher sensitivity of reservation wages to benefits implies a *larger* value of social insurance, contrary to the intuition embodied in earlier empirical studies (e.g., Feldstein and Poterba 1984), which view the sensitivity of reservation wages to UI benefits as a distortion. This point, like the moral hazard vs. liquidity decomposition above, illustrates the importance of connecting empirical estimates to models in order to fully understand the implications of empirical findings for policy.

Evidence on Reservation Wages. A long literature has sought to measure unemployed workers' reservation wages; see Devine and Kiefer (1991) for a review of early work and Krueger and Mueller (2011) for recent evidence. While these studies have shown that reservation wages have predictive power for unemployment durations and the types of offers that workers accept, they have also documented significant problems with self-reported reservation wage measures. For instance, a large fraction of workers end up accepting jobs that pay below their reported reservation wage. Moreover, most jobs have many characteristics that matter beyond the wage rate, such as the nature of the work or commuting distance. A one-dimensional reservation wage measure does not incorporate these other dimensions. When jobs have multiple characteristics, in order to implement (24), one would ideally like to measure how reservation *utilities* for jobs vary with the benefit rate. Because such reservation utilities cannot be easily measured, it is difficult to estimate $\frac{d\bar{w}_0}{db}$ accurately.

Perhaps because of these measurement issues, there is relatively little evidence on the impact of UI benefits on reservation wages. Feldstein and Poterba (1984) use survey data from the Current Population Survey to estimate $\frac{d\bar{w}_0}{db}$ by studying how changes in UI benefit levels affect reported reservation wages. Their point estimates imply a substantial correlation between UI benefit levels

³³This corresponds to equation (12) in Shimer and Werning (2007), where the unemployment rate is $u = 1 - e$. The slight difference between the formulas (the $\frac{1}{1-u}$ factor in the denominator) arises because Shimer and Werning write the formula in terms of a partial-derivative-based elasticity. Here, $\varepsilon_{1-e,b}$ is the elasticity including the UI tax response needed to balance the budget; in Shimer and Werning's notation, it is holding the tax fixed.

and reservation wages. However, their specification does not isolate purely exogenous variation in UI benefits due to law changes, as in more recent work, which raises concerns about omitted variable bias. Shimer and Werning implement (23) using an estimate of $\frac{d\bar{w}_0}{db}$ from Feldstein and Poterba (1984) and find a large, positive value for $M_W(b)$ at current benefit levels. They caution, however, that their exercise must be viewed as purely illustrative given the uncertainty in estimates of $\frac{d\bar{w}_0}{db}$ and call for further work on estimating this parameter using alternative methods.

One alternative approach is to use data on actual wages obtained at the next job and back out the implied distribution of reservation wages. There are now a large set of studies using administrative panel data that study whether increasing UI benefits raises subsequent wage rates. Card, Chetty, and Weber (2007b) use a regression discontinuity design in data from Austria to show that individuals eligible for 10 additional weeks of UI benefits take more time to find a job but do not have higher wages at their next job. Their estimates are sufficiently precise to rule out even a 1% increase in the wage rate at the upper bound of the 95% confidence interval. They also show that there are no detectable impacts on other observable job characteristics or on the number of years the worker spends at his next job, a summary measure of job match quality. Subsequent studies using similar RD designs have reached very similar conclusions. For instance, Lalive (2007) studies a 170 week benefit extension in Austria and shows that it has no impact on subsequent wages. van Ours and Vodopivec (2008) use a difference-in-difference approach using data from Slovenia and find that changes in benefit duration from 6 months to 3 months had no impact on subsequent earnings. These findings that ex-post observed wages are unaffected by benefit levels imply that reservation wages must be unaffected by benefit levels, i.e. $\frac{d\bar{w}_0}{db} = 0$.

How can UI have significant consumption-smoothing and liquidity benefits but little effect on ex-post wages? Lentz and Tranaes (2005), Chetty (2008), and Card, Chetty, and Weber (2007b) develop pure search intensity models with borrowing constraints that generate these patterns. We present a stylized version of this model in Section 3.3.1 below. In these models, workers can control the amount of effort they spend searching for a job, but have a fixed wage rate. Such models can be viewed as an approximation to an environment in which the arrival rate of suitable job offers is relatively low, so the option value of waiting for a better offer is small and most workers take the first offer they receive. One puzzling feature of the data given this explanation is that there is tremendous heterogeneity in observed wage changes from the previous job to the new job across unemployed workers. Understanding how the variance of job offers can be reconciled with the lack of mean impacts of UI benefits on reservation wages is an open question for future research.³⁴

In summary, there are now a variety of methods of calculating the welfare gains from increasing benefit levels in social insurance programs, but each of the methods yields different results because empirical evidence on many of the key parameters remains inadequate. Moreover, the three formulas discussed above are not an exhaustive list of potential approaches for connecting theory to data in analyzing optimal benefit levels for social insurance. There could be many other

³⁴One possibility is a model with impatient workers who wait to search for a job until they have no cash on hand. Such a model would generate significant liquidity effects but no impacts on ex-post wages, and would have very different welfare implications.

representations of the optimality condition in (13) that could be useful for applied work. The multiplicity of formulas for $M_W(b)$ is a general property of the sufficient-statistic approach (Chetty 2009). Because the positive model is not fully identified by the inputs to the formula, there are generally several representations of the formula for welfare gains. This flexibility allows researchers to use the representation most suitable for their applications given the available variation and data.

3.3 Generalizing the Static Model

The analysis above rests on a static model that makes several strong assumptions that are unlikely to hold in practice. Many of these assumptions turn out to have little impact on the formulas for optimal social insurance derived above, which is why these formulas have been widely applied. However, there are some assumptions that are more consequential and raise issues that remain unresolved. In this section, we consider the consequences of each of the main assumptions in turn. The general principle is that if agents' choices in the private sector maximize private surplus, then the formulas derived above continue to hold irrespective of the structure of the model. However, as soon as choices do not maximize private surplus – either because of externalities or imperfect optimization – the formulas no longer hold.

One important assumption that we do not relax below – not because it is inconsequential but rather because there is very limited normative work on this issue – is that wage rates and all other prices are fixed. That is, we do not consider the possibility that changes in social insurance policies will affect market clearing prices in general equilibrium. Acemoglu and Shimer (1999) analyze optimal unemployment insurance in an equilibrium search model with endogenous occupation choice and show that expanding UI can improve efficiency in this environment by encouraging more workers to take high-wage, high-risk jobs. There is no empirical evidence to date on the magnitude of such effects.³⁵ Perhaps an even greater challenge is that there are no results to date on how such empirical estimates could be connected to equilibrium models to make quantitative statements about optimal policy.

3.3.1 Dynamics: Endogenous Savings and Borrowing Constraints

The most important limitation of the model analyzed above is that it does not incorporate dynamics. In dynamic models, agents can smooth consumption across periods and thus “self insure” part of the income fluctuations they face, potentially reducing the value of social insurance. In addition, social insurance distorts not just agents' effort choices but also their consumption and savings decisions. Hubbard, Skinner, and Zeldes (1995) present simulation evidence suggesting that social insurance

³⁵Rothstein (2010) and Chetty et al. (2011) present evidence that tax and transfer policies affect wage rates and the distribution of jobs in equilibrium. Rothstein shows that expansions of the Earned Income Tax Credit (EITC) depress wages for workers who are ineligible for the EITC, while Chetty et al. show that the tax schedule in Denmark affects even the earnings of workers unaffected by the tax incentives because of firm responses. The effects of social insurance on equilibrium outcomes could potentially be uncovered using a similar research design that focuses on groups whose incentives are not directly affected by the program.

and means-tested transfer programs can substantially reduce savings rates. Engen and Gruber (2001) show that increases in UI benefits have small but significant effects on wealth accumulation prior to job loss among workers at risk of layoff. These intertemporal consumption-smoothing and savings responses may be further complicated by borrowing constraints, generating buffer stock behavior as in Deaton (1991) or Carroll (1997).

Models that incorporate all of these features have been widely studied and are complex and difficult to solve analytically. Surprisingly, however, the three simple formulas for the marginal welfare gain from raising social insurance benefits derived above continue to hold in such models with small modifications. Because this result underpins much of the modern literature on connecting theory to data in analyzing optimal social insurance, we provide a simple proof here.

We analyze the optimal level of unemployment benefits in a dynamic job search model with borrowing constraints, following Lentz and Tranaes (2005) and Chetty (2008). For instructive purposes, we structure the analysis to parallel the steps for the static case in Section 4.1: (1) model setup, (2) characterizing the agent's problem, (3) the planner's problem, (4) deriving a condition for optimal benefits by exploiting envelope conditions, and (5) deriving empirically implementable sufficient statistic formulas.

Setup. The agent lives for T periods $\{0, \dots, T - 1\}$. The interest rate and the agent's time discount rate are set to zero to simplify notation. The agent becomes (exogenously) unemployed at $t = 0$. An agent who enters a period t without a job first chooses search effort e_t . As in the static case, we normalize e_t to equal the probability of finding a job in the current period. Let $\psi(e_t)$ denote the cost of search effort, which is strictly increasing and convex. If search is successful, the agent begins working immediately in period t . All jobs last indefinitely once found.

We make two assumptions to simplify exposition: (1) the agent earns a fixed pre-tax wage of w when employed, eliminating reservation-wage choices and (2) assets prior to job loss (A_0) are exogenous, eliminating effects of UI benefits on savings behavior prior to job loss. Neither assumption affects the results below (Chetty 2008).

If the worker is unemployed in period t , he receives an unemployment benefit $b < w$. If the worker is employed in period t , he pays a tax τ . Let c_t^e denote the agent's consumption in period t if a job is found in that period. Note that the agent will optimally set consumption at c_t^e for all $t' > t$ as well because he faces no uncertainty once he finds a job and therefore smooths consumption perfectly. If the agent fails to find a job in period t , he sets consumption to c_t^u . The agent then enters period $t + 1$ unemployed and the problem repeats. Let $u(c_t)$ denote flow consumption utility when unemployed and $v(c_t)$ denote flow utility when employed.

Agent's Problem. We characterize the solution to the agent's problem using discrete-time dynamic programming. The value function for an individual who finds a job at the beginning of period t , conditional on beginning the period with assets A_t is

$$V_t(A_t) = \max_{A_{t+1} \geq L} v(A_t - A_{t+1} + w_t - \tau) + V_{t+1}(A_{t+1}). \quad (25)$$

where L is the borrowing constraint. The value function for an individual who fails to find a job

at the beginning of period t and remains unemployed is

$$U_t(A_t) = \max_{A_{t+1} \geq L} u(A_t - A_{t+1} + b_t) + J_{t+1}(A_{t+1}) \quad (26)$$

where

$$J_t(A_t) = \max_{e_t} e_t V_t(A_t) + (1 - e_t) U_t(A_t) - \psi(e_t) \quad (27)$$

is the value of entering period t without a job with assets A_t .³⁶

An unemployed agent chooses e_t to maximize expected utility at the beginning of period t , given by (27). Optimal search intensity is determined by the first-order condition

$$\psi'(e_t) = V_t(A_t) - U_t(A_t). \quad (28)$$

The condition parallels (10) in the static model, except that the marginal value of search effort is given by the difference between the optimized present values of employment and unemployment in the dynamic model rather than the difference in flow utilities.

Planner's Problem. The social planner's objective is to choose the unemployment benefit level b that maximizes the agent's expected utility at time 0, $J_0(A_0; b, \tau)$, subject to balancing the government's budget. Let $D = \sum_{t=0}^{T-1} \prod_{j=0}^t (1 - e_j)$ denote the agent's expected unemployment duration. The planner's problem is:

$$\max_{b, \tau} J_0(A_0; b, \tau) \text{ s.t. } Db = (T - D)\tau \quad (29)$$

Differentiating J_0 w.r.t. b yields:

$$\frac{dJ_0}{db} = e_0 \frac{\partial V_0}{\partial b} + (1 - e_0) \frac{\partial U_0}{\partial b} - (e_0 \frac{\partial V_0}{\partial \tau} + (1 - e_0) \frac{\partial U_0}{\partial \tau}) \frac{d\tau}{db} \quad (30)$$

The key step in obtaining an empirically implementable representation of (30) is to exploit the envelope conditions for e_t , c_t^u , and c_t^e . These variables are all chosen to maximize the agent's expected utility at each stage of his dynamic program. Because changes in these variables do not have a first-order impact on utility, one can ignore the impacts of changes in b and τ on these choices when calculating the derivatives in (30). Hence, the only terms that appear in the derivatives are the marginal utilities in which b and τ directly appear. To characterize these terms, define the average marginal utility of consumption while unemployed as

$$\mathbb{E}u'(c_t^u) = \frac{1}{D} \sum_{t=0}^{T-1} \prod_{i=0}^t (1 - e_i) u'(c_t^u)$$

and the average marginal utility of consumption while employed as

$$\mathbb{E}v'(c_t^e) = \frac{1}{T - D} (e_0 T v'(c_0^e) + \sum_{t=1}^{T-1} [\prod_{i=1}^t (1 - e_{i-1})] e_t (T - t) v'(c_t^e)).$$

³⁶It is easy to show that V_t is concave because there is no uncertainty following re-employment; however, U_t could be convex. Lentz and Tranaes (2005) and Chetty (2008) report that non-concavity never arises in their simulations for a broad range of plausible parameters. Therefore, we assume that U_t is globally concave in the parameter space of interest and use first-order conditions to identify the optimal level of benefits.

Some algebra yields:

$$\begin{aligned} e_0 \frac{\partial V_0}{\partial b} + (1 - e_0) \frac{\partial U_0}{\partial b} &= D \mathbb{E}u'(c_t^u) \\ e_0 \frac{\partial V_0}{\partial \tau} + (1 - e_0) \frac{\partial U_0}{\partial \tau} &= (T - D) \mathbb{E}v'(c_t^e) \end{aligned}$$

The government's budget constraint implies that $\frac{d\tau}{db} = \frac{D}{T-D}(1 + \frac{T}{T-D}\varepsilon_{D,b})$. Combining these expressions, it follows that

$$\frac{dJ_0}{db} = D \mathbb{E}u'(c_t^u) - D \mathbb{E}v'(c_t^e) \left(1 + \frac{T}{T-D}\varepsilon_{D,b}\right)$$

As in the static model, we normalize the welfare gain from a \$1 increase in the size of the government insurance program by the welfare gain from raising the wage bill in the high state by \$1 to obtain

$$\begin{aligned} M_W(b) &= \frac{\frac{dJ_0}{db}(b)/D}{\frac{dJ_0}{dw}(b)/(T-D)} \\ &= \frac{1}{\mathbb{E}u'(c_t^e)} \left\{ \mathbb{E}u'(c_t^u) - \mathbb{E}v'(c_t^e) \left(1 + \frac{T}{T-D}\varepsilon_{D,b}\right) \right\} \\ &= \frac{\mathbb{E}u'(c_t^u) - \mathbb{E}v'(c_t^e)}{\mathbb{E}v'(c_t^e)} - \frac{1}{1 - D/T} \varepsilon_{D,b}. \end{aligned} \tag{31}$$

This expression coincides with the formula for $M_W(b)$ from the static model in (13) with two changes. First, the fraction of time spent unemployed is measured over the entire life of the agent (D/T) instead of in a single period. Correspondingly, the relevant elasticity is $\varepsilon_{D,b}$ rather than $\varepsilon_{1-e,b}$. Second, and more importantly, the gap in marginal utilities that enters the formula is the difference between the *average* marginal utility when employed and unemployed. This is because the average value of a marginal dollar of UI benefits depends upon the mean marginal utility across all the periods over which the agent is unemployed during his life. Similarly, the average cost of raising the UI tax depends upon the cost of losing \$1 during all the periods over which the agent is employed.

Sufficient Statistics Implementation. Equation (31) can be implemented using each of the three approaches described above with modifications to account for the fact that one must measure the gap in average marginal utilities. To implement the consumption-smoothing approach, one must estimate the *mean* consumption drop between periods when the agent is employed and unemployed, $\frac{\Delta \bar{c}}{\bar{c}^e} = \frac{\mathbb{E}c_t^u - \mathbb{E}c_t^e}{\mathbb{E}c_t^e}$, as the gap in expected marginal utilities when utility is not state-dependent is $\frac{\mathbb{E}u'(c_t^u) - \mathbb{E}u'(c_t^e)}{\mathbb{E}u'(c_t^e)} = \gamma \frac{\Delta \bar{c}}{\bar{c}^e}$. Identifying $\frac{\Delta \bar{c}}{\bar{c}^e}$ is conceptually analogous to identifying $\frac{\Delta c}{c}$ in the static model. However, it may not directly correspond to the difference between consumption immediately before and after unemployment, as measured by Gruber (1997) and others, if consumption trends substantially over the lifecycle. The extent to which $\frac{\Delta \bar{c}}{\bar{c}^e}$ differs from estimates using Gruber's approach has not yet been investigated empirically and requires further work.

Similarly, to implement the liquidity vs. moral hazard approach, one must estimate the impacts of *annuities* to recover expected marginal utilities over the unemployment spell in a dynamic model. Again, this is conceptually no different than estimating the impact of lump sum cash grants, but

may be harder to implement empirically. In practice, Chetty (2008) translates his estimates of the impacts of cash grants on unemployment durations into the impacts of annuities by making assumptions about discount rates. Finally, Shimer and Werning (2007) show that their reservation wage approach goes through in a dynamic model provided that utility has a CARA specification that eliminates income effects.

The general lesson from the analysis of the dynamic model is that one does not need to fully characterize all the margins through which agents may respond to shocks to calculate the marginal welfare gains of social insurance. Even in complex dynamic models, the calculation of welfare gains can be distilled to two parameters: the gap in average marginal utilities and the elasticity that enters the government’s budget constraint $\varepsilon_{D,b}$. There is no need to identify additional structural parameters such as the tightness of the borrowing constraint (L) or the cost of job search ψ to calculate $\frac{dW}{db}$. Similarly, one can show that the formula in (31) is robust to a variety of other extensions as well. For example, Kaplan (2010) shows that many young unemployed workers move back in with their parents as a method of consumption smoothing. He shows that incorporating this margin of adjustment into a structural model has significant effects on consumption fluctuations and the benefits of unemployment insurance. However, this margin is automatically accounted for in the formulas derived above. Mathematically, the option to move back home is simply another choice variable for agents and thus has no impact on (31). Intuitively, optimizing agents account for the option to move back home in all their choices. Thus, empirically observed consumption patterns, liquidity effects, and reservation wages all already incorporate this margin.

Chetty (2006b) establishes the validity of (31) at its most general level by analyzing a dynamic model where transitions from the good state to the bad state follow an arbitrary stochastic process. Agents make an arbitrary number of choices and are subject to arbitrary constraints. The choices could include variables such as reservation wages, savings behavior, spousal labor supply, or human capital investments. Chetty shows that (31) holds in this environment as long as agents’ choices maximize private welfare.³⁷ This is the critical assumption underlying (31). In the rest of this section, we discuss a series of important externalities that violate this assumption.

3.3.2 Externalities on Private Insurers

Public insurance is motivated by market failures of the types discussed in Section 2. However, most of the literature on optimal public insurance simply assumes that private markets do not provide any insurance against risks rather than modelling the underlying sources of the market failure. This assumption – which we made when deriving (31) above – is a convenient technical simplification but

³⁷A related point is that the same formula also holds in the context of other social insurance programs where the structure of the positive model differs. For instance, in the context of health insurance, one may choose the intensity of care *after* getting sick rather than controlling ex-ante effort. The formula in (31) holds in such a model with an appropriate redefinition of the average marginal utilities and the elasticity that enters the planner’s budget constraint. Chetty (2006b) considers an example of “tenure review” that is formally analogous to this case.

has important consequences for optimal social insurance.³⁸ If there is a formal market for private insurance, agents' choices will no longer maximize *total* private surplus (including the agent and private insurer) because the private insurance contract will distort choices. As a result, marginal changes in agents' choices will have fiscal externalities on private insurers. For example, suppose that part of health expenditures are covered by public insurance and part by private insurance. When the government raises health insurance benefits, agents will spend more on health care, and this increased expenditure will raise costs for *both* the government and the private insurer. This added fiscal externality on the private insurer is a first-order effect that reduces the marginal value of insurance and is not accounted for in (31).

Before discussing how these fiscal externalities affect optimal benefits, it is worth emphasizing that only private insurance contracts that generate moral hazard alter the formulas for optimal public insurance. Private insurance that does not generate moral hazard – such as informal insurance between relatives that is well monitored – has no impact on the original formula in (31). To see this, suppose that the agent can transfer b_p between states at a cost $q(b_p)$, so that increasing consumption by b_p in the low state requires payment of a premium $\frac{1-e}{e}b_p + q(b_p)$ in the high state. As long as the level of b_p is chosen to maximize utility, it is simply another choice variable in Chetty's (2006b) framework, and thus has no impact on the optimal social insurance formula. The effects of any such insurance arrangement are automatically embedded in the empirically observed consumption drop and other parameters that enter the formula.

There is relatively little work analyzing optimal social insurance with private insurance that induces moral hazard. Golosov and Tsyvinski (2007) analyze optimal government policy in a model in which agents can purchase insurance from multiple private providers. They rule out market failures due to adverse selection by assuming that agents can sign contracts before private information is revealed. However, the private market still does not achieve an efficient outcome because of a “multiple dealing” externality across firms, originally discussed by Pauly (1974). Intuitively, each firm does not take into account the fact that its provision of insurance distorts the agent's behavior, thereby affecting other insurers' budgets and leading to over-provision of insurance in the decentralized private market equilibrium. Golosov and Tsyvinski show that the government can raise welfare by imposing a corrective tax that countervails the multiple dealing externality. They also show that the provision of public insurance partially crowds out private insurance and may reduce welfare because it leads to further over-provision of insurance. Using numerical calibrations, they demonstrate that these effects could be quite large quantitatively.

While these results demonstrate that endogenous private insurance could reduce the scope for government intervention in insurance markets, the implications of their analysis for optimal benefit levels in practice are less clear. Golosov and Tsyvinski rule out many of the rationales for publicly provided insurance we discussed above, such as adverse selection, by assumption. In addition, their

³⁸One potential rationale for this approach is that private markets are essentially at a corner of providing zero insurance against certain risks (e.g., unemployment), and thus marginal changes in public insurance can be evaluated under the assumption that there is no private insurance. Unfortunately, this is not an accurate description of insurance markets in many important applications such as health.

numerical calibrations rest on strong assumptions about the structure of the underlying model that are not directly grounded in empirical evidence.

Chetty and Saez (2010) attempt to connect the theory to the data more explicitly by deriving formulas for optimal benefits in terms of empirically estimable parameters with endogenous private insurance. They first analyze a case in which agents are homogeneous and purchase insurance from a private firm. Because there is no market failure in this model, if agents in the private sector optimize perfectly, the marginal gain from government intervention is strictly negative (Kaplow 1991). When the private sector does not offer the optimal level of insurance – e.g., because of behavioral biases of the types discussed in Section 3.3.5 below – there is a potential role for government intervention. Chetty and Saez assume that the planner maximizes the agent’s true expected utility (which they assume is not state-dependent) and show that the marginal welfare gain from public insurance b can be expressed as

$$M_W(b) = (1 - r) \left[\frac{u'(c_l) - u'(c_h)}{u'(c_h)} - \frac{\varepsilon_{1-e,b}}{e} \frac{1 + b_p/b}{1 - r} \right] \quad (32)$$

where b_p is the private insurance benefit level in equilibrium and $r = -\frac{db_p}{db}$ measures the extent to which private insurance is crowded out by public insurance. When $b_p = 0$, this formula reduces to (13) with state-independent utility. When $b_p > 0$, two additional terms enter the formula. First, the marginal welfare gain is scaled down by $(1 - r)$ because \$1 more of public insurance raises total insurance $(b + b_p)$ by only $\$(1 - r)$. This effect rescales but does not change the sign of $M_W(b)$ and thus does not affect the optimal public benefit level b^* . The second and more important effect of endogenous private insurance is captured by the added term $\frac{1+b_p/b}{1-r}$ that amplifies the elasticity. This term reflects the fiscal externality that expanding public insurance has on private insurers. When the agent reduces e in response to a \$1 increase in b , it not only has a cost to the government proportional to $\varepsilon_{1-e,b}$ but also a cost to the private insurer proportional to $\varepsilon_{1-e,b}b_p/b$. This externality effect reduces the optimal benefit level beyond what one would have calculated based on the formula in (31) that ignored private insurance.

Chetty and Saez implement (32) in the context of unemployment insurance and health insurance to illustrate their approach. For unemployment – where the only form of private insurance is severance pay – Chetty and Saez estimate $\frac{b_p}{b} = 0.2$ and $r = 0.14$ using cross-state variation in UI benefits. Because the share of private UI benefits and crowdout effects are small, endogenous private insurance has small effects on the optimal UI benefit level with plausible elasticities. For health, the share of private insurance is much larger ($\frac{b_p}{b} = 0.89$), as is the degree of crowdout. Cutler and Gruber (1996) estimate that a \$1 increase in public health insurance benefits reduces private insurance benefit levels by 50 cents, implying $r = 0.5$. As a result, accounting for endogenous private insurance reduces the marginal welfare gains of an aggregate health insurance expansion by more than an order of magnitude according to (32).

The shortcoming of the formula in (32) is that it is sensitive to the sources of private market failures. For instance, Chetty and Saez extend their baseline analysis to a case where there are heterogeneous agents who have private information about their risk types. This generates adverse

selection and partial private insurance provision by the market. In this setting, the formula for $M_W(b)$ has additional terms relative to (32) because it has the added benefit of pooling risks across individuals of different types through a mandate. Hence, (32) is not a robust “sufficient statistic” formula because it is sensitive to the structure of the positive model. Developing formulas that are robust to the underlying sources of private market failures and can be implemented empirically is among the most important priorities for future research on social insurance.

3.3.3 Externalities on Government Budgets

Fiscal externalities can also arise when the government itself provides multiple types of social insurance or levies taxes. For instance, expansions in disability insurance may reduce the probability that a worker with a high disutility of work chooses to search for a job and claim unemployment insurance. A more general source of fiscal externalities is taxation. Any reduction in labor supply induced by social insurance programs will have a negative fiscal externality on the government by reducing income tax revenue.

It is useful to divide fiscal externalities into two categories. The first are mechanical externalities that arise because the choice insured by the program (e) directly affects other parts of the government’s budget, such as tax revenue. The second are indirect fiscal externalities that arise because of *other* behavioral responses unrelated to e itself. For instance, an increase in UI benefits may induce agents to save less, which could reduce revenue from capital gains taxes.

To see how such fiscal externalities affect (31), let us return to the static model and add an initial period, $t = 0$, in which the agent chooses how much to save, which we denote by A , before he faces the risk of unemployment. Let Z denote the agent’s wealth at the beginning of period 0, which we take as fixed. In period 0, consumption is given by $c_0 = Z - A$. Let τ_A denote the tax levied on savings A (e.g. a capital gains tax) and τ_e denote an income tax levied on employed individuals. Tax revenue $R = \tau_A A + \tau_e e$ is rebated to the agent as a lump sum in the high state, so that $c_h = A + w_h - \tau(b) + R$.

The agent chooses effort e to maximize his expected utility taking the parameters of the tax system as fixed:

$$\max_{e,A} V(e, A) = v(Z - A) + ev(A + w_h - \tau(b) + R) + (1 - e)u(A + w_l + b) - \psi(e)$$

The social planner’s objective is to choose the benefit level b that maximizes the agent’s expected utility:

$$\begin{aligned} \max_b W(b) &= v(Z - A) + ev(A + w_h - \tau(b) + R) + (1 - e)u(A + w_l + b) - \psi(e) \\ \text{s.t. } e &= e(b) \text{ and } A = A(b) \text{ and } R = \tau_A A + \tau_e e \end{aligned}$$

Algebra analogous to that above yields:

$$M_W(b) = \frac{u'(c_l) - v'(c_h)}{v'(c_h)} - \frac{\varepsilon_{1-e,b}}{e} \left(1 + \frac{\tau_e}{b}\right) + \frac{1}{1-e} \frac{A}{b} \tau_A \varepsilon_{A,b}$$

This equation differs from the original formula in (13) in two ways. The first new term, $\frac{\tau_e}{b}$, reflects the mechanical externality arising from the fact that reductions in e reduce income tax revenue. The magnitude of this effect is proportional to $\frac{\varepsilon_{1-e,b}}{e}$. Accounting for this mechanical externality does not require estimating any additional parameters; one must simply take the income tax rate into account when calculating the impact of changes in e on the government budget. The second new term is proportional to $\tau_A \varepsilon_{A,b}$. This term reflects the indirect fiscal externality imposed by distorting the choice of savings A . Accounting for this indirect externality requires estimation of the additional elasticity $\varepsilon_{A,b}$, e.g., as in Engen and Gruber (2001).³⁹

Empirically, an important source of indirect fiscal externalities are behavioral responses that affect takeup of other social insurance programs. Autor and Duggan (2003) provide empirical evidence on the interaction between unemployment and disability insurance in the U.S. Using local employment shocks instrumented by industry shares, they show that more individuals exit the labor force and apply for DI when they are laid off when DI benefits are raised. As a result, unemployment rates rise less when disability insurance is expanded and, correspondingly, one would expect UI benefit payments to fall. Of course, the government also loses revenue by collecting fewer taxes from individuals who would have worked if the DI benefit level were lower.

Borghans, Gielen, and Luttmer (2010) quantify the magnitude of such fiscal externalities using administrative data from the Netherlands. They use a regression discontinuity design to show that reducing the generosity of DI benefits increases reliance on other forms of social insurance. They estimate that every \$1 saved in DI benefits via a benefit cut leads to approximately 50 cents of additional expenditure on other social insurance programs. Part of this cost is offset by increased tax revenue, but it is clear that accounting for such fiscal externalities is critical in designing optimal social insurance policy. More generally, one should ideally analyze social insurance and tax policies in a unified framework rather than optimizing each program (UI, DI, health insurance, etc.) separately.

3.3.4 Other Externalities

Agents in the private sector may have direct non-pecuniary externalities on each other independent of the fiscal channels discussed above. In Section 2.3, we discussed how such externalities – such as the spread of contagious diseases – could provide a motivation for social insurance. In this section, we discuss how such externalities affect the optimal design of social insurance policies.

While any externality would affect (31), the literature on optimal social insurance has focused on two types of externalities in particular. The first are social multiplier effects, which refer to the idea that one individual’s choices may affect the choices of those around him. For instance, an increase in disability benefits may induce some agents to stop working, which in turn may increase the value of leisure for their peers. The second are congestion externalities, which arise because of constraints that make one agent’s behavior change the returns to effort for another agent.

³⁹Consistent with the results in section 3.3.1, the distortion in A has no impact on the formula if $\tau_A = 0$; it is only because savings are taxed that this behavioral response matters.

For instance, in the context of unemployment, agents compete for a limited number of positions when jobs are rationed. In this setting, changes in UI benefits can have different effects at the macroeconomic level relative to the microeconomic level.

Theoretical work on social multipliers effects and social insurance has shown that the aggregate effects of insurance on behavior could be significantly larger than microeconomic estimates of a single individual's behavioral responses to a change in his marginal incentives. For instance, Lindbeck, Nyberg, and Weibull (1999) present a benchmark model of how social norms about labor supply change the impacts of the welfare state on the economy. In their model, individuals' taste for work depends not only upon their private disutility of work but also the fraction of individuals who rely on social support rather than work to make a living. They demonstrate that in this environment, small changes in policy can lead to dramatic, discontinuous changes in the size of the welfare state by shifting the equilibrium. For instance, the macroeconomic effect of an increase in disability benefits on labor supply, ε^M , could be much larger than the microeconomic elasticity ε for any single individual. Intuitively, when social insurance benefits are increased, the direct effect on each agent's behavior is amplified by the feedback mechanism of the change in the norm. Lindbeck, Nyberg, and Weibull argue that these non-linear responses could explain why some developed countries (e.g. Scandinavia) have much larger social insurance systems than others (such as the U.S.).

How do such social multipliers affect the formula for optimal benefits in (31)? If social multipliers affect behavioral responses but do not directly affect utility, (31) holds, but ε must be replaced by the macro elasticity ε^M . This is because the aggregate response ε^M is what determines how an increase in benefits affects the government budget. If the social externalities enter utility directly – e.g., if one agent not working increases the utility of leisure for others agents – then an additional term enters (31) to capture the impact of this externality on welfare, as the envelope conditions used to derive (31) do not take account of this effect.⁴⁰

Empirical evidence lends support to the presence of such social multipliers. For instance, Lalive (2003) presents evidence that an extension of unemployment benefits to workers over age 50 in Austria affected the labor supply of their peers below age 50, although his results rely on relatively strong identification assumptions. Clark (2003) uses data on subjective well-being from a panel of households in the United Kingdom and shows that unemployed individuals report higher levels of happiness when the local unemployment rate is higher, while employed individuals report lower levels of satisfaction. Bertrand, Luttmer, and Mullainathan (2000) show that individuals are more likely to take up welfare when living in neighborhoods with a large population of residents who speak their own language, suggesting that there are network effects in welfare takeup. A limitation of this analysis is that it relies purely on cross-neighborhood comparisons and thus rests on relatively strong identification assumptions. Chetty, Friedman, and Saez (2012) present evidence of social spillovers via learning using panel data. They show that individuals are more likely to *change* their earnings behavior in response to the incentives created by the Earned Income

⁴⁰See Kroft (2008) for an analysis of optimal UI with social spillovers in takeup decisions.

Tax Credit in the U.S. when they move to a neighborhood where other individuals respond to the program. Unfortunately, none of these studies provide estimates of how the micro impacts of policy (ε) changes differ from their macro impacts (ε^M).

While social multiplier effects amplify elasticities and potentially reduce the optimal level of social insurance, congestion externalities have the opposite effect. Landais, Michaillat, and Saez (2012) analyze optimal unemployment insurance in a job search model with rationing. A standard matching function maps the number of vacancies and job openings to the equilibrium level of unemployment. They assume that wages are rigid, so that when the economy is hit by a negative productivity shock, jobs are rationed in the sense that the labor market does not clear even if workers exert arbitrarily high search effort. In this environment, job seekers have a negative search externality on each other: each individual tries to outrun his peers to find a job, leading to excess search effort via a standard rat-race mechanism. Landais, Michaillat, and Saez demonstrate that this externality has two effects on (31). First, it reduces the macro elasticity of unemployment with respect to the benefit rate relative to the micro elasticity. Again, it is the macro elasticity that enters (31) for the reasons described above. Second, the negative job search externality creates an added benefit from raising UI benefits, because when each individual searches less, he has a positive spillover effect on other job seekers. Because job rationing is more severe in recessions, Landais, Michaillat, and Saez conclude that the optimal level of UI benefits should be countercyclical. They show that the micro and macro elasticities along with the other parameters in (31) are sufficient statistics to calculate the optimal level of benefits in a relatively general environment.⁴¹

Motivated by the question of whether unemployment benefits should vary over the business cycle, a recent empirical literature has investigated whether elasticities of unemployment durations with respect to benefit levels vary with labor market conditions. Most of the existing evidence uses identification strategies that are more likely to recover partial-equilibrium micro elasticities rather than general-equilibrium macro elasticities. This is because the need for a control group often makes it easier to identify partial equilibrium effects using quasi-experimental methods. Schmieder, von Wachter, and Bender (2012) show that UI benefit extensions have smaller effects on unemployment exit hazards in recessions than in booms using data for Germany, implying that the moral hazard effect of UI is smaller at the micro level in recessions. Landais (2011) uses a regression kink design to show that changes in UI benefit levels have similar effects on unemployment durations in recessions and booms in the United States. Kroft and Notowidigdo (2011) present evidence that the moral hazard effects of changes in UI benefits are smaller in recessions, while the consumption-smoothing benefits are roughly constant, although their estimates are somewhat imprecise due to a lack of power.

Crepon et al. (2012) present evidence that the macro elasticity ε^M is much smaller than the micro elasticity ε in weak labor markets because of congestion effects. They run a randomized experiment that provides job placement assistance to a large group of individuals in a given labor

⁴¹Landais, Michaillat, and Saez focus on benefits levels rather than the duration of benefits. However, the typical policy response to a recession is an extension in the duration of benefits rather than an increase in the level of benefits. Understanding why this is the case is an interesting open area for further work.

market. They find that the individuals who did *not* receive assistance have significantly *lower* probabilities of finding a job in weak labor markets when many of their peers receive job placement assistance. This constitutes direct evidence of congestion effects in the labor market and, based on the models discussed above, suggests that the optimal UI benefit level is higher when labor markets are weak.

3.3.5 Imperfect Optimization

Another important reason that agents' choices may not maximize total private surplus is failures of optimization. As we noted in Section 2.3 above, if agents optimized perfectly, we would observe very small consumption fluctuations $\frac{\Delta c}{c}$ for temporary shocks such as unemployment. Hence, (31) would imply that the welfare gains of social insurance are small. It is tempting to simply plug in a statistic such as the observed (large) values of $\frac{\Delta c}{c}$ and calculate the welfare gains of social insurance by applying (31) even if agents do not optimize. Unfortunately, (31) is not a valid formula for optimal benefits when agents do not optimize. This formula is valid only if agents optimize, and we know that if they optimized $\frac{\Delta c}{c}$ would not be large. Because a biased agent's choices do not maximize his own utility, the envelope conditions exploited to derive (31) no longer hold. Hence, one must modify the formulas to allow for imperfect optimization to obtain an internally consistent understanding of optimal social insurance.⁴²

At a broad level, there are two conceptual challenges in accounting for imperfect optimization. First, conducting welfare analysis requires recovering the individuals' true preferences, which is challenging when one cannot rely on the standard tools of revealed preference used above. One approach to solving this problem is to posit a structural model of behavioral failures, such as hyperbolic discounting (Laibson 1997), and estimate the parameters of that model to analyze optimal policy. Another approach, which is closer in spirit to the sufficient statistic methods we have focused on here, is to derive formulas for the welfare consequences of social insurance that do not rest on a specific positive model of optimization failures (Bernheim and Rangel 2009, Chetty, Looney, and Kroft 2009). The chapter by Bernheim in this volume discusses welfare analysis in behavioral models in greater detail. Second, from an empirical perspective, it is unclear how to systematically distinguish mistakes from unobserved attributes such as risk aversion or private information about expected losses under an insurance contract. While we are able to identify specific examples where we can reject the null of perfect optimization, it is less clear how we can quantify the degree to which individuals are biased, which is necessary for implementing corrective policies.

There is relatively little work on social insurance in behavioral models. DellaVigna and Paserman (2005) estimate a model of job search with impatient agents and show empirically that agents who are more impatient – as measured by variables that quantify tradeoffs between immediate and

⁴²This point illustrates a general weakness of the sufficient statistic approach, which is that the assumptions used to derive the formula are never explicitly tested because the parameters of the model are never fully identified. See Chetty (2009) for further discussion of these issues.

delayed payoffs – search less for jobs. In their model, increasing UI benefits would have additional costs because it would further reduce effort below the optimum. Fang and Silverman (2009) estimate a similar model of time inconsistency in the context of welfare program participation and estimate the model’s structural parameters. They then use their model to show that policies that limit the number of months for which individuals are eligible for social transfers could raise welfare. These papers are examples of the first approach to welfare analysis in behavioral models described above – positing a structural model and using it to analyze the welfare consequences of policy changes.

Spinnewijn (2010) is an example of the second approach to behavioral welfare analysis. He generalizes the sufficient statistic formulas for optimal benefit levels derived above to an environment in which agents are overoptimistic about their probability of finding a job. Spinnewijn shows that it is crucial to distinguish between two types of biases in beliefs: baseline bias, which is misestimating the probability of finding a job holding search effort fixed, and control bias, which is misestimating the impact of increased search effort on the probability of finding a job. Baseline bias has no impact on the formula in (31) because it does not affect the agent’s behavioral responses to benefit changes or his utility. In contrast, control bias introduces a new term in the formula that arises from the fact that the agent does not set e at its true optimum. As a result, changes in b have a first-order effect on utility by inducing changes in e . For instance, an agent who is control-pessimistic will undersupply effort in equilibrium, and increases in benefits will lower welfare by distorting e downward even further. Spinnewijn presents empirical evidence that agents are indeed control-pessimistic in practice, suggesting that calculations of optimal benefits based on the traditional formula in (31) will overstate the welfare gains of insurance.

The limitation of these papers is that each one characterizes the implications of a specific type of bias for policies. The challenge for research on social insurance in behavioral models is finding a framework that incorporates a broad set of biases yet offers empirically implementable results for optimal policy. Further work along these lines is a challenging but promising direction for future research.

3.4 Other Dimensions of Policy

While much of the literature connecting theory to data has focused on identifying the optimal level of social insurance benefits, there are many other important questions in the design of insurance programs. In this section, we briefly review three areas that have received attention in recent work: mandated savings accounts, increasing program takeup rates, and changing the path or duration of benefits.

3.4.1 Liquidity Provision and Mandated Savings Accounts

Our analysis thus far has focused exclusively on a social insurance system that transfers money from individuals in the high state to those in the low state via taxes and transfers. However, as we discussed above, there would be little need for social insurance against temporary shocks

such as unemployment if individuals had access to sufficient liquidity while unemployed. Recent research has therefore analyzed policies that provide liquidity to unemployed individuals or force them to build a buffer stock via mandated savings accounts rather than transferring resources across individuals. Intuitively, if the motive for social insurance benefits is a lack of liquidity, the optimal tool to correct this problem may be to directly provide liquidity rather than to provide state-contingent transfers.

Shimer and Werning (2008) analyze optimal unemployment insurance in a model where agents can freely save and borrow using a riskless asset while unemployed. Under constant absolute risk aversion preferences, they demonstrate that the optimal policy provides free access to liquidity and a relatively low level of unemployment benefits. Chetty (2008) numerically compares the welfare gains from the provision of zero-interest loans with the welfare gains of increasing unemployment benefits in a search model calibrated to match empirical estimates of liquidity and moral hazard effects. Consistent with Shimer and Werning's intuition, Chetty's simulations show that the provision of loans yields large welfare gains and greatly reduces the gains from raising unemployment benefit levels. While these results highlight the potential value of liquidity provision as a policy tool, the costs of providing liquidity (e.g., due to default risk) are not modeled in these studies. Hence, these studies do not shed light on the optimal combination of loans and unemployment insurance benefits. Identifying the optimal combination of these two policies in an environment where both policies have social costs is an important open question.

An alternative approach to providing liquidity is to directly address agents' failure to build buffer stocks by mandating savings prior to unemployment. Feldstein and Altman (2007) propose a system of UI savings accounts in which individuals are required to save a fraction of their wages in accounts designated to be used only in the event of unemployment spells. If they become unemployed, individuals would draw upon these accounts at standard benefit rates. If individuals run out of money in their savings account, they would automatically draw benefits from the state UI system, which would continue to be financed by a (smaller) payroll tax. Any remaining balance at the point of retirement would be refunded to the individual. The benefit of such a system is that individuals' incentives to search for a job while unemployed are not distorted by the provision of UI benefits, because each extra \$1 they use in UI benefits leaves them with \$1 less of wealth in retirement, provided that they do not fully deplete their savings account balance. Using data from the Panel Study of Income Dynamics, Feldstein and Altman show that less than half of UI benefits would be paid to individuals who hit the corner of a zero account balance, implying that incentives could potentially be improved for many people. Orszag and Snower (2002) present calibrations showing that the impacts of such a system on unemployment durations could be quite large.

Stiglitz and Yun (2005) present a more formal argument for the optimality of UI savings accounts. They analyze a model in which individuals pay money into a pension system that is taken as exogenous. They prove that permitting individuals to draw down these pension assets in the event of adverse shocks raises welfare if individuals are borrowing constrained. The intuition is that allowing individuals to borrow against retirement savings relaxes borrowing constraints and permits better consumption smoothing at a lower efficiency cost than state-contingent transfers.

An important limitation of existing theoretical work on optimal liquidity provision and mandated savings accounts is that they all analyze models with agents who optimize perfectly. Mandated savings can only be justified if agents suffer from biases such as impatience, as the government is simply restricting the choice set available to agents. However, if agents are impatient, it is not clear that they will fully internalize the benefits of having more wealth in retirement if they draw less UI benefits when they are young. Hence, in a model that justifies mandated savings, such policies could well have efficiency costs similar to traditional social insurance systems. Countervailing this effect, there may be other biases – such as the increased salience of UI savings accounts relative to eligibility for UI benefits – that may reduce distortions when agents do not optimize perfectly. These intuitions illustrate that imperfect optimization should be a central element of future work on optimal liquidity and insurance policies.

Empirical evidence on the impacts of UI savings accounts is scarce because few governments have implemented such systems. Hartley, van Ours, and Vodopivec (2011) present suggestive evidence that a UI savings account program introduced in Chile reduced unemployment durations. Their analysis uses individuals who endogenously choose to opt into a traditional unemployment insurance program as a comparison group. Their conclusions therefore rest on stronger identification assumptions than most of the empirical studies discussed above. Further work using quasi-experimental designs to study the impacts of mandated savings accounts is needed to fully understand their impacts.

3.4.2 Imperfect Takeup

Another important dimension of social insurance policy is program participation. In the analysis above, we assumed that all individuals who are eligible for a social insurance program automatically participate in the program. In practice, takeup rates for most programs are often well below 100% (Currie 2006). For instance, only 75% of individuals eligible for UI actually receive UI benefits when laid off. Incomplete takeup raises two questions for optimal program design. First, should the government seek to increase takeup rates, and if so what methods should be used to do so? Second, is the optimal level of benefits different when only some agents take up the benefit?

The social value of raising takeup rates – e.g., by reducing the costs of applying for a program – depends upon the value agents who choose not to participate get from the program relative to the takeup utility cost incurred by all agents. Imperfect takeup could in principle be socially desirable if the agents who choose not to take up value the program significantly below its social cost. For example, Nichols and Zeckhauser (1982) propose a stylized model in which implementing a hurdle to claim benefits raises social welfare by removing individuals who do not value the program highly from the pool of recipients. In this model, the costs of taking up a program can serve as a screening mechanism, allowing public funds to be targeted at the subgroups who are in greatest need of assistance (e.g., those with the least liquidity).⁴³ However, the screening induced by hurdles

⁴³Such screening through costly ordeals is only desirable if the ordeal is highly effective at separating agents with high vs. low valuations of the program. If the screening mechanism is not powerful, simply reducing benefit levels

to take up need not necessarily operate in this beneficial manner. For instance, suppose individuals who do not participate are uninformed about the program’s existence or make optimization errors. In this environment, policies that increase takeup rates could raise welfare, because they may bring the individuals who would value social support most highly into the system. The nature of the marginal agents who do not take up social insurance is ultimately an empirical question, and the answer could vary across programs and environments.

Empirical evidence on the determinants of takeup shows that imperfect takeup is driven by a mix of factors. Anderson and Meyer (1997) present evidence that individuals are more likely to take up UI benefits when net benefit levels are higher. Black et al. (2003) present experimental evidence that individuals are more likely to drop out of the UI system when they receive a mailing announcing that they must attend a training program in order to remain eligible for benefits. These results suggest that takeup is based on a rational cost-benefit conclusion and imply that the individuals who do not take up may be those who value the benefit the least. In contrast, Ebenstein and Stange (2010) show that the shift from in-person to telephone registration for UI benefits, which significantly reduced the time costs of takeup, had no impact on UI takeup rates. Bhargava and Manoli (2011) present experimental evidence showing that simplifying tax forms increases takeup of the Earned Income Tax Credit. These results raise the possibility that some individuals who do not take up may actually benefit significantly from the program despite getting screened out under the current program design. What remains unclear is the average valuation of benefits for those who do not take up, which is a key parameter for determining whether resources should be invested in increasing takeup rates. Developing methods to estimate this parameter and calculate the marginal welfare gains from increasing takeup is a promising area for future research.

How does incomplete takeup affect the optimal design of other parameters of the social insurance system? Kroft (2008) analyzes the optimal UI benefit level in a model with imperfect takeup. He generalizes the sufficient-statistic formulas derived above to a model with heterogeneous takeup costs and rational agents. The formula in (31) changes in two ways in this setting. First, $\mathbb{E}u'(c_t^u)$ must be computed for the subgroup of individuals who takes up benefits rather than all individuals in the low state. Second, the elasticity parameter that is relevant is the sum of the behavioral elasticity $\varepsilon_{1-e,b}$ and the takeup elasticity, as increases in benefit rates raise expenditures both through traditional margins and by raising the number of individuals claiming benefits. In general, accounting for endogenous takeup lowers the optimal benefit rate relative to computations that apply (31) to the subsample of individuals who take up benefits.

3.4.3 Path of Benefits

Most shocks have a variable time span. For instance, spells of work injury, disability, and unemployment all have uncertain durations. Although many health shocks are one-time events, there are some chronic conditions that generate expenditures over longer periods. In such settings, the

by an equivalent amount may yield larger welfare gains. See the chapter by Piketty and Saez in this volume for a discussion of related issues in the context of “workfare” models of taxation.

social planner can choose not just the level of benefits but also its path. At an abstract level, one can choose a different benefit level b_t in each period t during which the individual is in the low state. In practice, many social insurance systems have one or two-tiered benefit systems, in which benefits are provided at a constant level for a finite duration and then reduced or completely eliminated (e.g., the termination of UI benefits at 6 months in the U.S.). Our preceding analysis has focused on identifying the optimal level of b under the assumption that $b_t = b$ for all t . In this section, we discuss the small but growing body of work that has sought to characterize the optimal path of benefits.

The fundamental tradeoff in setting the path of benefits, originally described by Shavell and Weiss (1979), is again between consumption smoothing and moral hazard. Increasing benefits over time is desirable from a consumption smoothing perspective, as it provides the largest benefits to the agents who need it most. Intuitively, those who suffered from a long unemployment spells are likely to have depleted their assets and have a very high marginal utility of consumption. But increasing benefits over time is costly from an efficiency perspective, as it creates an incentive to “hold out” for higher benefits by prolonging one’s duration in the low state. Unfortunately, deriving formulas that map the relative magnitude of these two effects to quantitative predictions about the optimal path of b_t is challenging. The literature on the optimal path of benefits has thus been less successful in connecting theory to data because of the complexity of dynamic models and the high-dimensional nature of the problem. We therefore briefly describe some of the main theoretical and empirical results that bear on this problem, highlighting the scope for further work deriving sufficient statistic formulas in this area.

Shavell and Weiss (1979) characterize the optimal path of a benefits in a discrete time model in which agents make consumption and search effort choices in each period. When agents cannot save or borrow, the optimal path of benefits is declining because an upward sloping benefit path does not provide a smoother consumption path but creates moral hazard. When agents can save or borrow, Shavell and Weiss show that the optimal path could be upward or downward sloping depending upon the structural parameters of the model. Hopenhayn and Nicolini (1997) extend Shavell and Weiss’ analysis to allow for a wage tax after re-employment, but focus on the case with no saving or borrowing. They show that the optimal path of benefits is again declining in this environment, but that there is a significant welfare gain from levying a substantial tax upon re-employment.

Werning (2002) generalizes Hopenhayn and Nicolini’s analysis to a case with hidden savings, i.e. a more realistic environment where the planner can set benefit levels but cannot directly control consumption. Werning (2002) shows that when agents can save or borrow, optimal benefit levels may increase over time and are much lower in levels than those predicted by Hopenhayn and Nicolini’s analysis.⁴⁴ Shavell and Weiss, Hopenhayn and Nicolini, and Werning all study models where agents have fixed wages and only choose search effort. Shimer and Werning (2008) analyze a

⁴⁴The very high optimal replacement rates predicted by Hopenhayn and Nicolini’s model should essentially be interpreted as optimal *consumption* levels over the unemployment spell rather than optimal benefit levels.

standard reservation-wage model in which agents draw stochastic wage offers. They prove that the optimal path of benefits is constant ($b_t = b$) in a setting without liquidity constraints and CARA utility. Intuitively, under these assumptions, the agent’s problem while unemployed is stationary because his attitudes toward risk and incentives to find a job do not change as he depletes his assets. As a result, the moral hazard costs and consumption-smoothing benefits of a steeper benefit profile perfectly cancel out, yielding a constant optimal benefit path. Shimer and Werning argue that in cases where agents do not have access to perfect capital markets, the optimal government policy is to provide loans and constant benefits.

While this result is a useful benchmark, the optimal path of benefits could differ substantially under alternative assumptions about the structure of the positive model. In practice, agents are highly liquidity constrained while unemployed and governments tend not to offer loans to such agents, perhaps because of moral hazard problems in debt repayment. Moreover, there could be significant non-stationarities in responses to incentives over a spell, due both to selection in an environment with heterogeneity and changes in the cost of effort. For instance, individuals who remain unemployed after 6 months could be relatively elastic types but may also have a hard time controlling the arrival rate of offers, driving $\varepsilon_{1-e,b}$ down over the course of an unemployment spell. Such forces eliminate the stationarity that drives Shimer and Werning’s constant benefit result, as discussed in Shimer and Werning (2006). The quantitative magnitude of such effects is ultimately an empirical question whose answer could vary across environments and social insurance programs.

The empirical literature, which also focuses primarily on unemployment insurance, shows that the path of benefits has significant effects on agents’ behavior in a manner that is consistent with forward-looking decisions. Card and Levine (2000) analyze the impacts of a 13 week extension in UI benefits in New Jersey on unemployment durations. Their estimates imply that this extension would have raised durations on the UI system by 1 week had individuals been eligible for the extension from the start of their spells. Card, Chetty, and Weber (2007b) analyze a discontinuity in the Austrian UI benefit system based on work history that makes some individuals eligible for 20 weeks of benefits and others eligible for 30 weeks of benefits. Individuals eligible for 30 weeks of benefits have significantly lower job finding hazards even in the first five weeks of their unemployment spell, suggesting that the “holding out” moral hazard effect identified by Shavell and Weiss is indeed operative in the data. Lalive (2008) finds similar impacts of an age-based discontinuity in the Austrian unemployment system that varies eligibility much more dramatically, from 30 weeks to 209 weeks.

There is no evidence to date on the consumption-smoothing of extending benefits for a longer period due to the lack of high-frequency data on consumption. One approach to this problem may be to estimate liquidity effects at different points of a spell using variation in cash-on-hand from tax credits or other sources, as in Lalumia (2011). Investigating the path of asset decumulation and borrowing over the course of shocks would also shed light on the dynamic benefits of social insurance.

One approach for connecting the empirical evidence to theoretical models that make weaker assumptions may be to limit the set of benefit schedules that one considers. For instance, one

could analyze two-parameter systems in which benefits are paid at a constant rate b and then terminated at some date T or systems in which benefits follow a linear path $b_t = \alpha + \beta t$. It may be possible to derive analytic formulas in terms of the elasticities estimated in the empirical literature for the optimal two-parameter system. Given that governments are unlikely to implement highly variable benefits over time, such simplifications of the problem might not be very restrictive from a practical perspective.

4 Challenges for Future Work

Our review of the literature indicates that modern methods of connecting theory to data have proven to be quite fruitful, but much remains to be learned before one can draw strong conclusions about social insurance policy from this work. We conclude by summarizing some of the main outstanding challenges and open areas for future research.

First, there is a lack of empirical evidence on key parameters for many programs. Table 1 summarizes the current state of the literature for five major social insurance programs. For each program, we mark areas where we are aware of existing work on five topics: testing for selection (e.g., using positive correlation or cost curve tests), quantifying its welfare costs, estimating the benefits of social insurance (e.g., consumption smoothing or liquidity effects), estimating moral hazard costs (e.g., by studying behavioral responses to the program), and calculating optimal benefit levels (e.g., using sufficient statistic formulas). As discussed earlier, most empirical work to date on selection in insurance markets, and particularly its welfare costs, has concentrated on health insurance, with a few papers on annuities; noticeably missing is work on selection in unemployment insurance, disability insurance, or worker's compensation.⁴⁵ Research on government-provided insurance has focused primarily on measuring moral hazard costs rather than benefits. Unemployment insurance is one of the few programs in which both the benefits and costs of insurance have been studied extensively, with a corresponding well-developed literature on optimal policy.

A natural and quite valuable direction for further work is to fill in the empty boxes in Table 1 by applying the approaches surveyed here to these other important areas of social insurance. One particularly important program that has received relatively little attention in terms of measuring benefits and welfare consequences is disability insurance. Social insurance programs for health care – Medicaid and Medicare – are another area ripe for welfare analysis of optimal benefit levels. These programs have been the subject of a rich empirical literature examining their moral hazard impacts on health spending, and their benefits in terms of health and (to a more limited extent) consumption smoothing (see for Medicare e.g., Card et al. (2008, 2009), Finkelstein (2007), or Finkelstein and McKnight (2008) or for Medicaid e.g., Currie and Gruber (1996a, 1996b), Finkelstein et al. (2011) or Gross and Notowidigdo (2011)). Yet there have been virtually no attempts to translate these policy estimates into statements about welfare or optimal policy (Finkelstein and McKnight (2008)

⁴⁵The literature may have developed in this way because it is most natural study selection in markets where there is some private provision of insurance. In the U.S., which has been the primary focus on empirical work because of data availability, private insurance has been much more important in the health care market than the other markets.

is a notable, although highly stylized, exception). Another area in which modern tools connecting theory to data have not yet been applied is the analysis of social security programs. While there is a wealth of evidence of the reduced-form impacts of social security programs and a rich theoretical literature on optimal social security design, there is little work deriving robust formulas for optimal social security design in terms of empirically estimable parameters.

Second, the majority of the research we reviewed has focused on models with perfectly optimizing agents. However, as we argued above, it is difficult to rationalize some social insurance programs such as unemployment insurance in an environment with forward-looking agents because intertemporal consumption smoothing (“self insurance”) is quantitatively a very good substitute for insurance against temporary shocks. An important challenge for future work is to understand the consequences of behavioral economics for optimal insurance. For instance, if large liquidity and consumption-smoothing effects of unemployment insurance are driven by the fact that impatient workers postpone job search until they run out of cash, the gains from insurance may be much smaller than existing calculations suggest. Similarly, the optimal tools to correct behavioral biases may be very different from traditional insurance programs, which make state-contingent income transfers.

Third, most methods of connecting theory to data make restrictive assumptions that do not permit a full spectrum of general equilibrium responses. Results on measuring the welfare cost of selection have usually treated the offered contracts as fixed, and focused primarily on the distortions in prices created by selection. Likewise, formulas for optimal social insurance based on reduced-form elasticities all require that wage rates and other prices are unaffected by government policies. In practice, large-scale social insurance policies are likely to induce substantial supply-side responses, making it crucial to develop methods to account for such general equilibrium effects.

Fourth, and relatedly, the literature on optimal social insurance design we have reviewed has tended to focus on relatively specific design questions within the structure of existing programs, such as the optimal level of benefits given the existing time path of benefits or the optimal financing of benefits. It is generally silent on more global optimal design questions, such as those that would alter several features of a program simultaneously. One relatively under-studied design question is the choice of instrument used for government intervention. Governments can intervene in private insurance markets in at least three ways: direct public provision (e.g. Medicare, Social Security, and Unemployment Insurance), mandates on firms to provide insurance (e.g., Worker’s Compensation), and direct intervention in or regulation of private insurance markets (e.g., tax subsidies for employer provided health insurance or regulation of the allowable contracts and prices in the individual health insurance markets). The optimal choice of these instruments has not been well explored.⁴⁶

Fifth, the literature on motivations for social insurance and the optimal policy response are not yet well integrated. Existing empirical work on adverse selection analyzes selection and its

⁴⁶One exception is Summers (1989), who compares the effects of mandates and taxes in a static model without uncertainty. Extending Summers’ analysis to insurance markets would be an interesting direction for future research.

consequences *given the existing public policies*, such as tax subsidies to employer provided health insurance or publicly provided annuities through Social Security. This raises the question of whether selection would exist –and what its welfare consequences would be – in the absence of these public programs or under very different public programs than we currently have. Understanding, both theoretically and empirically, how public insurance programs or policies affect the existence and nature of adverse selection in the residual private markets is therefore an important area for further work. Similarly, in the optimal policy literature, fiscal externalities on private insurers and other parts of the government budget such as tax revenue could significantly reduce the gains from social insurance. Characterizing optimal public insurance in models that incorporate endogenous private market insurance failures is therefore essential to obtain a more precise understanding of the welfare consequences of social insurance.

Finally, a general conceptual challenge for the approaches we have discussed in this paper is that they are less informative about fundamental policy reforms, such as implementing universal health insurance, than local changes, such as changing copayment rates by 10%. This is because the reduced-form empirical estimates that the literature uses to make quantitative welfare predictions are identified using local variation around the currently observed environment. This makes it difficult to extrapolate out-of-sample to impacts of policies that have not been observed. The questions discussed above of how private markets would function in the absence of public insurance are one example of this problem. More generally, informing policy makers about significant reforms to social insurance programs will require a more global understanding of how these policies affect the economy. These issues are a fundamental challenge not just for the literature on social insurance but for empirical work more generally.

References

- Abaluck, J. and Gruber, J. (2011). Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *American Economic Review*, 101(4):1180–1210.
- Abbring, J. H., Chiappori, P.-A., and Pinquet, J. (2003a). Moral Hazard and Dynamic Insurance Data. *Journal of the European Economic Association*, 1(4):767–820.
- Abbring, J. H., Heckman, J. J., Chiappori, P.-A., and Pinquet, J. (2003b). Adverse Selection and Moral Hazard In Insurance: Can Dynamic Data Help to Distinguish? *Journal of the European Economic Association*, 1(2-3):512–521.
- Abel, A. B. (1986). Capital Accumulation and Uncertain Lifetimes with Adverse Selection. *Econometrica*, 54(5):1079–97.
- Acemoglu, D. and Shimer, R. (1999). Efficient Unemployment Insurance. *Journal of Political Economy*, 107(5):893–928.
- Adams, W., Einav, L., and Levin, J. (2009). Liquidity Constraints and Imperfect Information in Subprime Lending. *American Economic Review*, 99(1):49–84.
- Akerlof, G. A. (1970). The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 84(3):488–500.
- Anderson, P. M. and Meyer, B. D. (1993). Unemployment Insurance in the United States: Layoff Incentives and Cross Subsidies. *Journal of Labor Economics*, 11(1):70–95.
- Anderson, P. M. and Meyer, B. D. (1997). Unemployment Insurance Takeup Rates and the After-Tax Value of Benefits. *Quarterly Journal of Economics*, 112(3):913–37.
- Auerbach, A. J. (1985). The theory of excess burden and optimal taxation. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 1, pages 61–127. Elsevier, Amsterdam, North Holland.
- Autor, D. H. and Duggan, M. G. (2003). The Rise In The Disability Rolls And The Decline In Unemployment. *Quarterly Journal of Economics*, 118(1):157–205.
- Baicker, K., Goldin, C., and Katz, L. F. (1998). A Distinctive System: Origins and Impact of U.S. Unemployment Compensation. In Bordo, M. D., Goldin, C., and White, E. N., editors, *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*, pages 227–263. University of Chicago Press, Chicago.
- Baily, M. N. (1978). Some aspects of optimal unemployment insurance. *Journal of Public Economics*, 10(3):379–402.
- Baird, S., Hicks, J. H., Kremer, M., and Miguel, E. (2011). Worms at Work: Long-run Impacts of Child Health Gains. Mimeo. http://elsa.berkeley.edu/~emiguel/pdfs/miguel_wormsatwork.pdf.
- Bank of Sweden (2001). Press Release: The 2001 Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel, Advanced Information. http://www.nobelprize.org/nobel_prizes/economics/laureates/2001/advanced-economicsciences2001.pdf.
- Barseghyan, L., Molinari, F., O’Donoghue, T., and Teitelbaum, J. C. (2012). The Nature of Risk Preferences: Evidence from Insurance Choices. SSRN Working Paper 1646520.

- Barseghyan, L., Prince, J., and Teitelbaum, J. C. (2011). Are Risk Preferences Stable across Contexts? Evidence from Insurance Data. *American Economic Review*, 101(2):591–631.
- Barsky, R. B., Juster, F. T., Kimball, M. S., and Shapiro, M. D. (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study. *Quarterly Journal of Economics*, 112(2):537–579.
- Bernheim, B. D. and Rangel, A. (2009). Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics-super-. *Quarterly Journal of Economics*, 124(1):51–104.
- Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network Effects And Welfare Cultures. *Quarterly Journal of Economics*, 115(3):1019–1055.
- Bhargava, S. and Manoli, D. (2011). Why are Benefits Left on the Table? Assessing the Role of Information, Complexity, and Stigma on Take-up with an IRS Field Experiment. Working Paper.
- Black, D. A., Smith, J. A., Berger, M. C., and Noel, B. J. (2003). Is the Threat of Reemployment Services More Effective Than the Services Themselves? Evidence from Random Assignment in the UI System. *American Economic Review*, 93(4):1313–1327.
- Blanchard, O. J. and Tirole, J. (2008). The Joint Design of Unemployment Insurance and Employment Protection: A First Pass. *Journal of the European Economic Association*, 6(1):45–77.
- Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption Inequality and Partial Insurance. *American Economic Review*, 98(5):1887–1921.
- Borghans, L., Gielen, A. C., and Luttmer, E. F. P. (2010). Social Support Shopping: Evidence from a Regression Discontinuity in Disability Insurance Reform. IZA Discussion Paper 5412, Institute for the Study of Labor (IZA).
- Bound, J., Cullen, J. B., Nichols, A., and Schmidt, L. (2004). The welfare implications of increasing disability insurance benefit generosity. *Journal of Public Economics*, 88(12):2487–2514.
- Bronchetti, E. T. (2012). Workers’ compensation and consumption smoothing. *Journal of Public Economics*, 96(5):495–508.
- Brown, J., Duggan, M., Kuziemko, I., and Woolston, W. (2011). How does Risk Selection Respond to Risk Adjustment? Evidence from the Medicare Advantage Program. NBER Working Paper 16977, National Bureau of Economic Research.
- Brown, J. R. and Finkelstein, A. (2007). Why is the market for long-term care insurance so small? *Journal of Public Economics*, 91(10):1967–1991.
- Brown, J. R. and Finkelstein, A. (2008). The Interaction of Public and Private Insurance: Medicaid and the Long-Term Care Insurance Market. *American Economic Review*, 98(3):1083–1102.
- Brown, J. R. and Finkelstein, A. (2011). Insuring Long-Term Care in the United States. *Journal of Economic Perspectives*, 25(4):119–42.
- Brown, J. R., Kroszner, R. S., and Jenn, B. H. (2002). Federal Terrorism Risk Insurance. *National Tax Journal*, 55(3):647–657.

- Browning, M. and Crossley, T. F. (2001). Unemployment insurance benefit levels and consumption changes. *Journal of Public Economics*, 80(1):1–23.
- Buchanan, J. (1975). The Samaritan’s Dilemma. In Phelps, E., editor, *Altruism, Morality and Economic Theory*, pages 71–85. Russell Sage, New York.
- Buchmueller, T. and DiNardo, J. (2002). Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut. *American Economic Review*, 92(1):280–294.
- Bundorf, M. K., Levin, J., and Mahoney, N. (forthcoming). Pricing and Welfare in Health Plan Choice. *American Economic Review*.
- Bundorf, M. K. and Simon, K. I. (2006). The Effects of Rate Regulation on Demand for Supplemental Health Insurance. *American Economic Review*, 96(2):67–71.
- Card, D., Chetty, R., and Weber, A. (2007a). The Spike at Benefit Exhaustion: Leaving the Unemployment System or Starting a New Job? *American Economic Review*, 97(2):113–118.
- Card, D., Chetty, R., and Weber, A. (2007b). Cash-On-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market. *Quarterly Journal of Economics*, 122(4):1511–1560.
- Card, D., Dobkin, C., and Maestas, N. (2008). The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare. *American Economic Review*, 98(5):2242–58.
- Card, D., Dobkin, C., and Maestas, N. (2009). Does Medicare Save Lives? *Quarterly Journal of Economics*, 124(2):597–636.
- Card, D. and Levine, P. B. (2000). Extended benefits and the duration of UI spells: evidence from the New Jersey extended benefit program. *Journal of Public Economics*, 78(1-2):107–138.
- Cardon, J. H. and Hendel, I. (2001). Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey. *RAND Journal of Economics*, 32(3):408–27.
- Carlin, C. and Town, R. (2010). Adverse Selection: The Dog That Didn’t Bite. Mimeo, University of Minnesota.
- Carroll, C. D. (1997). Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis. *Quarterly Journal of Economics*, 112(1):1–55.
- Centeno, M. and Novo, A. (2009). Reemployment wages and UI liquidity effect: a regression discontinuity approach. *Portuguese Economic Journal*, 8(1):45–52.
- Chetty, R. (2006a). A New Method of Estimating Risk Aversion. *American Economic Review*, 96(5):1821–1834.
- Chetty, R. (2006b). A general formula for the optimal level of social insurance. *Journal of Public Economics*, 90(10-11):1879–1901.
- Chetty, R. (2008). Moral Hazard versus Liquidity and Optimal Unemployment Insurance. *Journal of Political Economy*, 116(2):173–234.

- Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics*, 1(1):451–488.
- Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics*, 126(2):749–804.
- Chetty, R., Friedman, J. N., and Saez, E. (2012). Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings. NBER Working Paper 18232, National Bureau of Economic Research.
- Chetty, R. and Looney, A. (2006). Consumption smoothing and the welfare consequences of social insurance in developing economies. *Journal of Public Economics*, 90(12):2351–2356.
- Chetty, R., Looney, A., and Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *American Economic Review*, 99(4):1145–77.
- Chetty, R. and Saez, E. (2010). Optimal Taxation and Social Insurance with Endogenous Private Insurance. *American Economic Journal: Economic Policy*, 2(2):85–114.
- Chetty, R. and Szeidl, A. (2007). Consumption Commitments and Risk Preferences. *Quarterly Journal of Economics*, 122(2):831–877.
- Chiappori, P.-A., Jullien, B., Salanié, B., and Salanié, F. (2006). Asymmetric information in insurance: general testable implications. *RAND Journal of Economics*, 37(4):783–798.
- Chiappori, P.-A. and Salanie, B. (2000). Testing for Asymmetric Information in Insurance Markets. *Journal of Political Economy*, 108(1):56–78.
- Clark, A. E. (2003). Unemployment as a Social Norm: Psychological Evidence from Panel Data. *Journal of Labor Economics*, 21(2):289–322.
- Coate, S. (1995). Altruism, the Samaritan’s Dilemma, and Government Transfer Policy. *American Economic Review*, 85(1):46–57.
- Cochrane, J. H. (1991). A Simple Test of Consumption Insurance. *Journal of Political Economy*, 99(5):957–76.
- Cohen, A. and Einav, L. (2007). Estimating Risk Preferences from Deductible Choice. *American Economic Review*, 97(3):745–788.
- Cohen, A. and Siegelman, P. (2010). Testing for Adverse Selection in Insurance Markets. *Journal of Risk and Insurance*, 77(1):39–84.
- Crocker, K. J. and Snow, A. (1985). The efficiency of competitive equilibria in insurance markets with asymmetric information. *Journal of Public Economics*, 26(2):207–219.
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2012). Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment. CEPR Discussion Papers 9251, C.E.P.R. Discussion Papers.
- Cullen, J. B. and Gruber, J. (2000). Does Unemployment Insurance Crowd Out Spousal Labor Supply? *Journal of Labor Economics*, 18(3):546–72.

- Currie, J. (2006). The Take-up of Social Benefits. In Auerbach, A., Card, D., and Quigley, J., editors, *Poverty, the Distribution of Income, and Public Policy*, pages 80–148. Russell Sage, New York.
- Currie, J. and Gruber, J. (1996a). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *Quarterly Journal of Economics*, 111(2):431–66.
- Currie, J. and Gruber, J. (1996b). Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy*, 104(6):1263–96.
- Cutler, D. M. (2002). Health care and the public sector. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, edition 1, volume 4, pages 2143–2243. Elsevier, Amsterdam, North Holland.
- Cutler, D. M., Finkelstein, A., and McGarry, K. (2008). Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance. *American Economic Review*, 98(2):157–62.
- Cutler, D. M. and Gruber, J. (1996). Does Public Insurance Crowd Out Private Insurance? *Quarterly Journal of Economics*, 111(2):391–430.
- Cutler, D. M. and Reber, S. J. (1998). Paying For Health Insurance: The Trade-Off Between Competition And Adverse Selection. *Quarterly Journal of Economics*, 113(2):433–466.
- Cutler, D. M. and Zeckhauser, R. J. (2000). The anatomy of health insurance. In Culyer, A. J. and Newhouse, J. P., editors, *Handbook of Health Economics*, edition 1, volume 1 of *Handbook of Health Economics*, pages 563–643. Elsevier, Amsterdam, North Holland.
- Davidoff, T., Brown, J. R., and Diamond, P. A. (2005). Annuities and Individual Welfare. *American Economic Review*, 95(5):1573–1590.
- de Meza, D. and Webb, D. C. (2001). Advantageous Selection in Insurance Markets. *RAND Journal of Economics*, 32(2):249–62.
- Deaton, A. (1991). Saving and Liquidity Constraints. *Econometrica*, 59(5):1221–48.
- DellaVigna, S. and Paserman, M. D. (2005). Job Search and Impatience. *Journal of Labor Economics*, 23(3):527–588.
- Devine, T. J. and Kiefer, N. M. (1991). *Empirical Labor Economics: The Search Approach*. Oxford University Press, New York.
- Diamond, P. A. (1977). A framework for social security analysis. *Journal of Public Economics*, 8(3):275–298.
- Ebenstein, A. and Stange, K. (2010). Does inconvenience explain low take-up? Evidence from unemployment insurance. *Journal of Policy Analysis and Management*, 29(1):111–136.
- Eckstein, Z., Eichenbaum, M., and Peled, D. (1985). Uncertain lifetimes and the welfare enhancing properties of annuity markets and social security. *Journal of Public Economics*, 26(3):303–326.
- Einav, L. and Finkelstein, A. (2011). Selection in Insurance Markets: Theory and Empirics in Pictures. *Journal of Economic Perspectives*, 25(1):115–38.

- Einav, L., Finkelstein, A., and Cullen, M. R. (2010a). Estimating Welfare in Insurance Markets Using Variation in Prices. *Quarterly Journal of Economics*, 125(3):877–921.
- Einav, L., Finkelstein, A., and Levin, J. (2010b). Beyond Testing: Empirical Models of Insurance Markets. *Annual Review of Economics*, 2(1):311–336.
- Einav, L., Finkelstein, A., Pascu, I., and Cullen, M. (forthcoming). How General Are Risk Preferences? Choices Under Uncertainty in Different Domains. *American Economic Review*.
- Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2011). Selection on Moral Hazard in Health Insurance. NBER Working Paper 16969, National Bureau of Economic Research.
- Einav, L., Finkelstein, A., and Schrimpf, P. (2010c). Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence From the U.K. Annuity Market. *Econometrica*, 78(3):1031–1092.
- Engen, E. M. and Gruber, J. (2001). Unemployment insurance and precautionary saving. *Journal of Monetary Economics*, 47(3):545–579.
- Fang, H., Keane, M. P., and Silverman, D. (2008). Sources of Advantageous Selection: Evidence from the Medigap Insurance Market. *Journal of Political Economy*, 116(2):303–350.
- Fang, H. and Silverman, D. (2009). Time-Inconsistency And Welfare Program Participation: Evidence From The NLSY. *International Economic Review*, 50(4):1043–1077.
- Feldstein, M. (1978). The Effect of Unemployment Insurance on Temporary Layoff Unemployment. *American Economic Review*, 68(5):834–46.
- Feldstein, M. and Altman, D. (2007). Unemployment Insurance Savings Accounts. In *Tax Policy and the Economy*, volume 21 of *NBER Chapters*, pages 35–64. National Bureau of Economic Research.
- Feldstein, M. and Liebman, J. B. (2002). Social security. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, edition 1, volume 4 of *Handbook of Public Economics*, pages 2245–2324. Elsevier, Amsterdam, North Holland.
- Feldstein, M. and Poterba, J. (1984). Unemployment insurance and reservation wages. *Journal of Public Economics*, 23(1-2):141–167.
- Finkelstein, A. (2004). The interaction of partial public insurance programs and residual private insurance markets: evidence from the US Medicare program. *Journal of Health Economics*, 23(1):1–24.
- Finkelstein, A. (2007). The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare. *Quarterly Journal of Economics*, 122(1):1–37.
- Finkelstein, A., Luttmer, E. F. P., and Notowidigdo, M. J. (2008). What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption. NBER Working Paper 14089, National Bureau of Economic Research.
- Finkelstein, A., Luttmer, E. F. P., and Notowidigdo, M. J. (2009a). Approaches to Estimating the Health State Dependence of the Utility Function. *American Economic Review*, 99(2):116–21.

- Finkelstein, A. and McGarry, K. (2006). Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market. *American Economic Review*, 96(4):938–958.
- Finkelstein, A. and McKnight, R. (2008). What did Medicare do? The initial impact of Medicare on mortality and out of pocket medical spending. *Journal of Public Economics*, 92(7):1644–1668.
- Finkelstein, A. and Poterba, J. (2002). Selection Effects in the United Kingdom Individual Annuities Market. *Economic Journal*, 112(476):28–50.
- Finkelstein, A. and Poterba, J. (2004). Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market. *Journal of Political Economy*, 112(1):183–208.
- Finkelstein, A. and Poterba, J. (2006). Testing for Adverse Selection with ‘Unused Observables’. NBER Working Paper 12112, National Bureau of Economic Research.
- Finkelstein, A., Poterba, J., and Rothschild, C. (2009b). Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities. *Journal of Financial Economics*, 91(1):38–58.
- Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Group, T. O. H. S. (2011). The Oregon Health Insurance Experiment: Evidence from the First Year. NBER Working Paper 17190, National Bureau of Economic Research, Inc.
- Fishback, P. V. and Kantor, S. E. (1998). The Adoption of Workers’ Compensation in the United States, 1900-1930. *Journal of Law and Economics*, 41(2):305–41.
- Friedman, B. M. and Warshawsky, M. J. (1990). The Cost of Annuities: Implications for Saving Behavior and Bequests. *Quarterly Journal of Economics*, 105(1):135–54.
- Froot, K. A. (2001). The market for catastrophe risk: a clinical examination. *Journal of Financial Economics*, 60(2-3):529–571.
- Gertler, P. and Gruber, J. (2002). Insuring Consumption Against Illness. *American Economic Review*, 92(1):51–70.
- Geruso, M. (2011). Selection in Employer Health Plans: Homogeneous Prices and Heterogeneous Preferences. Mimeo, Princeton University.
- Glied, S. (2000). Managed care. In Culyer, A. J. and Newhouse, J. P., editors, *Handbook of Health Economics*, volume 1, pages 707–753. Elsevier, Amsterdam, North Holland.
- Golosov, M. and Tsyvinski, A. (2007). Optimal Taxation With Endogenous Insurance Markets. *Quarterly Journal of Economics*, 122(2):487–534.
- Gross, T. and Notowidigdo, M. J. (2011). Health insurance and the consumer bankruptcy decision: Evidence from expansions of Medicaid. *Journal of Public Economics*, 95(7-8):767–778.
- Gruber, J. (1997). The Consumption Smoothing Benefits of Unemployment Insurance. *American Economic Review*, 87(1):192–205.
- Gruber, J. (2000). Disability Insurance Benefits and Labor Supply. *Journal of Political Economy*, 108(6):1162–1183.
- Gruber, J. (2009). *Public Finance and Public Policy*. Worth Publishers, New York, edition 3.

- Gruber, J. and Wise, D. A., editors (1999). *Social Security and Retirement around the World*. NBER Book Series - International Social Security. University of Chicago Press, Chicago.
- Hamermesh, D. S. (1982). Social Insurance and Consumption: An Empirical Inquiry. *American Economic Review*, 72(1):101–13.
- Handel, B. R. (2011). Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts. NBER Working Paper 17459, National Bureau of Economic Research.
- Hansen, G. D. and Imrohroglu, A. (1992). The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard. *Journal of Political Economy*, 100(1):118–42.
- Hartley, G. R., van Ours, J. C., and Vodopivec, M. (2011). Incentive effects of unemployment insurance savings accounts: Evidence from Chile. *Labour Economics*, 18(6):798–809.
- Hemenway, D. (1990). Propitious Selection. *Quarterly Journal of Economics*, 105(4):1063–69.
- Hendel, I. and Lizzeri, A. (2003). The Role Of Commitment In Dynamic Contracts: Evidence From Life Insurance. *Quarterly Journal of Economics*, 118(1):299–327.
- Hendren, N. (2012). Private Information and Insurance Rejections. NBER Working Paper 18282, National Bureau of Economic Research.
- Herring, B. (2005). The effect of the availability of charity care to the uninsured on the demand for private health insurance. *Journal of Health Economics*, 24(2):225–252.
- Hopenhayn, H. A. and Nicolini, J. P. (1997). Optimal Unemployment Insurance. *Journal of Political Economy*, 105(2):412–38.
- Hosseini, R. (2010). Adverse Selection in the Annuity Market and the Role for Social Security. Mimeo. http://www.public.asu.edu/~rhossein/Roozbeh_Hosseini/Research_files/Hosseini_Annuity_SS_Final.pdf.
- Hubbard, R. G., Skinner, J., and Zeldes, S. P. (1995). Precautionary Saving and Social Insurance. *Journal of Political Economy*, 103(2):360–99.
- Huggett, M. and Parra, J. C. (2010). How Well Does the U.S. Social Insurance System Provide Social Insurance? *Journal of Political Economy*, 118(1):76–112.
- Jappelli, T. and Pistaferri, L. (2010). The Consumption Response to Income Changes. *Annual Review of Economics*, 2(1):479–506.
- Kaplow, L. (1991). Incentives and Government Relief for Risk. *Journal of Risk and Uncertainty*, 4(2):167–75.
- Karlan, D. and Zinman, J. (2009). Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment. *Econometrica*, 77(6):1993–2008.
- Katz, L. F. and Meyer, B. D. (1990). Unemployment Insurance, Recall Expectations, and Unemployment Outcomes. *Quarterly Journal of Economics*, 105(4):973–1002.
- Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C., and Wrobel, M. V. (2012). Comparison Friction: Experimental Evidence from Medicare Drug Plans. *Quarterly Journal of Economics*, 127(1):199–235.

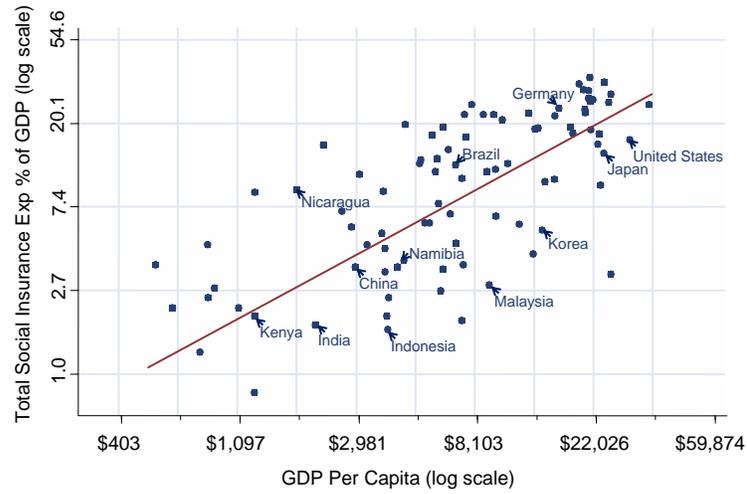
- Kocherlakota, N. R. (1996). The Equity Premium: It's Still a Puzzle. *Journal of Economic Literature*, 34(1):42–71.
- Kremer, M. and Miguel, E. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1):159–217.
- Kroft, K. (2008). Takeup, social multipliers and optimal social insurance. *Journal of Public Economics*, 92(3-4):722–737.
- Kroft, K. and Notowidigdo, M. J. (2011). Should Unemployment Insurance Vary With the Unemployment Rate? Theory and Evidence. NBER Working Paper 17173, National Bureau of Economic Research.
- Krueger, A. B. and Meyer, B. D. (2002). Labor supply effects of social insurance. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, edition 1, volume 4, pages 2327–2392. Elsevier, Amsterdam, North Holland.
- Krueger, A. B. and Mueller, A. I. (2011). Job Search and Job Finding in a Period of Mass Unemployment: Evidence from High-Frequency Longitudinal Data. IZA Discussion Paper 5450, Institute for the Study of Labor (IZA).
- Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics*, 112(2):443–77.
- Lalive, R. (2003). Social Interactions in Unemployment. IZA Discussion Paper 803, Institute for the Study of Labor (IZA).
- Lalive, R. (2007). Unemployment Benefits, Unemployment Duration, and Post-Unemployment Jobs: A Regression Discontinuity Approach. *American Economic Review*, 97(2):108–112.
- Lalive, R. (2008). How do extended benefits affect unemployment duration A regression discontinuity approach. *Journal of Econometrics*, 142(2):785–806.
- Lalive, R., van Ours, J. C., and Zweimüller, J. (2006). How Changes in Financial Incentives Affect the Duration of Unemployment. *Review of Economic Studies*, 73(4):1009–1038.
- LaLonde, R. (2007). The Case for Wage Insurance. Council Special Report 30, Council on Foreign Relations Press.
- LaLumia, S. (2011). The EITC, Tax Refunds, and Unemployment Spells. Department of Economics Working Paper 2011-09, Department of Economics, Williams College.
- Landais, C. (2011). Heterogeneity and Behavioral Responses to Unemployment Benefits over the Business Cycle. SIEPR working paper, Stanford Institute for Economic Policy Research.
- Landais, C. (2012). Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design. PEP working paper, London School of Economics.
- Landais, C., Michailat, P., and Saez, E. (2012). Optimal Unemployment Insurance over the Business Cycle. NBER Working Paper 16526, National Bureau of Economic Research.
- Lentz, R. (2009). Optimal Unemployment Insurance in an Estimated Job Search Model with Savings. *Review of Economic Dynamics*, 12(1):37–57.

- Lentz, R. and Tranas, T. (2005). Job Search and Savings: Wealth Effects and Duration Dependence. *Journal of Labor Economics*, 23(3):467–490.
- Lillard, L. A. and Weiss, Y. (1997). Uncertain Health and Survival: Effects on End-of-Life Consumption. *Journal of Business & Economic Statistics*, 15(2):254–68.
- Lindbeck, A., Nyberg, S., and Weibull, J. W. (1999). Social Norms And Economic Incentives In The Welfare State. *Quarterly Journal of Economics*, 114(1):1–35.
- Lucas, R. (1987). *Models of Business Cycles*. Basil Blackwell, New York.
- Lustig, J. (2011). The Welfare Effects of Adverse Selection in Privatized Medicare. Mimeo, Boston University. http://people.bu.edu/jlustig/lustig_march2011.pdf.
- Maestas, N., Mullen, K. J., and Strand, A. (forthcoming). Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *American Economic Review*.
- Mahoney, N. (2012). Bankruptcy as Implicit Health Insurance. NBER Working Paper 18105, National Bureau of Economic Research.
- Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., and Leibowitz, A. (1987). Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment. *The American Economic Review*, 77(3):251–277.
- McCarthy, D. and Mitchell, O. S. (2003). International Adverse Selection in Life Insurance and Annuities. NBER Working Paper 9975, National Bureau of Economic Research.
- Mehra, R. and Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161.
- Metrick, A. (1995). A Natural Experiment in ‘Jeopardy!’. *American Economic Review*, 85(1):240–53.
- Meyer, B. D. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, 58(4):757–82.
- Meyer, B. D., Viscusi, W. K., and Durbin, D. L. (1995). Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment. *American Economic Review*, 85(3):322–40.
- Mitchell, O. S., Poterba, J. M., Warshawsky, M. J., and Brown, J. R. (1999). New Evidence on the Money’s Worth of Individual Annuities. *American Economic Review*, 89(5):1299–1318.
- Newhouse, J. P. (2002). *Princing the Priceless: A Health Care Conundrum*. The MIT Press, Cambridge, Mass.
- Newhouse, J. P. and {RAND Corporation Insurance Experiment Group} (1996). *Free for All?: Lessons from the RAND Health Insurance Experiment*. Harvard University Press, Cambridge, Mass.
- Nichols, A. L. and Zeckhauser, R. J. (1982). Targeting Transfers through Restrictions on Recipients. *American Economic Review*, 72(2):372–77.

- Nyman, J. (2003). *The Theory of Demand for Health Insurance*. Stanford University Press, Stanford, Calif.
- Orszag, J. M. and Snower, D. J. (2002). Incapacity benefits and employment policy. *Labour Economics*, 9(5):631–641.
- Oster, E., Shoulson, I., Quaid, K., and Dorsey, E. R. (2010). Genetic adverse selection: Evidence from long-term care insurance and Huntington disease. *Journal of Public Economics*, 94(11-12):1041–1050.
- Pauly, M. V. (1974). Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection. *Quarterly Journal of Economics*, 88(1):44–62.
- Philipson, T. J. and Becker, G. S. (1998). Old-Age Longevity and Mortality-Contingent Claims. *Journal of Political Economy*, 106(3):551–573.
- Rabin, M. (2000). Risk Aversion and Expected-Utility Theory: A Calibration Theorem. *Econometrica*, 68(5):1281–1292.
- Rothschild, M. and Stiglitz, J. E. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics*, 90(4):630–49.
- Rothstein, J. (2010). Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence. *American Economic Journal: Economic Policy*, 2(1):177–208.
- Schmieder, J. F., Wachter, T. v., and Bender, S. (2012). The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years. *Quarterly Journal of Economics*, 127(2):701–752.
- Shavell, S. and Weiss, L. (1979). The Optimal Payment of Unemployment Insurance Benefits over Time. *Journal of Political Economy*, 87(6):1347–62.
- Sheshinski, E. (2007). *The Economic Theory of Annuities*, volume 1 of *Economics Books*. Princeton University Press.
- Shimer, R. and Werning, I. (2006). On the Optimal Timing of Benefits with Heterogeneous Workers and Human Capital Depreciation. NBER Working Paper 12230, National Bureau of Economic Research.
- Shimer, R. and Werning, I. (2007). Reservation Wages and Unemployment Insurance. *Quarterly Journal of Economics*, 122(3):1145–1185.
- Shimer, R. and Werning, I. (2008). Liquidity and Insurance for the Unemployed. *American Economic Review*, 98(5):1922–42.
- Simon, K. I. (2005). Adverse selection in health insurance markets? Evidence from state small-group health insurance reforms. *Journal of Public Economics*, 89(9-10):1865–1877.
- Social Security Administration (1997). Social Security Programs in the United States. SSA Publication 13-11758.
- Spinnewijn, J. (2010). Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs. Working Paper, London School of Economics. <http://personal.lse.ac.uk/spinnewi/biasedbeliefs.pdf>.

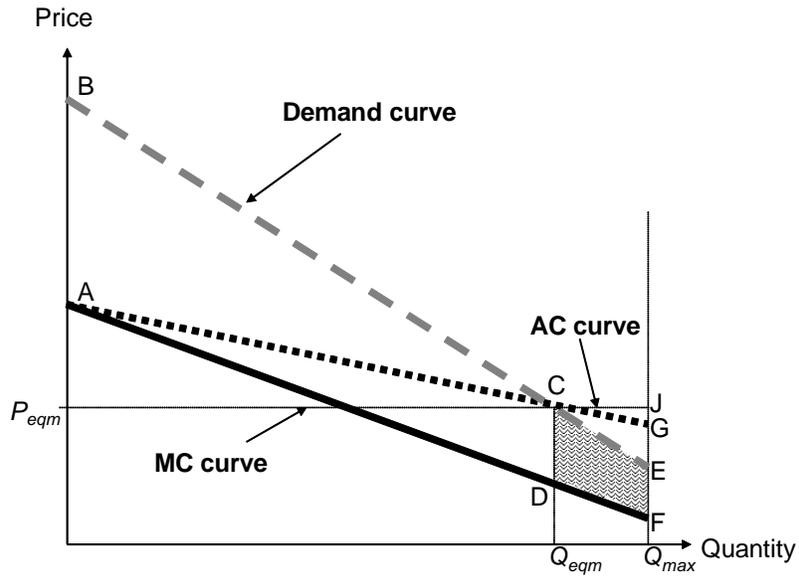
- Spinnewijn, J. (2012). Heterogeneity, Demand for Insurance and Adverse Selection. Working Paper, London School of Economics. http://personal.lse.ac.uk/spinnewi/perceptions_welfare.pdf.
- Starc, A. (2011). Insurer Pricing and Consumer Welfare: Evidence from Medigap. Mimeo. http://hcmg.wharton.upenn.edu/documents/research/Astarc_pricing.pdf.
- Stiglitz, J. E. and Yun, J. (2005). Integration of unemployment insurance with retirement insurance. *Journal of Public Economics*, 89(11-12):2037–2067.
- Summers, L. H. (1989). Some Simple Economics of Mandated Benefits. *American Economic Review*, 79(2):177–83.
- Sydnor, J. (2010). (Over)insuring Modest Risks. *American Economic Journal: Applied Economics*, 2(4):177–99.
- Topel, R. H. (1983). On Layoffs and Unemployment Insurance. *American Economic Review*, 73(4):541–59.
- Townsend, R. M. (1994). Risk and Insurance in Village India. *Econometrica*, 62(3):539–91.
- van Ours, J. C. and Vodopivec, M. (2008). Does reducing unemployment insurance generosity reduce job match quality? *Journal of Public Economics*, 92(3-4):684–695.
- von Wachter, T., Song, J., and Manchester, J. (2009). Long-Term Earnings Losses due to Mass-Layoffs During the 1982 Recession: An Analysis Using Longitudinal Administrative Data from 1974 to 2004. Working Paper, Columbia University.
- Wang, C. and Williamson, S. (1996). Unemployment insurance with moral hazard in a dynamic economy. *Carnegie-Rochester Conference Series on Public Policy*, 44(1):1–41.
- Werning, I. (2002). Optimal Unemployment Insurance with Unobservable Savings. Mimeo, University of Chicago.
- Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory*, 16(2):167–207.
- Wolpin, K. I. (1987). Estimating a Structural Search Model: The Transition from School to Work. *Econometrica*, 55(4):801–17.
- Yaari, M. E. (1965). Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. *The Review of Economic Studies*, 32(2):137–150.

Figure 1: Social Insurance vs. GDP Per Capita in 1996



Notes: Social Insurance statistics are from International Labour Organization (2000). Social insurance is defined as total expenditures on social security, disability insurance, unemployment insurance, insurance against work-related injuries, and government provided health insurance. GDP statistics are from the Penn World tables. GDP is measured in PPP-adjusted 1996 US dollars.

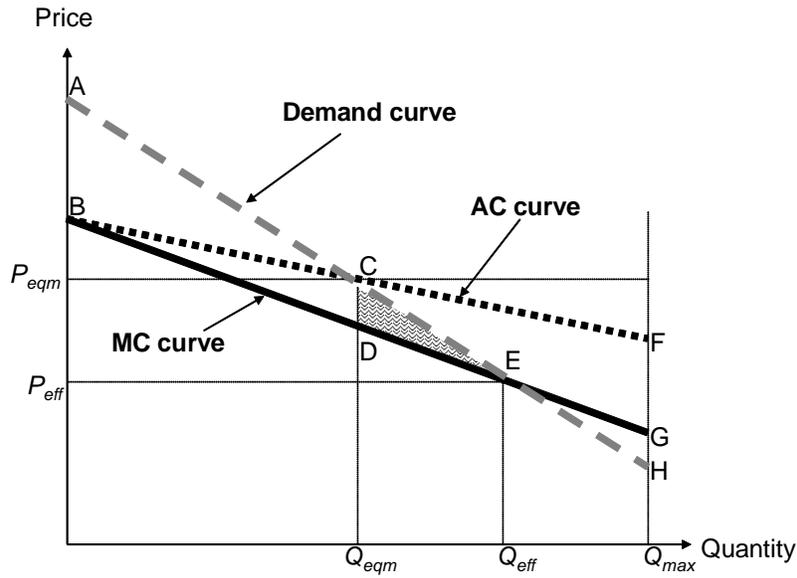
Figure 2: Adverse Selection in the Textbook Setting



Source: Einav and Finkelstein (2011).

Notes: Figure 2 shows the demand (willingness-to-pay) for a high coverage H relative to a lower coverage contract L , and the associated marginal and average incremental cost (i.e. expected insurance claims) curves. The downward sloping marginal cost curve indicates adverse selection. The efficient allocation is for everyone to be covered by H (since willingness to pay is always above marginal cost) but the equilibrium allocation covers only those whose willingness to pay is above average costs, creating the classic under-insurance result of adverse selection. The welfare loss from this under-insurance is given by the trapezoid CDEF, representing the excess of demand above marginal cost for those who are not covered by H in equilibrium.

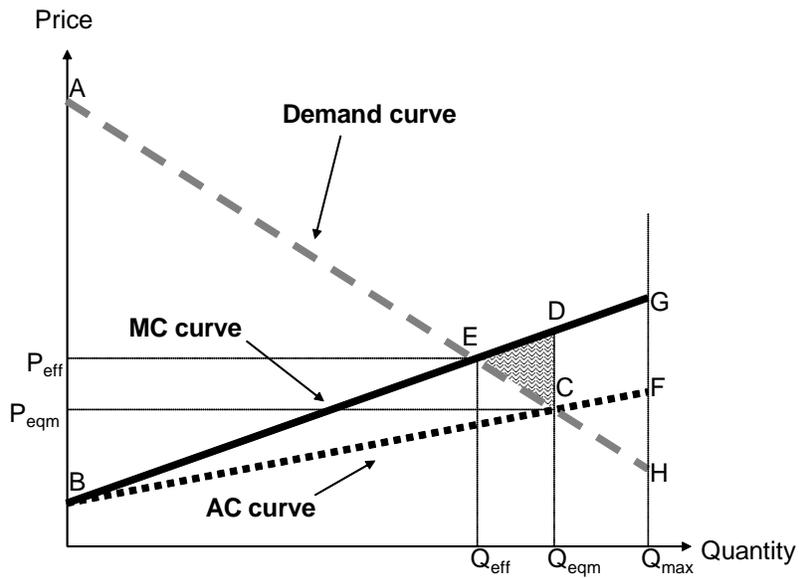
Figure 3: Adverse Selection with Additional Costs of Providing Insurance



Source: Einav, Finkelstein and Cullen (2010)

Notes: In this departure from the textbook case, we allow for the possibility of a loading factor on the insurance contract H . As a result, the marginal cost curve may now intersect the demand curve internally, in which case it is not efficient to cover all individuals with H . The efficient allocation is given by point E (where demand intersects the marginal cost curve) and the equilibrium allocation is given by point C (where demand intersects the average cost curve). Once again there is under-insurance due to adverse selection ($Q_{eqm} < Q_{eff}$) and the welfare loss from this under-insurance is given by the triangle CDE.

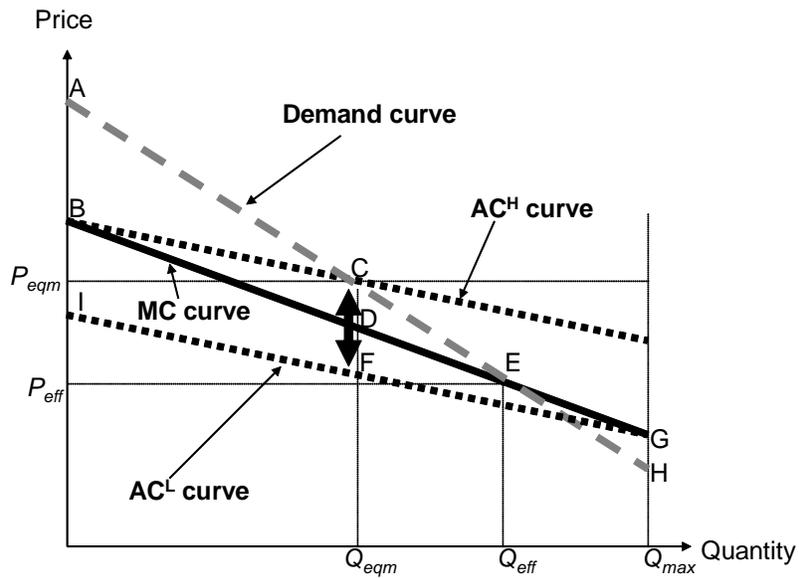
Figure 4: Advantageous Selection



Source: Einav, Finkelstein and Cullen (2010).

Notes: Advantageous selection is characterized by an upward sloping marginal cost curve. The average cost curve therefore lies below the marginal cost curve, resulting in over-insurance relative to the efficient allocation ($Q_{eff} < Q_{eqm}$). The welfare losses from over-insurance is given by the shaded area CDE and represents the excess of marginal cost over willingness to pay for people whose willingness to pay exceed the average costs of those covered by H .

Figure 5: The “Positive Correlation” Test for Selection



Source: Einav and Finkelstein (2011).

Notes: AC^H denotes the AC curve shown in previous figures (i.e. average costs of those with the higher coverage contract H). The AC^L curve represents the average costs of those with the lower coverage contract L . The difference in the average costs of those with more and less insurance in equilibrium is given by the line segment CF . The positive correlation test for asymmetric information tests whether average costs of those who in equilibrium have more insurance (point C) are above average costs of those who in equilibrium have less insurance (point F).

TABLE 1
Social Insurance: A Summary of Existing Applied Work

	(1) Testing for Selection	(2) Quantifying Welfare Costs of Selection	(3) Estimation of Benefits	(4) Estimation of Moral Hazard Costs	(5) Calculating Optimal Benefit Levels
Medicare / Medicaid	X	X	X	X	
Disability Insurance				X	
Unemp. Insurance			X	X	X
Workers' Comp.			X	X	X

NOTES: This table lists areas where we are aware of empirical work on the topic listed in each column for five major social insurance programs. Selected examples of this work would include the following. Column 1: Cutler and Reber (1998) on Medicare/Medicaid (more accurately: private health insurance markets). Column 2: Einav, Finkelstein and Cullen (2010) on health insurance markets. Column 3: Card et al. (2009) on Medicare and Finkelstein et al. (2012) on Medicaid; Gruber (1997) and Chetty (2008) on unemployment insurance, and Bronchetti (2012) on workers compensation. Column 4: Card et al. (2009) on Medicare and Finkelstein et al. (2012) on Medicaid; Gruber (2001) and Maestas, Mullen, and Strand (2012) on disability insurance, Meyer (1990) and Card, Chetty, and Weber (2007) on unemployment insurance, and Meyer, Viscusi, and Durbin (1995) on workers' compensation. Column 5: Gruber (1997) and Chetty (2008) on unemployment insurance and Bronchetti (2012) on workers compensation.