American Economic Review 2015, 105(10): 2986–3029
http://dx.doi.org/10.1257/aer.20110108

Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off†

By Hamish Low and Luigi Pistaferri*

We provide a life-cycle framework for comparing insurance and disincentive effects of disability benefits. The risks that individuals face and the parameters of the Disability Insurance (DI) program are estimated from consumption, health, disability insurance, and wage data. We characterize the effects of disability insurance and study how policy reforms impact behavior and welfare. DI features high rejection rates of disabled applicants and some acceptance of healthy applicants. Despite worse incentives, welfare increases as programs become less strict or generosity increases. Disability insurance interacts with welfare programs: making unconditional means-tested programs more generous improves disability insurance targeting and increases welfare. (*JEL* D14, J24, J65)

The disability insurance (DI) program in the United States is a large and rapidly growing social insurance program offering income replacement and health care benefits to people with work limiting disabilities. In 2012, the cash benefits paid by the DI program were more than three times larger than those paid by unemployment insurance (UI) ($136.9 billion versus $42.7 billion).[1] Between 1985 and 2012 the proportion of DI claimants in the United States has more than doubled (from about 2.4 percent to 5.9 percent of the working-age population), while the share of total old-age, survivors, and disability insurance (OASDI) spending accounted for by the DI program has grown from 10 percent to 17 percent. The key questions in thinking about the size and growth of the program are whether program claimants are genuinely unable to work, whether those in need of support are receiving insurance, and how valuable is the insurance provided vis-à-vis the inefficiencies created by the program.

* Low: Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD, UK (e-mail: Hamish.Low@econ.cam.ac.uk); Pistaferri: Department of Economics, Stanford University, Stanford, CA 94305 (e-mail: pista@stanford.edu). This paper previously circulated under the title "Disability Risk and the Value of Disability Insurance." Low thanks funding from the ESRC as a Research Fellow, grant number RES-063-27-0211. Pistaferri thanks funding from NIH/NIA under grant 1R01AG032029-01 and from NSF under grant SES-0921689. We have received useful comments from audiences at various conferences and departments in Europe and the United States. We are especially grateful to three anonymous referees, Tom Crossley, Pramila Krishnan, Costas Meghir, and Aleh Tsyvinski for detailed comments, and to Katja Kaufmann, Itay Saporta Eksten, and Tom Zawisza for research assistance. Supplementary material, including supporting evidence, data, and simulation programs, is contained in an online Appendix. All errors are our own. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

† Go to http://dx.doi.org/10.1257/aer.20110108 to visit the article page for additional materials and author disclosure statement(s).

[1] The relative size of DI is even larger if we add the in-kind health care benefits provided by the Medicare program to DI beneficiaries.

In this paper we evaluate the welfare consequences of reforming key aspects of the DI program that are designed to alter the dynamics of the trade-off between the incentive costs and insurance aspects of the program. This evaluation requires a realistic model of individual behavior; a set of credible estimates of preferences, risks, and of the details of the program; and a way to measure the welfare consequences of the reforms.

To address these aims, we first propose a life-cycle framework that allows us to study savings, labor supply, and the decision to apply for DI under nonseparable preferences. We consider the problem of an individual who faces several sources of risk: a disability or work limitation shock which reduces the ability to work (distinguishing between severe and moderate shocks), a permanent productivity shock unrelated to health (such as a decline in the price of skills), and labor market frictions. Individuals differ ex ante due to unobserved productivity that may potentially be correlated with the probability of developing a work limitation. We assume that the DI program screens applicants with errors and reassesses them probabilistically following award. Second, we obtain estimates of the parameters of the model using microeconomic data from the Panel Study of Income Dynamics (PSID). We show that the model replicates well salient features of reality both internally (targeted moments) as well as externally (reduced form elasticities measuring the costs of the program, screening errors, exit flows, and life-cycle patterns of consumption and wealth). Finally, we analyze the impact on welfare and behavior of varying key policy parameters: (i) the generosity of disability payments; (ii) the stringency of the screening process; (iii) the generosity of alternative social insurance programs; and (iv) the reassessment rate. The ability to evaluate these questions in a coherent, unified framework is one of the main benefits of the paper. Our metric for household welfare is the consumption equivalent that keeps expected utility at the start of life constant as policy changes. We show that the welfare effects are determined by the dynamics of insurance for severely work limited individuals ("coverage") and of application rates by individuals who are not severely work limited ("false applications") as the policy changes.

We document a number of important findings. First, the disability insurance program is characterized by substantial false rejections, but by fewer false acceptances. Our distinction between those with no work limitation versus a moderate limitation highlights that false acceptances exist among the moderately disabled, but are negligible for those without any limitation. Second, in terms of policy reforms, the high fraction of false rejections associated with the screening process of the disability insurance program leads to an increase in welfare when the program becomes less strict, despite the increase in false applications. This is because coverage among those most in need (and especially those less equipped against disability risk due to lack of self-insurance through savings) goes up. Similarly, welfare is higher if the generosity of DI is increased and if reassessment is less frequent. Both of these reforms have a large impact increasing the number of applications from those with only a moderate disability, but this is outweighed by the benefit of improved insurance for those most in need. It is the difference in responsiveness to incentives among the moderately disabled compared to the severely disabled which underlies our policy conclusions. This distinction is novel to our paper and explains the difference between our findings and those elsewhere in the literature where responsiveness is not disaggregated by the severity of disability. Finally, DI interacts in important ways with welfare programs. We show that an increase in generosity of welfare programs (such as food stamps) reduces DI application rates by nondisabled workers and increases insurance coverage among disabled workers. This positive combination is due to the fact that marginal undeserving applicants use the means-tested program as a *substitute* for DI (they switch to a program that is increasingly as generous as DI but has less uncertainty), while truly disabled workers treat the means-tested program as a *complement* (they use the more generous income floor to finance the waiting time of application and also consumption in case of rejection).

The literature on the DI program, surveyed in the next section, contains both reduced form papers attempting to separately estimate the extent of inefficiencies created by the program and its insurance value, as well as sophisticated structural analyses geared toward assessing the consequences of reforming the program. As with most structural models, the value of our approach relative to reduced form analyses is that we can evaluate the consequences of potential reforms to the DI program, i.e., we can examine counterfactual cases that have not been experienced in the past or that are too costly to assess in a randomized evaluation context. Relative to existing structural analyses, we stress the importance of a number of model features: the different degrees of work limitation, early life-cycle choices, nonseparable preferences, fixed costs of work that depend on work limitation status, permanent skill shocks, and interactions with social welfare programs. Further, we study the effects of novel policy reforms, and subject our model to various validity tests. For our structural model to deliver credible policy conclusions, we require that it fits the data in a number of key dimensions (internal validity) and that it can replicate the estimates prevailing in the reduced form literature without targeting these estimates directly (external validation). We show to what extent our model passes these tests.

The rest of the paper is structured as follows. Section I reviews the relevant literature on the DI program. Section II presents the life-cycle model and discusses how we model preferences, the sources of risk faced by individuals, and the social insurance programs available to them. Section III summarizes the data used in the estimation of the model, focusing on the data on work limitation status. Section IV discusses the identification strategy, presents the estimates of the structural parameters, and discusses both the internal and external fit of the model in a number of key dimensions. Section V carries out counterfactual policy experiments, reporting the effects on behavior and average household welfare of potential reforms of DI, along with sensitivity tests of these experiments. Section VI concludes and discusses limitations and directions for future work. The online Appendix contains further robustness checks and experiments.

I. Literature Review

The literature on DI has evolved in three different directions: (i) papers that estimate, typically in a reduced form way, the disincentive effects of the DI program; (ii) papers that estimate, again using reduced form strategies, the welfare benefits of the program; and (iii) papers that estimate structural models in order to evaluate the welfare consequences of reforming the program. Our paper belongs to the third line of research but we stress the importance of matching evidence from the first and second lines.

Incentive Effects of DI.—There is an extensive literature estimating the costs of the DI program in terms of inefficiency of the screening process and the disincentive effects on labor supply decisions.

Since disability status is private information, there are errors involved in the screening process. The only direct attempt to measure such errors is Nagi (1969), who uses a sample of 2,454 initial disability determinations. These individuals were examined by an independent medical and social team. Nagi (1969) concluded that, at the time of the award, about 19 percent of those initially awarded benefits were undeserving, and 48 percent of those denied were truly disabled. To the extent that individuals recover but do not flow off DI, we would expect the fraction falsely claiming to be higher in the stock than at admission. This is the finding of Benítez-Silva, Buchinsky, and Rust (2006a) who use self-reported disability data on those aged over 50 from the Health and Retirement Study (HRS): over 40 percent of recipients of DI are not truly work limited. We compare these estimates of the screening errors to the estimates of our model. These errors raise the question of whether the "cheaters" are not at all disabled or whether they have only a partial work limitation. With our distinction between severe work limitations and moderate limitations, we are able to explore this issue. Moreover, we assume that disability evolves over the life cycle, which allows for both medical recoveries and further health declines.

In terms of labor supply effects, the incentive for individuals to apply for DI rather than to work has been addressed by asking how many DI recipients would be in the labor force in the absence of the program.[2] Identifying an appropriate control group has proved difficult (see Parsons 1980; Bound 1989). Bound (1989) uses rejected DI applicants as a control group and finds that only one-third to one-half of rejected applicants are working, and this is taken as an upper bound of how many DI beneficiaries would be working in the absence of the program. This result has proved remarkably robust. Chen and van der Klaauw (2008) report similar magnitudes. As do French and Song (2014) and Maestas, Mullen, and Strand (2013), who use the arguably more credible control group of workers who were not awarded benefits because their application was examined by "tougher" disability examiners (as opposed to similar workers whose application was examined by more "lenient" adjudicators). In addition, von Wachter, Song, and Manchester (2011) stress that there is heterogeneity in the response to DI, and that younger, less severely disabled workers are more responsive to economic incentives than the older groups usually analyzed. Further, this growth in younger claimants has been a key change in the composition of claimants since 1984.[3] We compare the implied elasticity of employment with respect to benefit generosity that comes from our model with the estimates of such elasticity in the literature.

[2] Some of the costs of the program derive from beneficiaries staying on the program despite health improvements. Evidence on the effectiveness of incentives to move the healthy off DI is scant: Hoynes and Moffitt (1999) conclude via simulations that some of the reforms aimed at allowing DI beneficiaries to keep more of their earnings on returning to work are unlikely to be successful and may, if anything, increase the number of people applying for DI.

[3] These incentive effects have implications for aggregate unemployment. Autor and Duggan (2003) find that the DI program lowered measured US unemployment by 0.5 percentage points between 1984 and 2001 as individuals moved onto DI. They argue that this movement was firstly because the rise in wage inequality in the United States, coupled with the progressivity of the formula used to compute DI benefits, implicitly increased replacement rates for people at the bottom of the wage distribution (increasing demand for DI benefits); and second, because in 1984 the program was reformed and made more liberal (increasing the supply of DI benefits).

A further dimension of the incentive cost of the program is the possibility that poor labor market conditions (such as declines in individual productivity due to negative shocks to skill prices or low arrival rates of job offers), increase applications for the DI program. Black, Daniel, and Sanders (2002) use the boom and bust in the mining industry in some US states (induced by the exogenous shifts in coal and oil prices of the 1970s) to study employment decisions and participation in the DI program. They show that participation in the DI program is much more likely for permanent than transitory skill shocks. In our framework, we distinguish between these different types of shocks.

Estimates of the Benefits of the Program.—The literature on the welfare benefits of DI is more limited. Some papers (e.g., Meyer and Mok 2014, and Stephens 2001, for the United States; and Ball and Low 2014, for the United Kingdom) first quantify the amount of health risk faced by workers and then measure the value of insurance by looking at the decline in consumption that follows a poor health episode. Chandra and Sandwick (2006) use a standard life-cycle model, add disability risk (which they model as a permanent, involuntary retirement shock) and compute the consumer's willingness to pay to eliminate such risk. These papers interpret any decline in consumption in response to uninsured health shocks as a measure of the welfare value of insurance, ignoring the question of whether preferences are nonseparable and health-dependent. However, consumption may fall optimally even if health shocks are fully insured, for example because consumption needs are reduced when sick, leading to consumption and poor health being substitutes in utility. We allow explicitly for health-dependent preferences which provides a better assessment of the welfare benefits of the DI program.

The Value of Reforming the DI Program.—The broader issue of the value of DI and the effects of DI reform requires combining estimates of the risk associated with health shocks alongside the evaluation of the insurance and incentives provided by DI. Similar to our paper, previous work by Bound et al. (2004); Bound, Stinebrickner, and Waidmann (2010); Benítez-Silva, Buchinsky, and Rust (2006b); and Waidmann, Bound, and Nichols (2003) has also highlighted the importance of considering both sides of the insurance/incentive trade-off for welfare analysis and conducted some policy experiments evaluating the consequences of reforming the program. These papers differ in focus and this leads to differences in the way preferences, risk, and the screening process are modeled; and in the data and estimation procedure used.[4]

Benítez-Silva, Buchinsky, and Rust (2006b) use the HRS and focus on older workers. Their model is used to predict the implications of introducing the "$1 for $2 benefit offset," i.e., a reduction of $1 in benefits for every $2 in earnings a DI beneficiary earns above the "substantial gainful activity" (SGA) ceiling. Currently, there is a 100 percent tax (people get disqualified for benefits if earning more than

[4] There is a purely theoretical literature on optimal disability insurance, such as the model of Diamond and Sheshinski (1995) and the Golosov and Tsyvinsky (2006) result on the desirability of asset testing DI benefits. Our focus is on the estimation of the value and incentives of the actual DI program. We relate our results to the theoretical literature in Section V.

the SGA). The effect of the reform is estimated to be small. Their model is very detailed in numerous dimensions, but one important caveat is that there is no disaggregation by health. As stressed by von Wachter, Song, and Manchester (2011), behavioral responses to incentives in the DI program differ by age and by health status, with the young being the most responsive.

The paper closest to ours is Bound, Stinebrickner, and Waidmann (2010). They specify a dynamic programming model that looks at the interaction of health shocks, disposable income, and the labor market behavior of men. The innovative part of their framework is that they model health as a continuous latent variable for which discrete disability is an indicator. This is similar to our focus on different degrees of severity of health shocks. However, the focus of their paper is on modeling behavior among the old (aged 50 and over from the HRS), rather than over the whole life cycle. Further, the decline in labor market participation among the old is not disaggregated by health status and does not match the decline in the data. The point of our paper is that we need a life-cycle perspective to capture fully the insurance benefits, and we need an accurate characterization both of labor supply behavior and applications to the program to capture fully the incentive costs of the program.

II. Life-Cycle Model

A. *Individual Problem*

We consider the problem of an individual who maximizes lifetime expected utility:

$$\max_{c,P,DI^{App}} V_{it} = E_t \sum_{s=t}^{T} \beta^{s-t} U(c_{is}, P_{is}; L_{is}),$$

where β is the discount factor, E_t the expectations operator conditional on information available in period t (a period being a quarter of a year), P a discrete $\{0,1\}$ employment indicator, c_t consumption, and L_t a discrete work limitation (disability) status indicator $\{0,1,2\}$, corresponding to no limitation, a moderate limitation and a severe limitation, respectively. Work limitation status is often characterized by a $\{0,1\}$ indicator (as in Benítez-Silva, Buchinsky, and Rust 2006a). We use a three state indicator to investigate the importance of distinguishing between moderate and severe work limitations. Individuals live for T periods, may work T^W years (from age 23 to 62), and face an exogenous mandatory spell of retirement of $T^R = 10$ years at the end of life. The date of death is known with certainty and there is no bequest motive.

The intertemporal budget constraint during the working life has the form

$$\begin{aligned} A_{it+1} = R\Big[&A_{it} + \big(w_{it}h(1-\tau_w) - F(L_{it})\big)P_{it} \\ &+ \big(B_{it}Z_{it}^{UI}(1-Z_{it}^{DI}) + D_{it}Z_{it}^{DI} + SSI_{it}Z_{it}^{DI}Z_{it}^{W}\big)(1-P_{it}) + W_{it}Z_{it}^{W} - c_{it}\Big], \end{aligned}$$

where A is the beginning of period assets, R is the interest factor, w the hourly wage rate, h a fixed number of hours (corresponding to 500 hours per quarter), τ_w

a proportional tax rate that is used to finance social insurance programs, F the fixed cost of work that depends on disability status, B unemployment benefits, W the monetary value of a means-tested welfare payment, D the amount of disability insurance payments obtained, SSI the amount of Supplemental Security Income (SSI) benefits, and Z^{DI}, Z^{UI}, and Z^{W} are recipiency $\{0, 1\}$ indicators for disability insurance, unemployment insurance, and the means-tested welfare program, respectively.[5] We assume that unemployment insurance is paid only on job destruction and only for one quarter; the means-tested welfare program is an anti-poverty program providing a floor to income, similar to food stamps, and this is how we will refer to it in the rest of the paper. Recipiency Z_{it}^{W} depends on income being below a certain (poverty) threshold. The way we model both programs is described fully in the online Appendix.

The worker's problem is to decide whether to work or not. When unemployed, the decision is whether to accept a job that may have been offered or wait longer. The unemployed person will also have the option to apply for disability insurance (if eligible). Whether employed or not, the individual has to decide how much to save and consume. Accumulated savings are used to finance consumption at any time, particularly during spells out of work and retirement.

We assume that individuals are unable to borrow: $A_{it} \geq 0 \, \forall t$. This constraint has bite because it precludes borrowing against social insurance and means-tested programs. At retirement, people collect Social Security benefits which are paid according to a formula similar to the one we observe in reality, and is the same as the one used for DI benefits (see below). Social Security benefits, along with assets that people have voluntarily accumulated over their working years, are used to finance consumption during retirement. The structure of the individual's problem is similar to life-cycle models of savings and labor supply, such as Low, Meghir, and Pistaferri (2010). The innovations in our setup are to consider the risk that arises from work limitation shocks, distinguishing between the severity of the shocks, the explicit modeling of disability insurance, and the interaction of disability insurance with welfare programs.

While eligibility and receipt of disability insurance are not means-tested, in practice high-education individuals are rarely beneficiaries of the program. In our PSID dataset individuals with low and high education have similar DI recipiency rates only until their mid-30s (about 1 percent), but after that age, the difference between the two groups increases dramatically. By age 60, the low educated are 2.5 times more likely to be DI claimants than the high educated (17 percent versus 7 percent).[6] Figure 4 in the online Appendix provides the details. Given these large differences, in the remainder of the paper we focus on low-education individuals (those with at most a high school degree), with the goal of studying the population group that is more likely to be responsive to changes in the DI program. We do however introduce heterogeneity in individual productivity: as detailed in the subsection on

[5] We do not have an SSI recipiency indicator because that is a combination of receiving DI and being eligible for means-tested transfers.

[6] The low DI participation rates among the high educated is partly due to the vocational criterion used by the Social Security Administration (SSA) for awarding DI (described later).

wages below, individuals differ ex ante in terms of the level of productivity as well as differing ex post due to idiosyncratic shocks.

While our model is richer than existing characterizations in many dimensions, there are certain limitations. First, we model individual behavior rather than family behavior and hence neglect insurance coming from, for example, spousal labor supply. On the other hand, we assume that social insurance is always taken up when available. Second, in our model health shocks result in a decline in productivity which indirectly affects consumption expenditure, but we ignore direct health costs (i.e., drugs and health insurance) that may shift the balance across consumption spending categories. Third, we do not allow for health investments which may reduce the impact of a health shock. This assumption makes health risk independent of the decision process and so can be estimated outside of the model. In practice most heterogeneity in health investment occurs between education groups. On the other hand, we allow the transition matrix describing health shocks to differ according to an individual's type.

We now turn to a discussion of the three key elements of the problem: (i) preferences, (ii) wages, and (iii) social insurance.

B. *Preferences*

We use a utility function of the form

$$u(c_{it}, P_{it}; L_{it}) = \frac{\left(c_{it}\exp(\theta L_{it})\ \exp(\eta P_{it})\right)^{1-\gamma}}{1-\gamma}. \tag{1}$$

To be consistent with disability and work being "bads," we require $\theta < 0$ and $\eta < 0$, two restrictions that as we shall see are not rejected by the data. The parameter θ captures the utility loss for the disabled in terms of consumption. Employment also induces a utility loss determined by the value of η. This implies that consumption and work are Frisch complements (i.e., the marginal utility of consumption is higher when working) and that the marginal utility of consumption is higher when suffering from a work limitation.[7]

If individuals were fully insured, they would keep marginal utility constant across states. $\theta < 0$ implies that individuals who are fully insured want more expenditure allocated to the "disability" state, for example because they have larger spending needs when disabled (alternative transportation services, domestic services, etc.).[8]

Consumption in equation (1) is equivalized consumption. We introduce demographics by making household size at each age mimic the average family size in the data (rounded to the nearest integer). We then equivalize consumption in the utility function using the OECD equivalence scale.

[7] In addition to the nonseparable effect of disability, there may be an additive utility loss associated with disability. Since disability is not a choice, we cannot identify this additive term. Further, such an additive utility loss would be uninsurable because only consumption can be substituted across states.

[8] Lillard and Weiss (1997) also find evidence for $\theta < 0$ using HRS savings and health status data. On the other hand, Finkelstein, Luttmer, and Notowidigdo (2013) use health data and subjective well-being data to proxy for utility and find $\theta > 0$.

C. *The Wage Process and Labor Market Frictions*

We model the wage process as a combination of observable characteristics X, shocks to work limitation status L, general productivity (skill) shocks ε, as well as unobserved fixed heterogeneity f:

$$\ln w_{it} = X_{it}' \mu + \sum_{j=1}^{2} \varphi_j L_{it}^j + f_i + \varepsilon_{it} \tag{2}$$

where

$$\varepsilon_{it} = \varepsilon_{it-1} + \zeta_{it},$$

and $L_{it}^j = \mathbf{1}\{L_{it} = j\}$ is an indicator for work limitation status $j = \{0, 1, 2\}$.

We assume that ex ante heterogeneity f_i may be potentially correlated with the work limitation status. This captures the idea that there may be a group of individuals with both low productivity and high propensity to develop a disability. In Section V we discuss estimation of the parameters of (2). While in estimation f_i is continuous, in the simulations we assume that there are three discrete "types" of workers, corresponding to the bottom quartile, the two middle quartiles, and the top quartile of the distribution of f_i.

We assume that the work limitation status of an individual evolves according to a three state first-order Markov process. Upon entry into the labor market, all individuals are assumed to be healthy ($L_{i0} = 0$). Transition probabilities from any state depend on age and the unobserved heterogeneity type. These transition probabilities are assumed to be exogenous (conditional on type).

Finally, we interpret ε_{it} as a measure of time-varying individual unobserved productivity that is independent of health shocks—examples would include shocks to the value and price of individual skills—and interpret ζ_{it} as a permanent productivity shock.

Equation (2) determines the evolution of individual productivity. Productivity determines the offered wage when individuals receive a job offer. The choice about whether or not to accept an offered wage depends in part on the fixed costs of work, which in turn also depends on the extent of the work limitation, $F(L)$. In addition, there are labor market frictions which means that not all individuals receive job offers. First, there is job destruction, δ, which forces individuals into unemployment for (at least) one period. Second, job offers for the unemployed arrive at a rate λ and so individuals may remain unemployed even if they are willing to work.

This wage and employment environment implies a number of sources of risk, from individual productivity, work limitation shocks, and labor market frictions. These risks are idiosyncratic, but we assume that there are no markets to provide insurance against these risks. Instead, there is partial insurance coming from government insurance programs (as detailed in the next section) and from individuals' own saving and labor supply.

D. *Social Insurance*

The DI Program.—The Social Security Disability Insurance program (DI) is an insurance program for covered workers, their spouses, and dependents that pays benefits related to average past earnings. The purpose of the program is to provide insurance against persistent health shocks that impair substantially the ability to work. The difficulty with providing this insurance is that health status and the impact of health on the ability to work is imperfectly observed. The policy we focus on is the program in place since the major reform of 1984, although the program has gone through minor revisions since. It would be interesting to allow for the policy itself to be stochastic, but that is beyond the scope of this paper.

The award of disability insurance depends on the following conditions: (i) an individual must file an application; (ii) there is a work requirement on the number of quarters of prior employment: workers over the age of 31 are disability-insured if they have 20 quarters of coverage during the previous 40 quarters;[9] (iii) there is a statutory five-month waiting period out of the labor force from the onset of disability before an application will be processed; and (iv) the individual must meet a medical requirement, i.e., the presence of a disability defined as "Inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment, which can be expected to result in death, or which has lasted, or can be expected to last, for a continuous period of at least 12 months (Social Security Administration 2014, p. 2)."[10]

The actual DI determination process consists of sequential steps. After excluding individuals earning more than a so-called "substantial gainful amount" ((SGA) $1,010 a month for non-blind individuals as of 2012), the SSA determine whether the individual has a medical disability that is severe and persistent (per the definition above).[11] If such disability is a listed impairment, the individual is awarded benefits without further review.[12] If the applicant's disability does not match a listed impairment, the DI evaluators try to determine the applicant's residual functional capacity. In the last stage the pathological criterion is paired with an economic opportunity criterion. Two individuals with identical work limitation disabilities may receive

[9] There are two tests that individuals must pass that involve work credits: the "recent work test" and the "duration of work test." The "recent work test" requires that individuals aged 31+ have worked at least 5 of the last 10 years. The "duration of work test" requires people to have worked a certain fraction of their lifetime. For people aged 40+, representing the bulk of DI applications, the fraction of their lifetime that they need to have worked is about 25 percent.

[10] Despite this formal criterion changing very little, there have been large fluctuations over time in the award rates: for example, award rates fell from 48.8 percent to 33.3 percent between 1975 and 1980, but then rose again quickly in 1984, when eligibility criteria were liberalized, and an applicant's own physician reports were used to determine eligibility. In 1999, a number of work incentive programs for DI beneficiaries were introduced (such as the Ticket to Work program) in an attempt to push some of the DI recipients back to work.

[11] The criteria quoted above specifies "any substantial gainful activity": this refers to a labor supply issue. However, it does not address the labor demand problem. Of course, if the labor market is competitive this will not be an issue because workers can be paid their marginal product whatever their productivity level. In the presence of imperfections, however, the wage rate associated with a job may be above the disabled individual's marginal productivity. The Americans with Disability Act (1990) tries to address this question but tackles the issue only for incumbents who become disabled.

[12] The listed impairments are described in a blue-book published and updated periodically by the SSA ("Disability Evaluation under Social Security"). They are physical and mental conditions for which specific disability approval criteria has been set forth or listed (for example, amputation of both hands, heart transplant, or leukemia).

different DI determination decisions depending on their age, education, general skills, and even economic conditions faced at the time the determination is made.

In our model, we make the following assumptions in order to capture the complexities of the disability insurance screening process. First, we require that the individuals make the choice to apply for benefits. Second, individuals have to have been at work for at least the period prior to becoming unemployed and making the application.[13] Third, individuals must have been unemployed for at least one quarter before applying. Successful applicants begin receiving benefits in that second quarter. Unsuccessful individuals must wait a further quarter before being able to return to work, but there is no direct monetary cost of applying for DI. Finally, we assume that the probability of success depends on the individual's work limitation status and age:

$$(3) \qquad \Pr\left(DI_{it} = 1 \,|\, DI_{it}^{App} = 1, L_{it}, t\right) = \begin{cases} \pi_L^{Young} & \text{if } t < 45 \\ \pi_L^{Old} & \text{if } 45 \leq t \leq 62. \end{cases}$$

We make the probability of a successful application for DI dependent on age because the persistence of health shocks is age dependent.[14] Individuals leave the disability program either voluntarily (which in practice means into employment) or following a reassessment of the work limitation and being found to be able to work (based on (3)). We depart from the standard assumption made in the literature that DI is an absorbing state because we want to be able to evaluate policies that create incentives for DI beneficiaries to leave the program.

DI beneficiaries have their disability reassessed periodically through Continuing Disability Reviews (CDR). By law, the SSA is expected to perform CDRs every 7 years for individuals where medical improvement is not expected, every 3 years for individuals where medical improvement is possible, and every 6 to 18 months for individuals where medical improvement is expected. In this way, the probability of reassessment depends on perceived work limitation status. To capture this, we would ideally allow the probability of reassessment to vary with the assessment of true health status that the SSA made on acceptance onto the program, with the most healthy-seeming reassessed most quickly. We approximate this by setting the probability of being reassessed, P_L^{Re}, to be 0 for the first year, then varying the assessment rate with true work limitation status, L.

DI benefits are calculated in essentially the same fashion as Social Security retirement benefits. Beneficiaries receive indexed monthly payments equal to their Primary Insurance Amount (PIA), which is based on taxable earnings averaged over the number of years worked (known as AIME). Benefits are independent of the

[13] This eligibility requirement is weaker than the actual requirement. We check in our simulations how many applicants would satisfy the requirement to have worked at least 50 percent of possible quarters. In our simulations below, 96 percent of applicants satisfy this requirement. Further, 99 percent of applicants have worked at least 25 percent of possible quarters.

[14] The separation at age 45 takes into account the practical rule followed by DI evaluators in the last stage of the DI determination process (the so-called Vocational Grid, see Appendix 2 to Subpart P of Part 404—Medical-Vocational Guidelines, as summarized in Chen and van der Klaauw 2008).

extent of the work limitation, but are progressive.[15] We set the value of the benefits according to the actual schedule in the US program (see the online Appendix).

We assume that the government awards benefits to applicants whose signal of disability exceeds a certain stringency threshold. Some individuals whose actual disability is less severe than the threshold may nonetheless wish to apply for DI if their productivity is sufficiently low because the government only observes a noisy measure of the true disability status. In contrast, some individuals with true disability status above the threshold may not apply because they are highly productive despite their disability. Given the opportunity cost of applying for DI, these considerations suggest that applicants will be predominantly low-productivity individuals or those with severe work limitations (see Black, Daniels, and Sanders 2002, for a related discussion).

Supplemental Security Income (*SSI*).—Individuals who are deemed to be disabled according to the rules of the DI program and who have income (comprehensive of DI benefits but excluding the value of food stamps) below the threshold that would make them eligible for food stamps, receive also supplemental security income (SSI). The definition of disability in the SSI program is identical to the one for the DI program, while the definition of low income is similar to the one used for the food stamps program.[16] We assume that SSI generosity is identical to the food stamps program described in the online Appendix.

E. *Solution*

There is no analytical solution for our model. Instead, the model must be solved numerically, beginning with the terminal condition on assets, and iterating backwards, solving at each age for the value functions conditional on work status. The solution method is discussed in detail in the online Appendix, which also provides the code to solve and simulate the model. The approach is similar to Attanasio, Low, and Sanchez-Marcos (2008) and Low, Meghir, and Pistaferri (2010).

III. Data

The ideal dataset for studying the issues discussed in our model is a longitudinal dataset covering the entire life cycle of an individual, while at the same time containing information on consumption, wages, employment, disability status, the decision to apply for DI, and information on receipt of DI. Unfortunately, none of the US datasets typically used by researchers working on DI satisfy all these requirements at once. Most of the structural analyses of DI errors have used data from the HRS or the Survey of Income and Program Participation (SIPP). The advantage of the

[15] Caps on the amount that individuals pay into the DI system as well as the nature of the formula determining benefits make the system progressive. Because of the progressivity of the benefits and because individuals receiving DI also receive Medicare benefits after two years, the replacement rates are substantially higher for workers with low earnings and those without employer-provided health insurance.

[16] In particular, individuals must have income below a "countable income limit," which typically is slightly below the official poverty line (Daly and Burkhauser 2003). As in the case of food stamp eligibility, SSI eligibility also has an asset limit which we disregard.

HRS is that respondents are asked very detailed questions on disability status and DI application, minimizing measurement error and providing a direct (reduced form) way of measuring screening errors. However, there are three important limitations of the HRS. First, the HRS samples only from a population of older workers and retirees (aged above 50). In Figure 6 of the online Appendix, we show that in recent years an increasing fraction of DI awards have gone to younger individuals, which highlights that capturing the behavior of those under 50 is an important part of our understanding of disability insurance, as also discussed in von Wachter, Song, and Manchester (2011). Second, the HRS asks questions about application to DI only to those individuals who have reported having a work limitation at some stage in their life course. Finally, the HRS has no consumption data. The SIPP has the advantage of being a large dataset covering the entire life cycle, but it also lacks consumption data. This is problematic because an important element of our model is the state dependence in utility induced by health. Moreover, the longitudinal structure of the SIPP makes it difficult to link precisely the timing of wages with those of changes in work limitations.

Our choice is to use all the waves of the PSID between 1986 and 2009.[17] Data are collected annually between 1985 and 1997, and bi-annually after 1997. The PSID offers repeated, comparable data on disability status, disability insurance recipiency, wages, employment, and consumption. The quality of the data is comparable to SIPP and HRS and the panel is long.

One important advantage of the PSID over the SIPP and the HRS is that (at least in recent waves) it contains rich information on household consumption. In particular, before the 1999 wave, the only measure of consumption available was food. Starting with the revision of the survey in 1999, however, a more comprehensive measure of consumption was collected—which included information on utilities, gasoline and other vehicle expenses, transportation, health expenditures, education, child care, and housing.[18] The main items that are missing are clothing, recreation, alcohol, and tobacco.[19] We treat missing values in the consumption subcategories as zeros and aggregate all nondurable and services consumption categories to get the household consumption series. Blundell, Pistaferri, and Saporta-Eksten (2014) discuss descriptive statistics on the various components of aggregate consumption and how it compares with national accounts (see Table 2 in the online Appendix). To get as close as possible to the consumption concept of the model, our consumption regressions only use the 1999–2009 PSID waves containing the more comprehensive measure of household consumption.

There are also disadvantages from using the PSID, and here we discuss how important they are and what we do to tackle them. The first problem is that the sample of people likely to have access to disability insurance is small. Nevertheless,

[17] Due to the retrospective nature of the questions on earnings and consumption, this means our data refer to the 1985–2008 period. We do not use data before 1985 because major reforms in the DI screening process were implemented in 1984 (see Autor and Duggan 2003; and Duggan and Imberman 2009).

[18] While housing rent is reported for tenants, there is no information on housing services for homeowners. To construct a series of housing services for homeowners we impute rent expenditures using the self reported house price and assume that the rent equivalent is 6 percent of the self-reported house price (see Flavin and Yamashita 2002).

[19] Other consumption categories have been added starting in 2005 (such as clothing). We do not use these categories to keep the consumption series consistent over time.

it is worth noting that estimates of disability rates in the PSID are similar to those obtained in other, larger datasets (CPS, SIPP, NHIS—and HRS conditioning on age, see Bound and Burkhauser 1999, and Figure 2 in the online Appendix). Moreover, PSID DI rates by age and over time compare well with administrative data. For rates by age, see the online Appendix, Figure 3. For rates over time, consider that in the population the proportion of male workers on DI has increased from 2.46 percent to 4.98 percent between 1986 and 2008; in the PSID the increase over the same time period is almost identical, 2.10 percent to 4.97 percent.

The second problem is that the PSID does not provide information on DI application. We use our indirect inference procedure to circumvent this problem: For a given set of structural parameters, we simulate DI application decisions and the resulting moments that reflect the DI application decision (such as DI recipiency by age and disability status, disability state of DI recipients by age, and transitions into the program). These moments, crucially, can be obtained both in the actual and simulated data and the fit of these moments is an explicit way of checking how well our model approximates the decision to apply for DI.

Finally, the frequency with which data are collected switches from annual to biannual starting with the 1999 wave. In some cases (estimation of year-to-year transitions across disability categories) we use only the data before 1999; in other cases (estimation of the consumption equation) we use only the data since 1999 because of more comprehensive information; and in some cases we use the entire sample period (estimation of the wage process). Additionally, the timing of disability status and DI recipiency are not synchronized: Disability status refers to the time of the interview, while DI recipiency (and earnings) refers to the previous calendar year. We use longitudinal information to align the timing of the information available. We describe these various choices below whenever relevant.

The PSID sample we use excludes the Latino subsample, female heads, and people younger than 23 or older than 62. Further sample selection restrictions are discussed in the online Appendix.[20]

Disability Data.—We define a discrete indicator of work limitation status (L_{it}), based on the following set of questions: (i) *Do you have any physical or nervous condition that limits the type of work or the amount of work you can do?* To those answering "Yes," the interviewer then asks: (ii) *Does this condition keep you from doing some types of work?* The possible answers are: "Yes," "No," or "Can do nothing." Finally, to those who answer "Yes" or "No," the interviewer then asks: (iii) *For work you can do, how much does it limit the amount of work you can do?* The possible answers are: "A lot," "Somewhat," "Just a little," or "Not at all,"

We assume that those without a work limitation ($L_{it} = 0$) either answer "No" to the first question or "Not at all" to the third question. Of those that answer "Yes" to the first question, we classify as severely limited ($L_{it} = 2$) those who answer question 2 that they "can do nothing" and those that answer question 3 that they are limited "a lot." The rest have a moderate limitation ($L_{it} = 1$): their answer to

[20] While PSID data refer to a calendar year, our model assumes that the decision period is a quarter, as events like unemployment, wage shocks, etc., happen at a frequency that is shorter than the year. We match timing in the model with that available in the data by converting quarterly data in our simulations into the frequency of the PSID.

question 3 is that they are limited either "Somewhat" or "Just a little." This distinction between severe and moderate disability enables us to target our measure of work limitation more closely to that intended by the SSA.[21] In particular, we interpret the SSA criterion as intending DI for the severely work limited rather than the moderately work limited.[22]

The validity of work limitation self-reports is somewhat controversial for three reasons. First, subjective reports may be poorly correlated with objective measures of disability. However, Bound and Burkhauser (1999) survey a number of papers that show that self-reported measures are highly correlated with *clinical* measures of disability. We provide additional evidence in support of our self-reported measure of work limitation in Table 1 in the online Appendix.

Second, individuals may over-estimate their work limitation in order to justify their disability payments or their nonparticipation in the labor force. Benítez-Silva et al. (2004) show that self-reports are unbiased predictors of the definition of disability used by the SSA ("norms"). In other words, there is little evidence that, for the sample of DI applicants, average disability rates as measured from the self-reports are significantly higher than disability rates as measured from the SSA final decision rules. However, Kreider (1999) provides evidence based on bound identification that disability is over-reported among the unemployed.

Third, health status may be endogenous, and nonparticipation in the labor force may affect health (either positively or negatively). Stern (1989) and Bound (1991) both find positive effects of nonparticipation on health, but the effects are economically small. Further, Smith (2004) finds that income does not affect health once one controls for education (as we do implicitly by focusing on a group of homogeneous individuals with similar schooling levels). Similarly, Adda, Banks, and von Gaudecker (2009) find that innovations to income have negligible effects on health.

Sample Summary Statistics.—Table 1 reports descriptive statistics for our sample (pooling data for all years), stratifying it by the degree of work limitation. The severely disabled are older and less likely to be married or white. They have lower family income but higher income from transfers (most of which come from the DI or SSI program). They are less likely to work, have lower earnings if they do so, are more likely to be a DI recipient, and have lower consumer spending than people without a disability.

These statistics underpin the moments used in the indirect inference estimation. Two particularly important descriptive statistics are the fraction of DI recipients who are not severely disabled ("false claimants") and the fraction of individuals with a severe disability who receive DI ("coverage"). Figure 1 plots the life-cycle patterns

[21] Our three-way classification uses the responses to the multiple questions (1)–(3), and hence reduces the measurement error associated with using just the "Yes/No" responses associated to question (1). An alternative way to reduce such error is to classify as disabled only those who answer "Yes" to question (1) for two consecutive years, as in Burkhauser and Daly (1996).

[22] The distinction between moderate and severe disability is a key step in achieving identification of the error rates in the DI application process. However, our distinction does not take into account that the vocational criterion of DI implies that eligibility potentially varies across time and space for workers with similar disabilities because of market conditions. On the other hand, as noticed by Benítez-Silva et al. (2004), these measures have the unique advantage of being sufficient statistics for use in the structural modeling of individual behavior under disability risk.

TABLE 1—SUMMARY STATISTICS BY WORK LIMITATION STATUS

Variable	$L = 0$	$L = 1$	$L = 2$
Age	38.88	44.05	47.30
Percent married	0.78	0.77	0.69
Percent white	0.58	0.65	0.54
Large SMSA	0.48	0.49	0.47
Family size	3.23	3.14	2.94
Family income	46,446	39,780	25,897
Income from transfers	1,794	5,091	8,281
Percent employed at the time of interview	0.91	0.61	0.11
Percent annual wages > 0	0.96	0.72	0.24
Hours | Hours > 0	2,163	1,913	1,510
Wages | Hours > 0	30,539	26,463	18,478
Percent DI recipient	0.01	0.13	0.45
Total food (missing in 1987–1988)	5,510	5,883	4,060
Total spending (1998–2008)	24,682	25,738	18,286
Observations	19,682	1,739	1,532

Note: Monetary values are in 1996 dollars.

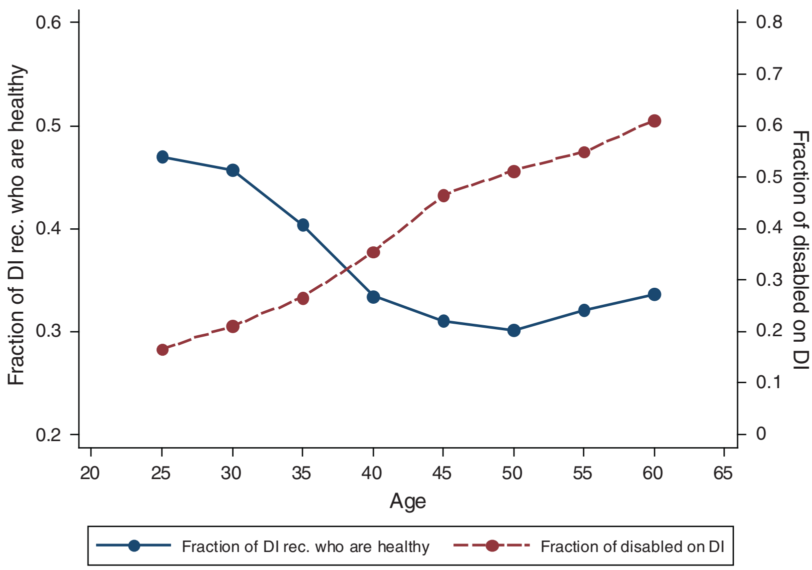

FIGURE 1. COVERAGE VERSUS FALSE CLAIMANTS

for each: the fraction of claimants who are healthy is particularly high early in the life cycle, while "coverage" becomes more effective at the end of the working life cycle. This suggests the DI program is less effective at screening younger workers.

IV. Identification and Results

Identification of the unknown parameters proceeds in three steps. First, some parameters are predetermined or calibrated using established findings from the literature. We check the sensitivity of our policy experiment results to assuming different values for key predetermined parameters. Second, some parameters are estimated outside the structure of the model. For some parameters, this is because no structure is needed: disability risk can be estimated directly from transitions between disability states because of the exogeneity assumption. For other parameters, we use a reduced form approach to reduce the computational burden when there are plausible selection correction processes, as is the case for the wage parameters. The remaining parameters are estimated structurally using an indirect inference procedure.

This mixed identification strategy is not novel to our paper. For example, to make estimation feasible, Bound, Stinebrickner, and Waidmann (2010) estimate, in a context very similar to ours, the parameters of the earnings equations and health equations outside the behavioral model. This mixed strategy has been used more generally in a number of papers looking at consumption choices under uncertainty: Gourinchas and Parker (2002); Attanasio et al. (1999); Low, Meghir, and Pistaferri (2010); Alan and Browning (2009); and Guvenen and Smith (2011).

A. *Predetermined and Calibrated Parameters*

We fix the relative risk aversion coefficient γ and the intertemporal discount rate β to realistic values estimated elsewhere in the literature. In principle, one could identify γ and β using asset data. We use the asset data available in the PSID at certain intervals to test the out-of-sample behavior of our model.

We set $\gamma = 1.5$ in our baseline and we later examine the sensitivity of our results to setting $\gamma = 3$.[23] As for the estimate of β, we use the central value of estimates from Gourinchas and Parker (2002) and Cagetti (2003), two representative papers of the literature and set $\beta = 0.9756$ on an annual basis.[24]

In principle, the arrival rate of offers when unemployed (λ) parameter could be identified using unemployment duration by age and disability states. However, there are important censoring issues, and hence we take the estimate of λ from Low, Meghir, and Pistaferri (2010), who use a very similar empirical strategy and estimate a quarterly arrival rate $\lambda = 0.73$.

We allow the reassessment rate of disability status to vary with true work limitation status to approximate the approach and frequency that the SSA follows with its CDRs. Therefore, $P^{RE}_{L=2} = 0.036$, $P^{RE}_{L=1} = 0.083$, and $P^{RE}_{L=0} = 0.222$. If we weight these probabilities by the numbers on DI in each health category, we obtain an unconditional probability of reassessment equal to 0.066. This is very similar to the reported aggregate rate of the SSA.

Finally, we set the interest factor to a realistic value, $R = 1.016$ (on an annual basis), and assume that a life span is 50 years, from age 22, with the last 10 years in compulsory retirement.

[23] Attanasio et al. (1999); Blundell, Browning, and Meghir (1994); Attanasio and Weber (1995); and Banks, Blundell, and Brugiavini (2001); report estimates of 1.35, 1.37, 1.5, and 1.96, respectively. Our choice $\gamma = 1.5$ is a central value of these estimates.

[24] Both use annual data and we convert their annual discount rate in a quarterly discount rate. The estimates we use from their papers refer to their low education (high school or less) sample. Gourinchas and Parker's (2002) estimate is 0.988; Cagetti's (2003) estimates range between 0.987 and 0.951 depending on the definition of wealth, the dataset used (PSID and SCF), and whether mean or median assets are used.

B. *The Wage Process and Productivity Risk*

For estimation, we augment the wage process (2) to include an additional error term ω_{it}:

$$\ln w_{it} = X_{it}'\mu + \sum_{j=1}^{2} \varphi_j L_{it}^j + f_i + \varepsilon_{it} + \omega_{it} \tag{4}$$

with $\varepsilon_{it} = \varepsilon_{it-1} + \zeta_{it}$ as before. We assume that ω_{it} reflects measurement error. We do this because measurement error is not separately identifiable from transitory shocks. Despite the lack of transitory shocks in wages, there will be transitory shocks to *earnings* because of the frictions which induce temporary loss of income for a given productivity level. We make the assumption that the two errors ζ_{it} and ω_{it} are independent. Our goal is to identify the variance of the productivity shock, σ_ζ^2, the effect of disability on productivity, φ_1 and φ_2, and the distribution of unobserved heterogeneity types.

There are two issues to tackle in the empirical estimation of (4). The first is potential correlation between the fixed unobserved heterogeneity and the work limitation variable. A standard solution to this problem is to remove the fixed effect by differencing the data. A second complication is selection effects because wages are not observed for those who do not work and the decision to work depends on the wage offer. Further, the employment decision may depend directly on disability shocks as well as on the expectation that the individual will apply for *DI* in the subsequent period (which requires being unemployed in the current period). We observe neither these expectations, nor the decision to apply.

Our selection correction is based on a reduced form rather than on our structural model, although the structural model is consistent with the reduced form.[25] We assume that "potential" government transfers and its interaction with disability status serve as exclusion restrictions. The interaction accounts for the fact that the disincentive to work that government transfers are intending to capture may be different for people who have a physical cost to work. We also interact the exclusion restriction with a post-1996 welfare reform dummy. This is to account for the fact that the 1996 welfare reform may have changed the nature of the interaction between DI and social welfare programs, and hence also affected the decision to apply for DI for people with different levels of disability (see, e.g., Blank 2002). "Potential" government transfers are the sum of food stamp benefits, AFDC/TANF payments, unemployment insurance benefits, and EITC payments that individuals would receive in case of program application. These potential benefits are computed using the formulae coded in the federal (for food stamps and EITC) and state (for AFDC/TANF and UI) legislation of the programs. Full details on how we construct potential benefits are in the online Appendix. The use of this variable is in the spirit of the "simulated IV" literature in empirical public finance. In general, realized public income transfers are endogenous because the individual's take-up decision is a choice and their value may depend on past wages. Since the parameters behind

[25] Estimating the wage process jointly with preferences and DI parameters is computationally burdensome, as it would require adding several additional parameters. In the online Appendix we show that if we use our simulated data to mimic this reduced form empirical strategy, we get very similar results.

these public programs are exogenous, however, we use the amount of benefits a representative individual working part-time at the minimum wage would be eligible for if applying in his state of residence. This way, the only variation we exploit is by exogenous characteristics: state of residence, year, and demographics (number and age of children, if entering the formulae for computing benefits).

In Table 2, column 1, we report marginal effects from a probit regression for employment. Throughout the exercise, standard errors are clustered at the individual level. Employment is monotonically decreasing in the degree of work limitations. Absent potential transfers, the probability of working declines by 27 percentage points at the onset of a moderate work limitation, and by 74 percentage points at the onset of a severe work limitation. Regarding our exclusion restrictions, they are jointly statistically significant (p-value 3 percent). The disincentives to work in states with more generous welfare programs are stronger and more significant after the 1996 tax reform.

Estimation of the probit for employment allows us to construct an estimate of the inverse Mills' ratio term. We then estimate the wage growth equation only on the sample of workers. The resulting estimates of φ_1 and φ_2, with the selection correction through the inverse Mills' ratio, should be interpreted as the estimates of the effect of work limitations on *offered* wages.

Results are shown in column 2 of Table 2. The key coefficients are the ones on $\{L = 1\}$ and $\{L = 2\}$. A moderate work limitation reduces the observed wage rate by 6 percentage points, whereas a severe limitation reduces the offered wage by 18 percentage points. As we discuss in the online Appendix, ignoring selection effects and unobserved heterogeneity would induce opposite biases. In particular, selection attenuates the apparent impact of disability shocks because those who remain at work despite their work limitations have higher-than-average permanent income. By contrast, low unobserved productivity types tend to be more likely to develop disabilities, in which case the omission of fixed effects exaggerates the impact of a disability on wages.[26]

Productivity Risk.—To identify the variance of productivity shocks, we define first the "adjusted" error term:

$$\begin{aligned} g_{it} &= \Delta\left(\ln w_{it} - X_{it}'\mu - \sum_{j=1}^{2} \varphi_j L_{it}^j\right) \\ &= \zeta_{it} + \Delta\omega_{it}. \end{aligned} \tag{5}$$

From estimation of μ, φ_1, and φ_2 described above we can construct the "adjusted" residuals, and use them as if they were the true adjusted error terms (5) (MaCurdy 1982). We can then identify the variance of productivity shocks and the variance of measurement error using the first and second moments and the autocovariances of g_{it}, as discussed fully in the online Appendix. The identification idea is simple.

[26]To account for possible deviations from normality, we also experiment using a semi-parametric correction suggested by Newey (2009), detailed in more detail in the online Appendix. We find that the results remain very similar.

TABLE 2—ESTIMATING WAGE GROWTH

	Employment (1)	Wage growth (2)
$L = 2$	−0.744***	−0.177**
	(0.106)	(0.080)
$L = 1$	−0.270***	−0.057**
	(0.118)	(0.025)
Age	0.010***	0.052***
	(0.002)	(0.015)
Age sq./100	−0.016***	−0.067***
	(0.002)	(0.008)
p-value exclusion restrictions	0.032	
p-value selection corrections		0.000
Observations	22,953	17,771

Note: Clustered standard errors in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Neglect for a moment the issue of selection. With measurement error, the variance of g_{it} reflects two sources of innovations: permanent productivity shocks and measurement error. The autocovariances identify the contribution of measurement errors (which are mean-reverting), and hence the variance of productivity shocks is identified by stripping from the variance of wage growth the contribution of measurement error. Without selection, second moments conditional on working would just reflect variances of shocks. With selection, conditional variances are less than unconditional variances (which are the parameters of interest) by a factor that depends on the degree of selection in the data. Conditional means help pinning down the latter. We use the first and second moment, and the autocovariance of wage growth (conditional on working and controlling for selection) in a GMM framework to estimate the parameters of interest.

The results are in Table 3. As before, we report standard errors clustered at the individual level. The estimate of the variance of productivity shocks is 0.027 and is measured precisely. We also report, for completeness, the variance of measurement error (0.044).

Unobserved Heterogeneity.—The last part of the estimation process consists of recovering the distribution of unobserved heterogeneity in wages. To do so, we use the estimates of μ and φ_j from the difference specification reported in Table 2, and compute $\hat{f}_i = T_i^{-1}\sum_t\left(\ln w_{it} - X_{it}'\mu - \sum_{j=1}^{2}\hat{\varphi}_j L_{it}^j\right)$, where T_i is the number of years individual i is observed working. For the purpose of identifying unobserved heterogeneity "types" in the model, we divide the distribution of f_i into three parts, corresponding to low productivity (f_L, those with values of $\hat{f}_i$ in the bottom quartile), medium productivity (f_M, with a value of $\hat{f}_i$ in the intermediate 50 percent), and high productivity (f_H , a value of $\hat{f}_i$ in the top quartile). The main problem with this procedure is that $\hat{f}_i$ is unavailable for people who, during our sample period, are

Table 3—Variances of the Productivity Shocks

Parameter	Estimate
σ^2_ζ	0.027***
	(0.002)
σ^2_ω	0.044***
	(0.002)

Note: Clustered standard errors in parenthesis.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

never observed at work (4 percent of the sample). This event is likely to be strongly correlated with disability status, and we assume that these individuals are drawn from the bottom part of the distribution of unobserved productivity heterogeneity.

C. *Disability Risk*

Disability risk is independent of any choices made by individuals in our model, and independent of productivity shocks, but its evolution over the life cycle differs by heterogeneity type. This means that the disability risk process can be identified structurally without indirect inference.

In principle, since we have three possible work limitation states, there are nine possible transition patterns for each unobserved heterogeneity type $\Pr\left(L_{it} = j \mid L_{it-1} = k, f_q\right)$, $j,k = \{0,1,2\}$, $q = \{L,M,H\}$. In Figure 2 we plot only selected estimates, with the remainder reported in the online Appendix.[27] These estimates are informative about work limitation risk. For example, $\Pr\left(L_{it} = 2 \mid L_{it-1} = 0, f_H\right)$ is the probability that a high-productivity individual with no work limitations is hit by a shock that puts him in the severe work limitation category. Whether this is a persistent or temporary transition can be assessed by looking at the value of $\Pr\left(L_{it} = 2 \mid L_{it-1} = 2, f_H\right)$.

Panel A of Figure 2 plots $\Pr\left(L_{it} = 0 \mid L_{it-1} = 0, f_q\right)$, i.e., the probabilities of staying healthy by age and type. This probability declines over the working part of the life cycle, but the decline is much more rapid for the low-productivity type, even though the three types start from very similar levels. The decline is equally absorbed by increasing probabilities of transiting in moderate and severe work limitations. Panel B plots the latter, $\Pr\left(L_{it} = 2 \mid L_{it-1} = 0, f_q\right)$. This probability increases over the working life, and the increase is again faster for the low-productivity type, whose probability of moving from no disability to severe disability changes from 2 percent around age 25 to 20 percent around age 60. The probability of full recovery following a severe disability (shown in panel C) declines over the life cycle, gradually for the two top productivity types and extremely quickly for the low-productivity type. Finally, the probability of persistent severe work limitations,

[27] To obtain these plots, we regress (separately by type) an indicator for the joint event $\{L_{it} = j, L_{it-1} = k\}$ against a full set of age dummies using the sample of individuals with $L_{it-1} = k$. The predicted values of these regressions (after smoothing by simple local regression) are our estimates of the transition probabilities $\Pr(L_{it} = j \mid L_{it-1} = k, f_q)$ (and what we plot in the figure). Note that these are 1-year transition probabilities, so can only be estimated using data before 1999.

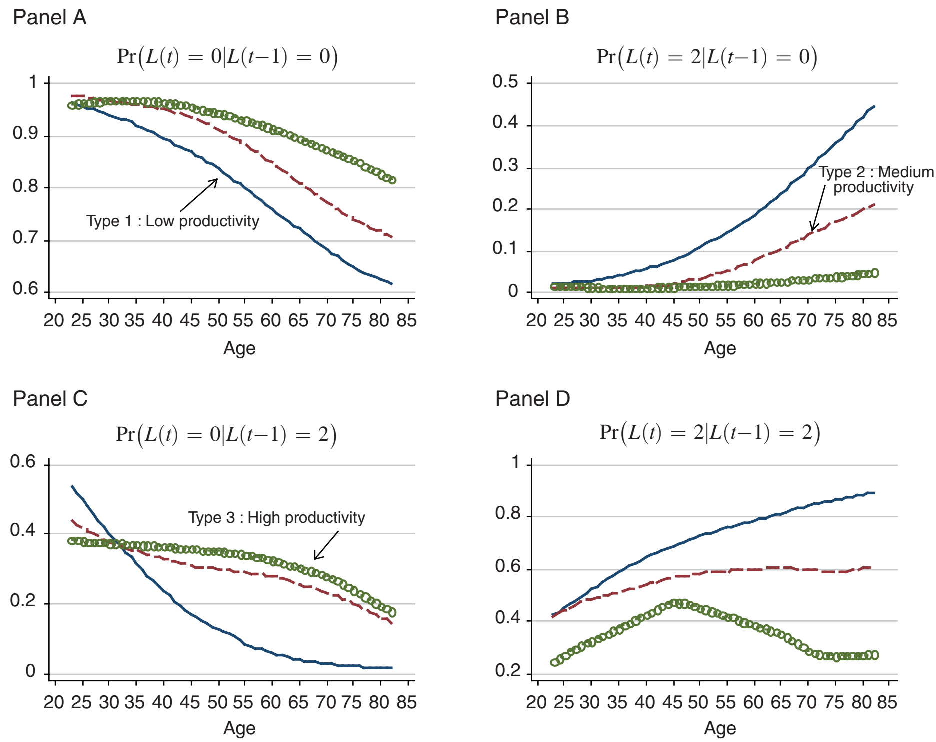

FIGURE 2. SELECTED TRANSITIONS

$\Pr\left(L_{it} = 2 \,|\, L_{it-1} = 2, f_q\right)$ (panel D) increases strongly with age, especially among the low-productivity type.[28]

D. *Identification of Preferences and Disability Insurance Parameters*

Identification of the remaining structural parameters of interest $(\eta, \theta, \delta, F_{L=0}, F_{L=1}, F_{L=2})$ and the DI policy parameters $(\pi_{L=0}^{Young}, \pi_{L=1}^{Young}, \pi_{L=2}^{Young}, \pi_{L=0}^{Old}, \pi_{L=1}^{Old}, \pi_{L=2}^{Old})$ is achieved by indirect inference (see Gourieroux, Monfort, and Renault 1993). Indirect inference relies on matching moments from an approximate model (known as *auxiliary model*) which can be estimated on both real and simulated data, rather than on moments from the correct data generating process. The moments of the auxiliary model are related (through a so-called binding function) to the structural parameters of interest. The latter are estimated by minimizing the distance between the moments of the auxiliary model estimated from the observed data and the moments of the auxiliary model estimated from the simulated data. Any bias in estimates of the auxiliary model on actual data will be mirrored by bias in estimates of the auxiliary model on simulated data, under the null that the structural model is correctly specified. However, the closer the link

[28] Low-educated individuals face worse health risk than high-educated individuals, with higher probabilities of bad shocks occurring and a lower probability of recovery. These differences across education, alongside the much greater prevalence of DI among the low educated, are the reasons why we focus our analysis on the subsample of individuals with low education.

between the moments of the auxiliary equations and the structural parameters, the more reliable is estimation.

The key question is how to choose which auxiliary moments to match. In our theoretical model, individuals make three decisions: how much to consume, whether to work, and whether to apply for DI. We also know that age is an important discriminant of admission into the program. Chen and van der Klauww (2008) show that the medical vocational grid used by the SSA in the assessment of applicants sets admission thresholds as a function of age. We hence choose auxiliary moments that reflect the choices individuals make and condition on their health status and age.[29] In particular, we use: (i) the composition of the stock of recipients of DI in terms of work limitation status and age; (ii) the fractions of people of different work limitation status and age who are on DI; (iii) the flows into the DI program by age and disability status; (iv) a regression of log consumption on work limitation, disability insurance, employment (and interactions), controlling for a number of other covariates; and (v) employment rates, conditional on disability status and age. These choices give us 30 moments overall.

Moments: Disability Insurance.—There are three ways in which we calculate moments involving DI recipients. First, we consider the composition of DI recipients by health status and age. This identifies the fraction of DI recipients who are not truly disabled and helps to pin down the incentive cost of the program. Second, we consider the DI status of individuals within work limitation-types. For the severely work limited, this identifies coverage: the fraction of the truly needy who benefit from DI. Third, we consider the flow rates onto DI by individuals within work limitation types.

These moments can be directly related to the probabilities of a successful application, the structural parameters of the DI program. Intuitively, a higher probability of success for a given type would generate higher flows into the program and larger stocks on DI for that type. For example, we use the fraction of those with a severe limitation and not on DI in $t-1$ who move onto DI in t to help identify $\pi_{L=2}^{Old}$. The fraction we observe, and use as auxiliary moment, is $Fr\left(DI_t = 1 \mid \Omega\right)$, where $\Omega = \left\{DI_{t-1} = 0, L = 2, Old\right\}$. We can show that

$$
\text{(6)}\quad Fr\left(DI_t = 1 \mid \Omega\right) = \Pr\left(DI_t = 1 \mid \Omega, DI^{App} = 1\right) \times \Pr\left(DI_t^{App} = 1 \mid \Omega\right)
$$

$$
= \pi_{L=2}^{Old} \Pr\left(DI_t^{App} = 1 \mid \Omega\right).
$$

The observed fraction would be particularly informative if all $L = 2$ individuals applied $\left(\text{i.e., if } \Pr\left(DI_t^{App} = 1 \mid \Omega\right) = 1\right)$. However, because not everyone applies, the moment we use (the left-hand side of (6)) is a lower bound on the probability of acceptance, the structural parameter of interest. To move from a bound on the

[29] We do not have data on DI application, and hence use moments reflecting participation in the DI program.

probability of acceptance to the actual probability requires a model of the application decision, which will itself be affected by the probabilities of acceptance, as well as the availability of other insurance programs and wage offers.

Consider the following example: the flow fraction $Fr(DI_t = 1|\Omega) = 0.28$ in the data. Suppose we start the iteration with $\pi^{Old}_{L=2} = 0.1$. This probability will not match the data regardless of what the application probability is. Since the probability of applying for DI is not greater than 1, it is clear that $\pi^{Old}_{L=2}$ must exceed 0.28 to make sense of the data, and this is indeed the area where the algorithm will search. For any value of $\pi^{Old}_{L=2}$, the structural model simulates a different $\Pr(DI^{App}_t = 1|\Omega)$, where at the margin more people apply as $\pi^{Old}_{L=2}$ increases. If the fraction (6) were the only moment to match, the algorithm would pick the $\pi^{Old}_{L=2}$ such that $\pi^{Old}_{L=2}\Pr(DI^{App}_t = 1|\Omega)$ is as close as possible to 0.28. In practice, the probabilities and application rates also affect the stock of DI recipients, which are more precisely measured, but which are affected by the flows off DI and by changes in health status over time. We use both flows and stocks by work limitation status as our auxiliary moments.

Moments: Consumption Regression.—A work limitation is likely to have two separate effects on consumption: first, the work limitation affects earnings and hence consumption through the budget constraint. The size of this effect will depend on the persistence of the shock and the extent of insurance, both self-insurance and formal insurance mechanisms such as DI. The second effect of the work limitation is through nonseparabilities in the utility function (measured by the parameter θ in (1)). For example, if being disabled increases the marginal utility of consumption (e.g., through increased needs) then consumption will rise on disability even if there is full insurance and marginal utility is smoothed over states of disability. It is important to separate out these two effects. Stephens (2001) calculates the effect of the onset of disability on consumption, but does not distinguish whether the effect is through nonseparability or through the income loss directly.

The identification of θ comes from a regression of consumption on work limitation. Of course, the presence of an effect on consumption through the budget constraint means this does not itself identify the nonseparability, θ. However, if we were able to identify a (control) group of individuals who are fully insured against disability shocks, then the consumption response to the work limitation for those individuals would capture only preference effects. Since no group is completely insured,[30] our method for separating out the two effects is to use the parameters of the following auxiliary regression:

$$\ln c_{it} = \alpha_0 + \alpha_1 L^1_{it} + \alpha_2 L^2_{it} + \alpha_3 L^1_{it} DI_{it} + \alpha_4 L^2_{it} DI + \alpha_5 DI_{it} + \alpha_6 P_{it} + \alpha_7 t + \alpha_8 t^2 + \upsilon_{it}.$$

[30] The extent of insurance from DI obviously depends on being admitted into the program, but conditional on receiving DI, the extent of insurance is greater for low-income individuals because replacement rates for our low-educated sample can be fairly high: (i) DI payments are progressive (the replacement rate is about 85 percent for people at the 25th percentile of the AIME distribution, and about 65 percent at the median); (ii) DI covers medical expenses through the Medicare program after two years on the program; (iii) unlike wages, benefits are untaxed up to a certain limit; (iv) lifetime replacement rates may potentially be higher because DI payments are received with certainty while employment is random due to labor market frictions.

The effect of a (severe) work limitation on consumption for individuals who are not in receipt of DI is given by the parameter α_2. This captures both the income effect and the nonseparability effect. For individuals who are in receipt of DI, however, the effect of a severe disability on consumption is $(\alpha_2 + \alpha_4)$. If DI provided full insurance, $(\alpha_2 + \alpha_4)$ would capture the effect of the nonseparability, with the parameter α_4 negating the income effect in α_2. The split between α_2 and α_4 is less clear if insurance is partial; in which case $(\alpha_2 + \alpha_4)$ captures both the nonseparable part and the lack of full insurance for those receiving *DI*. Indirect inference exploits this identification intuition without putting a structural interpretation directly on the α parameters. The coefficients α_1 and α_3 correspond to the effects of a moderate disability. We use an adult-equivalent measure of consumption and control for a quadratic in age to account for life-cycle evolution of family composition and tastes.[31,32]

Employment can also provide insurance against disability shocks. In addition, employment has a direct effect on the marginal utility of consumption (the parameter η). We use the auxiliary parameter α_6 to help capture this nonseparability between consumption and labor supply. Intuitively, whether consumption and employment covary positively or negatively (controlling for health status and point on the life cycle) is informative about whether they are Frisch complements or substitutes in utility.

Moments: Employment Rates over the Life Cycle.—We calculate employment rates by age and by work limitation status, using older (45 and above) and younger workers (less than 45). The moments that we use are the employment rates for the three work limitation groups in each age group, giving six moments overall. These moments are related to fixed cost of work with different work limitations, $F(L)$, the utility cost of working, η, and the labor market frictions.

In particular, unemployment rates among the healthy in the early life cycle are informative about the job destruction rate δ because assets are very low at young ages and so very few decide voluntarily not to work. The differences in employment by disability status is informative about the extent that work is more costly for disabled than for healthy workers and thus how the fixed cost of work differs by work limitation status.

E. *Indirect Inference Results*

In this section we present results on the moments matched by indirect inference in Tables 4 and 5, and the estimates of the structural parameters in Table 6. For each targeted moment, we present its value in the data, its simulated value (evaluated at the estimated structural parameters), and the 95 percent confidence interval for the

[31] Our measure of consumption is per adult equivalent (using the OECD equivalence scale $1 + 0.7(A - 1) + 0.5K$, where A is the number of adults and K the number of children in the household).

[32] We need to add two caveats to our identification strategy. First, as stressed by Meyer and Mok (2013), consumption is measured at the family level, but we observe changes in disability at the individual level. To partly account for this, we use a measure of adult equivalent consumption. The second caveat is that disability insurance is only one form of insurance against disability risk (SSI and workers' compensation being others).

TABLE 4—TARGETED MOMENTS: CONSUMPTION AND EMPLOYMENT

Variable/moment	Data	Simulations	95% C.I. difference
Panel A. The log consumption regression			
$\{L_{it} = 1\}$	−0.051 (0.035)	−0.091	(−0.03, 0.11)*
$\{L_{it} = 2\}$	−0.162 (0.070)	−0.172	(−0.13, 0.15)*
$\{L_{it} = 1\} \times DI_{it}$	0.177 (0.134)	0.154	(−0.23, 0.27)*
$\{L_{it} = 2\} \times DI_{it}$	0.260 (0.148)	0.374	(−0.40, 0.17)*
DI_{it}	−0.105 (0.123)	−0.336	(−0.01, 0.47)*
Employed	0.390 (0.055)	0.242	(0.04, 0.25)
Panel B. Employment by disability status			
Fr $\{P_{it} = 1 \mid L_{it} = 0,\ t \leq 45\}$	0.927 (0.0034)	0.917	(0.00, 0.017)
Fr $\{P_{it} = 1 \mid L_{it} = 0,\ t > 45\}$	0.868 (0.0074)	0.914	(−0.06, −0.03)
Fr $\{P_{it} = 1 \mid L_{it} = 1,\ t \leq 45\}$	0.701 (0.0217)	0.683	(−0.03, 0.06)*
Fr $\{P_{it} = 1 \mid L_{it} = 1,\ t > 45\}$	0.499 (0.0277)	0.516	(−0.07, 0.03)*
Fr $\{P_{it} = 1 \mid L_{it} = 2,\ t \leq 45\}$	0.161 (0.0185)	0.169	(−0.04, 0.03)*
Fr $\{P_{it} = 1 \mid L_{it} = 2,\ t > 45\}$	0.077 (0.0125)	0.087	(−0.03, 0.01)*

Note: The confidence interval is computed with the block bootstrap.
* Significant at the 5 percent level.

difference between data and simulations.[33] Table 4 presents consumption moments and employment moments, while Table 5 presents DI coverage moments, moments related to the composition of DI recipients, and DI flows moments.

Starting with Table 4, we find that our auxiliary model estimates of the consumption regression suggest that consumption falls when people become disabled and there is no insurance. However, those who are insured against the disability shock (those who are receiving DI) increase their consumption, consistent with the idea that consumption and poor health are Frisch complements ($\theta < 0$ in our utility specification). This may arise, for example, because a disability that induces a work limitation may also reduce an individual's opportunities for home production, such as in preparing food, housework, and in accessing the cheapest shops. These auxiliary regression results are very closely replicated by our simulated moments. The effect of employment on consumption is higher in the data than in the simulation, but qualitatively the effect is similar. None of the health and DI-related moments are statistically different in the data relative to the simulations.

[33] To obtain this confidence interval we compute standard errors of the auxiliary moments estimated in the data by the block bootstrap. Call $s_{\hat{\beta}_{data}}$ this standard error. The standard error of the difference $\left(\hat{\beta}_{data} - \hat{\beta}_{sims}\right)$ is computed (using asymptotic results) as: $\sqrt{\left(1 + \frac{1}{S}\right) s^2_{\hat{\beta}_{data}}}$, where S is the number of simulations ($S = 40$ in our case).

Table 5—Targeted Moments: Disability Stocks and Flows

Variable/moment	Data	Simulations	95% C.I. diff.
Panel A. DI coverage			
$\mathrm{Fr}\{DI_{it} = 1 \mid L_{it} = 2, t \leq 45\}$	0.308 (0.032)	0.298	(−0.05, 0.07)*
$\mathrm{Fr}\{DI_{it} = 1 \mid L_{it} = 2, t > 45\}$	0.552 (0.030)	0.544	(−0.05, 0.06)*
$\mathrm{Fr}\{DI_{it} = 1 \mid L_{it} = 1, t \leq 45\}$	0.081 (0.014)	0.091	(−0.04, 0.02)*
$\mathrm{Fr}\{DI_{it} = 1 \mid L_{it} = 1, t > 45\}$	0.187 (0.021)	0.182	(−0.04, 0.05)*
$\mathrm{Fr}\{DI_{it} = 1 \mid L_{it} = 0, t \leq 45\}$	0.003 (0.001)	0.003	(−0.00, 0.00)*
$\mathrm{Fr}\{DI_{it} = 1 \mid L_{it} = 0, t > 45\}$	0.016 (0.003)	0.014	(−0.00, 0.01)*
Panel B. Composition of DI recipients			
$\mathrm{Fr}\{L_{it} = 2 \mid DI_{it} = 1, t \leq 45\}$	0.638 (0.0390)	0.605	(−0.05, 0.12)*
$\mathrm{Fr}\{L_{it} = 2 \mid DI_{it} = 1, t > 45\}$	0.678 (0.0224)	0.691	(−0.06, 0.03)*
$\mathrm{Fr}\{L_{it} = 1 \mid DI_{it} = 1, t \leq 45\}$	0.243 (0.0314)	0.266	(−0.09, 0.04)*
$\mathrm{Fr}\{L_{it} = 1 \mid DI_{it} = 1, t > 45\}$	0.209 (0.0189)	0.220	(−0.05, 0.03)*
$\mathrm{Fr}\{L_{it} = 0 \mid DI_{it} = 1, t \leq 45\}$	0.120 (0.0228)	0.128	(−0.06, 0.04)*
$\mathrm{Fr}\{L_{it} = 0 \mid DI_{it} = 1, t > 45\}$	0.113 (0.0147)	0.090	(−0.01, 0.05)*
Panel C. Flows into DI			
$\mathrm{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 2, t \leq 45\}$	0.168 (0.022)	0.158	(−0.03, 0.05)*
$\mathrm{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 2, t > 45\}$	0.279 (0.024)	0.283	(−0.05, 0.04)*
$\mathrm{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 1, t \leq 45\}$	0.039 (0.008)	0.030	(−0.01, 0.02)*
$\mathrm{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 1, t > 45\}$	0.067 (0.011)	0.043	(0.00, 0.05)
$\mathrm{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 0, t \leq 45\}$	0.001 (0.0003)	0.0005	(−0.00, 0.00)*
$\mathrm{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 0, t > 45\}$	0.007 (0.001)	0.002	(0.00, 0.01)

Note: The confidence interval is computed with the block bootstrap.
* Significant at the 5 percent level.

Turning to panel B of Table 4, the model is capable of matching well the employment behavior of people with severe and moderate disabilities, but it tends to overpredict employment of older nondisabled workers. Nevertheless, the differences appear economically small. These discrepancies arise for the group of healthy individuals that is least affected by the type of policy experiments we consider.

In Table 5 we present disability moments. In panels A and B we look at the two sides of the insurance/disincentives trade-off of DI. Our model is capable of matching all the moments with great accuracy. For example, it matches closely the

TABLE 6—ESTIMATED PARAMETERS

Frictions and preferences		Disability insurance program	
Parameter	Estimate	Parameter	Estimate
θ	−0.448*** (0.126)	$\pi_{L=0}^{Young}$	0.006 (0.964)
η	−0.185 (0.160)	$\pi_{L=0}^{Old}$	0.075 (0.800)
δ	0.062*** (0.002)	$\pi_{L=1}^{Young}$	0.171*** (0.025)
		$\pi_{L=1}^{Old}$	0.180*** (0.032)
$F_{L=0}$	0.000 [$0] (0.371)	$\pi_{L=2}^{Young}$	0.331*** (0.031)
$F_{L=1}$	0.547*** [$2,472] (0.111)	$\pi_{L=2}^{Old}$	0.626*** (0.046)
$F_{L=2}$	0.952*** [$4,301] (0.109)		

Notes: Fixed costs are reported as the fraction of average offered wage income at age 23 and also in 1992 dollars per quarter. Standard errors in parentheses (see the online Appendix for definitions).
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

proportions of "false recipients," $\text{Fr}(L = 0|DI = 1, t)$, as well as the proportion of disabled individuals "insured" by the DI program, $\text{Fr}(DI = 1|L = 2, t)$, which are the reduced form equivalents of the incentive cost/insurance benefit trade-off. In panel C we examine the flows into the program by work limitation and age. Once more, the model fits these moments well, and the statistical rejections are not economically significant.[34]

In Table 6 we report the indirect inference parameter estimates corresponding to these moments. We estimate that a moderate (severe) work limitation corresponds to a 36 percent (59 percent) loss of utility in terms of consumption. Working corresponds to a 17 percent loss of consumption, but the estimate of η is imprecise. The fixed costs of work per quarter rise substantially with the degree of work limitation. We estimate that a job is destroyed on average every 16 quarters. The probability of success of DI application increases with disability status, and it increases markedly by age for the severely disabled. There is no evidence of age increases among the non-severely disabled suggesting the efficiency of the program is greater among the old. The estimates of the success probabilities by type (age and work limitation status) provide information on the extent of type I and type II errors, which we discuss further in the next section. All estimates are statistically significant except for the probabilities of success among those without any work limitation and the fixed cost of work for the nondisabled, which are however economically insignificant.

[34] For the moments in Table 4, panel B we use reported employment status and self-reported work limitation status at the time of the interview (and hence use all waves). For the moments in Table 5, panels A and B we use DI recipiency status reported in wave t (referring to calendar year $t-1$) and self-reported work limitation status at the time of the interview in wave t (and so again use all available PSID waves). Finally, for panel C, we use two-year flows from all waves (DI recipiency reported in waves t and $t+2$, and work limitation status reported at the time of the interview in wave $t+1$). Note that the moments computed in the simulations replicate exactly these timings.

F. *Implications and External Fit of the Model*

In this section we discuss the implications of our estimates for the success of the DI screening process, for behavioral responses to DI program parameters, and the extent of self-insurance. We also show the importance of our estimates about the role of work limitations. We compare predictions of our model with evidence from the predominantly reduced form literature. This is a way to verify that the model can reproduce statistics about the DI program that were not explicitly targeted by the estimation procedure (external validity).

Success of the DI Screening Process.—One important issue is to evaluate the success rate of the existing DI screening process. We first look at the award rate at the point of entry in the system (i.e., award of initial application).[35] We simulate this rate (using our structural model and estimated parameters) to be 0.42. French and Song (2014) use administrative SSA data on the outcome of DI applications and report a very similar success rate for the initial application (0.39). In practice, applicants who are rejected can appeal at four different successive levels: DDS reconsideration, administrative law judges (ALJ), federal court, and at the council review level. While we do not model the appeal process formally, we do allow individuals to reapply for DI following rejections. This allows us to compare award rates in the short and long run in the model and in reality. According to French and Song (2014) the award rate after 2 years from the initial application is 0.53 (0.52 in our model); and after 10 years is 0.67 (0.73 in our model). Hence, our model captures quite well short- and long-run award rates.

These award rates do not condition on the applicant's health. Given that the true disability status of an applicant is private information, SSA evaluators are likely to commit two types of errors: admitting onto the DI program undeserving applicants and rejecting those who are truly disabled. Our structural estimates of the success rates show how large these errors are. Consider first the extent of false positives (the proportion of healthy applicants who are awarded DI). From Table 6, these type II errors have probabilities ranging from 0.6 percent (for the nondisabled) to 18 percent (for those with only a moderate disability). Similarly, we can use our model to estimate the award error: the fraction of successful applicants to DI who are not severely disabled, given by $\Pr(L = \{0,1\} \,|\, DI = 1, DI^{App} = 1) = 0.12$. In the literature, one finds reduced form estimates that are slightly larger, 0.16 to 0.22 in Benítez-Silva et al. (1999), depending on the statistical assumptions made, and 0.19 in Nagi (1969). In our simulations, 70 percent of applicants are severely disabled. Those who are healthy and yet are on DI are predominantly those who have recovered while on DI but not left the program.

Consider next the probability of false negatives (i.e., the proportion of the severely work limited who apply and do not receive DI). From Table 6, our estimate is that the type I errors are 67 percent for the younger and 37 percent for the older workers. The fraction of rejected applicants who are severely work limited, the rejection error, is

[35] We restrict the sample to be between age 35 and 60 for consistency with French and Song (2014).

given by $\Pr(L = 2 | DI = 0, DI^{App} = 1) = 0.57$. This is similar to Benítez-Silva et al. (1999), who report 0.52–0.58, and Nagi (1969), 0.48.

These comparisons confirm that our structural model is capable of replicating reduced form estimates obtained using direct information on the application and award process. Taken together, these estimates suggest substantial inefficiencies in providing coverage for the severely work limited especially among the under 45s, but less inefficiencies in terms of identifying false claimants.[36]

Elasticities.—The reduced form literature on DI has analyzed the incentive cost of DI by looking at a number of behavioral responses, in particular the response of DI application and labor force participation (or employment) to an increase in generosity of the DI program. In Tables 7 and 8 we report elasticity estimates from representative papers in the literature (surveyed in the authoritative surveys of Bound and Burkhauser 1999, and Haveman and Wolfe 2000) and we compare these estimates with those that we can compute in our model. These are obtained by simulating individual responses as we change marginally the generosity of the DI program.

In Table 7 we report the elasticity of DI applications with respect to benefit generosity. As surveyed by Bound and Burkhauser (1999), empirical analyses using aggregate time series data from the 1960s and 1970s (such as Halpern 1979; Lando, Coate, and Kraus 1979) in general tend to find smaller elasticities (around 0.5) than those obtained from cross-sectional data (such as Kreider 1998 and Halpern and Hausman 1986), which however display more variability. A central estimate from Table 13 of Bound and Burkhauser (1999) is about 0.6 (with a 0.2–1.3 range). We vary the generosity of DI in our simulations and calculate the effect on the number of applications made. Our estimate of the elasticity (using all individuals) is 0.62. However, this figure masks considerable heterogeneity by health and productivity type. We break down the change in behavior into changes in behavior when moderately disabled and changes in behavior when severely disabled. The moderately disabled are very elastic in their response to generosity, whereas the severely disabled have very little response. Further, the response of applications to changes in generosity (conditioning on health) varies by productivity type, but while it decreases with productivity among the moderately disabled, it increases with productivity among the severely disabled. This reflects the fact that the "marginal" individual (as far as applying for DI is concerned) is the low-productivity type among the $L = 1$ individuals (as high-productivity types still face a high opportunity cost of applying), and the high-productivity type among the $L = 2$ individuals (as the low-productivity types apply regardless of benefit generosity). As we shall see, this difference across groups plays an important role when assessing the welfare implications of changing DI benefits generosity.

Table 8 shows the elasticity of the nonemployment rate with respect to benefit generosity. In the literature, the response of nonemployment to benefits is generally estimated to be smaller than the response of DI applications. For example, the range of estimates reported by Bound and Burkhauser (1999) in their Table 16 and by

[36] One caveat to this conclusion is the possibility of nonclassical measurement error. This might arise for example if people tend to exaggerate their report of work limitations if in receipt of DI or unemployed. If that was the case, our estimates of type I error will be overestimated and our estimates of type II error underestimated.

Table 7—Reduced Form Elasticities for Application onto DI

Range of estimates from literature	0.2 – 1.3			
	Whole sample	$f_q = f_L$	$f_q = f_M$	$f_q = f_H$
Our model				
All individuals	0.62	0.88	0.48	0.28
Moderately disabled $L = 1$	2.22	2.99	1.74	0.72
Severely disabled $L = 2$	0.018	−0.082	0.072	0.24

Notes: The range of estimates from the literature comes from Bound and Burkhauser (1999, Table 13). The pool of possible applicants is those individuals not on disability insurance already. This pool shrinks as generosity increases.

Table 8—Reduced Form Elasticities of Nonemployment

Range of estimates from literature	0.06 – 0.93			
	Whole sample	$f_q = f_L$	$f_q = f_M$	$f_q = f_H$
Our model				
All individuals	0.056	0.023	0.084	0.053
Moderately disabled $L = 1$	0.20	0.079	0.356	0.288
Severely disabled $L = 2$	0.023	0.002	0.042	0.146

Note: The range of estimates from the literature comes from Bound and Burkhauser (1999, Table 16) and Haveman and Wolfe (2000, Table 10).

Haveman and Wolfe (2000) in their Table 10 is between 0.06 and 0.93. In our model this elasticity is right at the bottom of this range of estimates (our estimate is 0.06). We also break our sample by work limitation and find a differential effect on the moderately and severely disabled: the moderately disabled are more sensitive but neither group has a large response. The response of nonemployment to generosity is smaller for the low types than the others. This is because there is substantial nonemployment among the low types regardless of whether on disability insurance or not.

Flows off DI.—We use our model to simulate the rate of flows off the DI program by work limitation status, and we compare these to rates in the data. We did not use these rates in the estimation because these moments are imprecisely estimated given the size of our sample. However, we reproduce in Table 9 the main annual flow statistics and the simulated counterparts as an indication of the performance of the model. Simulated flows off DI match the decline by age observed in the data. The difference between actual and simulated outflows is statistically insignificant.

Life-Cycle Profiles and Asset Accumulation.—We consider two aspects of life-cycle profiles that we do not target directly but which are key checks on the validity of the model. First, we consider consumption and earnings over the life cycle, then we consider asset profiles.

It is well-known that in the data, consumption has a hump shape over the life cycle. The left-hand side of Figure 3 shows the match for spending between data and simulations. Similarly, the right-hand side shows the match for earnings between data and simulations, where earnings reflects the wage process and the labor supply

Table 9—Flows off DI

Moment	Data	Simulated	95 percent CI difference
$\mathrm{Fr}(DI_t = 0 \mid DI_{t-1} = 1, t < 45)$	0.119 (0.031)	0.110	(−0.063, 0.081)*
$\mathrm{Fr}(DI_t = 0 \mid DI_{t-1} = 1, t \geq 45)$	0.088 (0.016)	0.074	(−0.014, 0.042)*

Note: Block bootstrap standard errors in parenthesis.
* Significant at the 5 percent level.

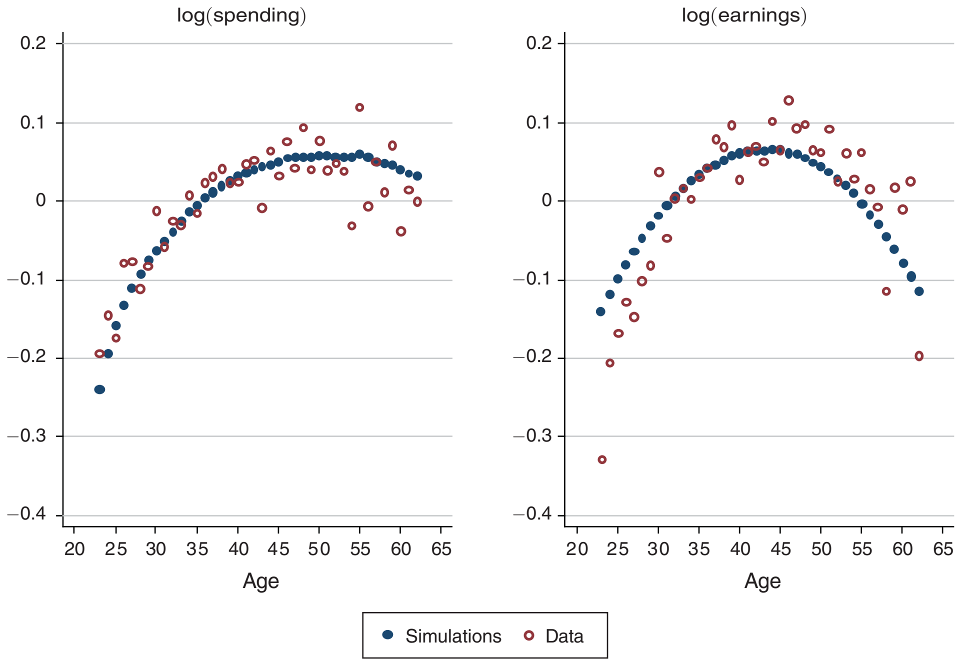

Figure 3. log Spending and log Earnings over the Life Cycle

decision. The figure shows that both simulated profiles match the data closely.[37] This is a minimum requirement of any life-cycle model and our model passes this test: we did not target the life-cycle fit of the model explicitly.

An important part of the model is individuals' ability to self-insure through asset accumulation. Unfortunately, it is difficult to compare asset accumulation when there is only one liquid asset available (as in the model), with data where individuals have both liquid financial wealth (bank deposits and stocks) as well as illiquid assets (housing and pension wealth). It is precisely for this reason that we use data on consumption and income, rather than assets, in our estimation. Moreover, during

[37] These pictures are obtained as deviation from life-cycle means; we drop the bottom 2.5 percent and top 2.5 percent of the relevant data distributions to reduce the influence of outliers.

the sample period covered by our data, asset information was asked only irregularly (in 1989, 1994, and every wave since 1999). Nonetheless we can compare simulated life-cycle asset profiles by health with those we obtain in the data as a form of external validation.

Our definition of assets in the data includes both housing wealth and liquid financial wealth.[38] Median asset holdings around retirement (in the five year interval centered at 60) in the simulations are close to the data. In the model, those without work limitations have accumulated median assets that are 1.93 times those of the severely disabled; those with moderate limitations have median assets that are 1.33 times the severely disabled. In the data, these numbers are 1.78 and 1.39, respectively. Another useful statistic is wealth dynamics by health status. In online Appendix Figure 11, we plot the evolution of average wealth over the life cycle for the three health groups. We find that the model approximates well the asset profiles we observe in the data. For the nondisabled, the growth over the life cycle is more rapid at young ages and less rapid in middle age than in the data. For those with some limitations there is a closer fit.

Sensitivity: The Importance of Health.—In our structural model, health status affects behavior in two ways: it shifts preferences (nonseparability) and it changes the fixed cost of work.[39] We consider here whether both of these mechanisms are necessary. First, we consider the case where the fixed cost of work does not vary with health status. Second, we consider switching off the nonseparability between consumption and health. In both cases, we reestimate the structural model to match the same set of moments as in the baseline.

When the fixed cost of work does not vary with health, the structural estimates of the model imply large and numerous deviations between data and simulated moments (or auxiliary parameters). The very poor fit of the model is because without heterogeneity in the fixed cost of work by health status it is very difficult to generate differences in employment across disability groups: too many of the disabled remain at work compared to the data. The bad fit for the employment numbers cascades onto the number of DI applicants and this in turn affects the DI moments and so forth.

When we assume separability between consumption and work limitations, or $\theta = 0$, we also obtain a worse fit relative to the baseline, but the model does fairly well in most dimensions (there are 10 rejections out of 30 moments, compared to just 5 rejections in the baseline). Estimation details are in the online Appendix in Tables 9 and 10. The poor fit in this case is coming from the consumption equation. The coefficients on the work limitation variables $L = 1$ and $L = 2$ are much more negative: statistically we reject the null that data and simulations produce similar estimates of the auxiliary parameters of the consumption regression. This is expected: our estimate of θ implies that the marginal utility of consumption is higher when disabled and so resources are moved into periods in which people suffer a work limitation shock to keep marginal utility of consumption smoothed. When θ

[38] We calculate median asset holdings at different ages and for different work limitation status.

[39] There is also the effect of health on wages and the effect of health on the variance of productivity shocks which are estimated directly.

is set to 0 and the nonseparability is removed, there is a larger negative effect on consumption because there is only an income effect with no offsetting substitution.

V. Reform of the DI Process

The most important use of our model and structural estimates is the ability to analyze the effects on welfare and behavior of changing the main parameters of the DI program. We consider four main changes: (i) changing the generosity of disability payments; (ii) making the program "stricter" by increasing the threshold that needs to be met in order to qualify for benefits; (iii) changing the generosity of the means-tested (food stamp-type) program; and (iv) changing the reassessment rate of disability recipients. For each scenario, we show the implications for the coverage of the severely disabled, the extent of false applications by the nondisabled, welfare, aggregate output, and asset accumulation. We calculate the welfare implications by measuring the willingness to pay for the new policy through a proportional reduction in consumption at all ages which makes the individual indifferent ex ante between the status quo and the policy change considered.[40] In all the experiments below the impact on the government budget is neutralized by adjusting the proportional wage tax iteratively (see equation 4 in the online Appendix). We also examine the sensitivity of our policy experiment conclusions to changes in the value of risk aversion, one of the key exogenous parameters. We stress that we cannot draw conclusions about optimal policy from these experiments. Our policy experiments are best seen as showing partial effects of reform because, although reform is revenue neutral, we do not take account of general equilibrium effects, nor do we consider introducing multiple reforms simultaneously.

A. *Generosity of DI Payments*

In the first experiment, we consider the effects of revenue-neutral, proportional changes in DI generosity, with the proportional changes ranging from a cut to 60 percent of its current value to a 40 percent increase.[41] Figure 4 shows the effects of these changes. The left-hand side of Figure 4 shows the effects of the policy on the fraction of applications that are from $L = 0$ or $L = 1$ individuals (the solid line labeled "False applications") and on the fraction of severely disabled who are receiving insurance (the dashed lines labeled "$L = 2$ insured," plotted separately for older and younger workers). Both false applications and coverage of the severely disabled increase as generosity increases. However, the fraction of false applicants is much more responsive to changes in generosity than the coverage of the severely disabled (as also evident from the first column of Table 7).[42]

[40] This is obtained by calculating expected utility at the start of the life cycle before the resolution of any uncertainty ("behind the veil of ignorance").

[41] This proportional increase in generosity also applies to benefits from SSI. If the generosity of SSI is unchanged, then the left-hand side of these graphs is different: when DI is very ungenerous, then SSI will replace it and so there is not the same decline in applicants for disability support.

[42] The fraction of the severely disabled aged under 45 receiving insurance is at a lower level, but similarly varies with generosity much less than the number of false applicants.

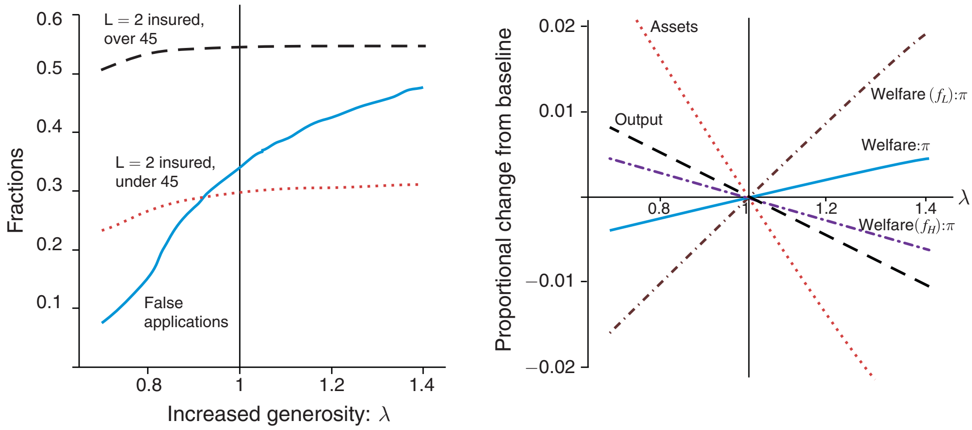

Figure 4. Changing DI Generosity

The right-hand graph shows the effects on welfare, output, and assets. The solid line shows the overall effect on welfare (weighting the expected utility at time 0 by the sample size of each skill type). Despite the rise in false applications with generosity, welfare increases with increased generosity.[43] The cost of increasing generosity is the extra tax that has to be raised: there is no direct welfare cost from "cheating" in this framework. The key point is that the greater insurance value of more generous payments dominates the cost of the revenue needed to pay the false claimants. We split this welfare effect by skill type: for the lowest skill $(f_q = f_L)$, welfare is sharply increasing with generosity because the disability insurance is partly redistributive toward the low types. The corollary of this is that welfare is decreasing in generosity for the high types $(f_q = f_H)$, who have to pay more toward disability insurance than the actuarially fair amount. The greater false applications are associated with lower labor force participation and so lower output. Output falls despite welfare increasing partly because of the utility value associated with increased leisure and partly because there is better insurance associated with increased generosity. The assets line shows the effect of generosity on the maximum assets held over the lifetime. The fall in assets with generosity partly reflects the fall in output reducing saving for consumption smoothing and, to the extent that assets are more sensitive than output, the additional crowding out of self-insurance.

B. *Strictness of DI Admissions*

Increases in the strictness of the screening process for DI implemented in 1980 led to sharp declines in inflows onto DI and significant removal of DI recipients, although the criteria were relaxed again in 1984. The issue is whether the benefit induced by greater strictness in terms of reduced incentives for false applications outweighs the worsening insurance of truly disabled workers. To tackle this

[43] Meyer and Mok (2013) reach a similar conclusion. They apply a variant of the benefit optimality formula derived by Chetty (2008) to conclude that the current level of DI benefits is lower than the optimal level (i.e., that it is welfare improving to increase DI generosity).

issue, we need first to define a measure of strictness of the program. As discussed in Section IID, DI evaluators decide whether to award DI as a function of a noisy signal about the severity of the applicant's disability status, which has some distribution g:

$$S_{it} \sim g(L,t).$$

Our estimates of the success probabilities imply that the properties of the distribution of the signal S vary by age and by work limitation status L. Assume that the Social Security DI evaluators make an award if $S_{it} > \overline{S}$. The parameter $\overline{S}$ can be interpreted as a measure of the strictness of the DI program: other things equal, an increase in $\overline{S}$ reduces the proportion of people admitted into the program. We assume that S lies between 0 and 1 and has a beta distribution, $Beta(q_{L,t}, r_{L,t})$, whose parameters q and r vary with age and work limitation status. The values of $q_{L,t}$ and $r_{L,t}$ and of $\overline{S}$ are pinned down by the six structural probabilities (π_L^t) estimated above:[44]

$$1 - \pi_L^t = \Pr(\text{Rejection} \mid t, L, Apply) = CDF\big(Beta(q_{L,t}, r_{L,t})\big).$$

Figure 5 illustrates the resulting distributions of S for those over 45 by work limitation status, and illustrates some of the errors under the estimated DI program. The area on the left of $\overline{S}$ under the light gray curve (labeled $f(S|L = 2, t \geq 45)$ measures the probability of rejecting a deserving DI applicant. The area on the right of $\overline{S}$ under the dark gray curve (labeled $f(S|L = 1, t \geq 45)$) measures the probability of accepting into the DI program a DI applicant with only a moderate disability. Increasing the strictness of the test (increasing $\overline{S}$) reduces the probability of false positives (reduces the extent of the incentive problem), but increases the probability of false negatives (reduces the extent of insurance provided by the program). It also can have substantial effects on who applies. A policy of changing $\overline{S}$ therefore has both benefits and costs, trading off incentives against insurance, and we use our model to determine which dominates when the strictness of the test changes.[45]

Figure 6 reports the results of changing the level of strictness as measured by $\overline{S}$. The left-hand graph shows the implications for the DI program in terms of the coverage/disincentive trade-off, while the right-hand graph shows implications for welfare, output, and asset accumulation. Increasing $\overline{S}$ from 0.6 to 0.8 reduces the probability of acceptance for the severely disabled over 45 (under 45) from 90 percent to 30 percent (70 percent to 13 percent, respectively). Furthermore, the increase

[44] We impose that the parameter r is identical across age and work limitation status. We need to impose two further normalizations, and choose to normalize the mean of the signal for the severely disabled old and that of healthy young workers (those with the highest and lowest probability of success in the data). These normalizations, alongside the use of the beta distribution, impose a particular distribution on the signals which we do not have the data to test. The intuitive advantage of the beta distribution is that the precision of the signal increases as true disability status worsens. We considered alternative assumptions, such as a lognormal distribution and find very similar results. See the online Appendix for a discussion of the beta distribution and the results using a lognormal distribution.

[45] An alternative policy might be to reduce the noise involved in the evaluation of the signal. We do not evaluate such a policy. In theory, we could take the cost of extra SSA evaluations as being the same as the cost of a review. However, the difficulty is estimating the effect of evaluations on reducing the noise.

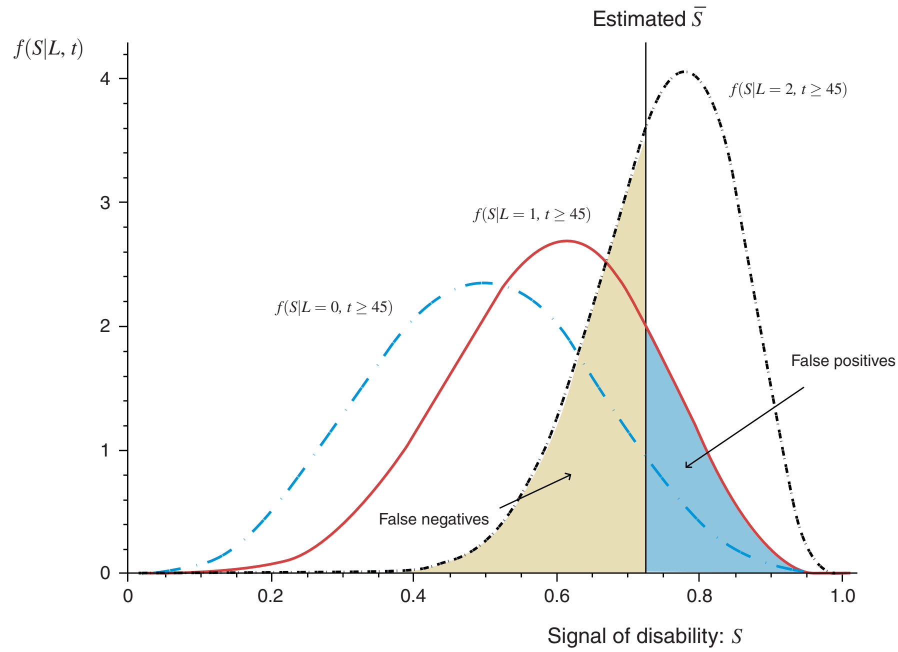

FIGURE 5. THE DISTRIBUTION OF S FOR THE OLDER WORKERS BY WORK LIMITATION STATUS

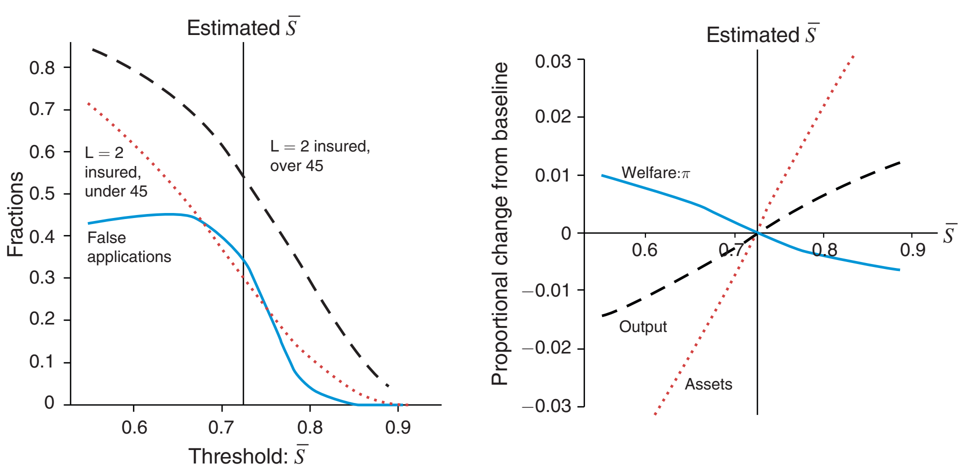

FIGURE 6. INCREASING STRICTNESS OF THE SCREENING PROCESS

in $\bar{S}$ reduces the proportion of applicants from those with no or only a moderate disability. This is shown in the downward sloping line labeled "False applications."[46]

The right-hand graph shows the willingness to pay for the alternative $\bar{S}$ in expected utility terms (the welfare measure π). The willingness to pay increases

[46] Corresponding to this fall in healthy applicants and lower rate of acceptance, there is a clear decline in the fraction of awards being made to the healthy or moderately disabled (the award error, not reported).

as $\overline{S}$ decreases from its estimated value: the gain in improved insurance from making the program less strict dominates the loss associated with increased numbers of false applicants and a greater award error. The magnitude of the gain in terms of consumption equivalent arising from reducing strictness from its estimated value ($\overline{S} = 0.72$) to $\overline{S} = 0.6$ is about 0.01 (1 percent). This net gain is the result of two offsetting effects: there is a benefit of increased insurance against disability which individuals are willing to pay for, but this is partly offset by a loss arising from output being lower as individuals work less. Part of the benefit of the relaxed strictness arises from the moderately disabled and the severely disabled young being offered better insurance.

The key to this conclusion of reduced strictness being welfare increasing is, however, the low acceptance rate of young, severely disabled individuals onto DI in the baseline (see Table 6). The subgroup of young, severely disabled individuals are particularly ill-equipped to insure against disability risk because these individuals face high rejection rates when applying for DI and yet have not had time to accumulate enough assets to self-insure. Hence reduced strictness that increases the chance to get into the program is highly valued.[47]

French and Song (2014) and Maestas, Mullen, and Strand (2013) consider the extent of labor force participation by DI applicants who have been denied benefits because their application was dealt with by "tougher" disability examiners. We can interpret this empirical strategy as similar to the effect of changing the strictness of the regime in our experiment, as shown in Figure 6. A higher level of strictness in our model leads to a reduction in DI receipt, and a corresponding higher level of labor force participation. For the severely disabled, among those who do not receive DI because of greater strictness, we calculate that approximately 5 percent will be working. Among the moderately disabled, the percentage who will be working is significantly higher at about 30 percent, and this percentage is higher for the young. This range is similar to the range found by Maestas, Mullen, and Strand (2013).[48]

C. *Generosity of the Food Stamp Program*

The DI program may interact in important ways with government welfare programs, such as food stamps. Here we investigate how important such interactions might be. Figure 7 shows the effects of changing the generosity of food stamps (from a 40 percent reduction to a 40 percent increase relative to the status quo). For false applicants, food stamps are substitutes for disability insurance and generally application to DI falls as food stamps' generosity increases. This is because at some point food stamps provide such a sufficiently generous support (without the uncertainty and inconvenience of application for DI) that false applications for DI fall. Those with only a moderate work limitation use the increasingly generous means-tested program. By contrast, for severely disabled workers food stamps are complementary to DI: the fraction of the severely disabled who receive DI increases as food stamps

[47] Denk and Michau (2013) obtain a similar result using a dynamic mechanism design approach to the insurance-incentive trade-off.

[48] Some caution is needed in making this comparison: the fraction in the model is calculated by comparing two steady-states, whereas in Maestas, Mullen, and Strand (2013) the fraction is calculated using randomization due to the allocation of lenient assessors.

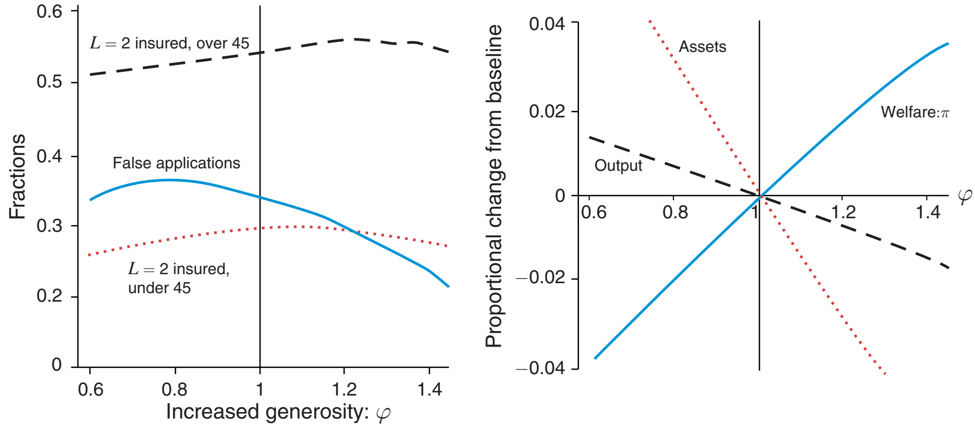

Figure 7. Changing the Generosity of Food Stamps

become more generous. This is because the consumption floor increases, making application for DI less costly for the severely disabled who were marginal between working and applying for DI. The effect of increasing food stamps is nonmonotonic, especially for the younger severely disabled who face high rejection rates from the DI program and food stamps becomes a good substitute at high levels of generosity. Together, these effects imply substantial welfare increases as the generosity of food stamps increases. A 10 percent increase in generosity implies a welfare gain of 0.9 percent of consumption. This increase is despite the fall in output and savings that greater generosity induces. It is important to stress that this movement onto food stamps is funded by a change in the tax rate and so, although the saving on DI may appear a false saving because of the greater spending on the food stamp program, our calculations are that this is welfare increasing despite the tax rise required. What this simulation highlights is the value of food stamps in providing long term support for those whose productivity is too low to be able to work for a reasonable wage. Part of the reason for this result is that the food stamps program is less distortionary than DI because it does not require people to disengage from the labor force and to stop working altogether.

D. *Reassessment Rates*

As a final policy change, we consider changing the reassessment rate. This is a policy that instead of affecting the nature of the screening process at the point of entry in the DI program, tries to affect exit rates from the program (which are notoriously quite low). Given our estimate of the cost per reassessment,[49] this has a direct impact on the budget, as well as the effect induced by changes in the number of recipients and in labor supply. We assume that the probabilities of success, conditional on

[49] For the period 2004–2008, the SSA spent \$3.985 billion to conduct 8.513 million "continuing disability reviews." This means a review costs on average \$468, and we deflate this back to 1992 prices and include this cost in the government's budget constraint.

work limitation status and age, are the same at reassessment as at initial application. The details are in the online Appendix, but are briefly summarized here, because the effects are not substantial. A proportional increase in the reassessment rate across all disability groups discourages false applications by those who are not severely disabled, but also reduces coverage for the severely disabled: reassessment causes some severely disabled to be removed from DI and this directly reduces coverage, as well as discouraging applications, as the frequency of reassessment increases. The reduced false applications lead to greater labor force participation and output, and increased asset accumulation as individuals have to self-insure further. The increase in reassessment rates has a small negative effect on welfare.

E. *Sensitivity to Risk Aversion*

The welfare and behavioral conclusions on policy experiments may be affected by the degree of risk aversion, which we take from previous literature rather than estimating it. In this section, we consider how differences in risk aversion affect the policy conclusions. We set the coefficient of relative risk aversion γ to equal 3 (compared to the baseline where $\gamma = 1.5$), and re-estimate the structural parameters of the model (i.e., those reported in Table 6). Details of the moments are in Table 12 in the online Appendix and details of the parameter estimates in Table 13. We find that the fit of our model is somewhat worse than in our baseline, but that we can still match the moments fairly well. The structural parameter estimates are somewhat different. First, the probability of success is higher and the cost of work is higher when γ is higher and individuals are more risk averse. This higher probability and higher fixed cost is necessary to induce risk averse agents not to work and instead to apply, which is needed to match the DI and participation moments in the data.

We use these new estimates of the structural parameters to redo our three main counterfactual policy experiments, varying generosity, strictness, and food stamps. As the generosity of the program increases, the fraction of the truly disabled who receive DI increases and the fraction of false applicants also increases, much as in the left-hand side of Figure 4.[50] Similarly, this translates into lower output and lower asset accumulation. The welfare consequences of the increased generosity are strengthened: more generous DI increases welfare by more when individuals are more risk averse because the value of the insurance goes up much more. The effects of changing strictness are qualitatively similar in all dimensions when risk aversion is higher: coverage and false applications both fall as strictness increases; similarly assets and output increase as individuals work harder and save more in response to the tougher policy. However, the magnitudes are different. In particular, the welfare benefit of reducing strictness is substantially greater than in the baseline: the insurance value of reducing the uncertainty about success for the severely work limited is much greater. When food stamps become more generous, the fraction of the truly disabled goes up as in Figure 7, and as in our baseline estimates, the number of false applications is not monotonic: false applications rise with food stamp generosity at low levels of generosity as this provides a subsidy to applications, but

[50] Figures 4, 6, and 7 are reproduced in the online Appendix for the case $\gamma = 3$.

false applications decline when food stamp generosity becomes large enough and there is less need to go onto DI. Output falls and asset holdings fall as generosity increases in a similar way to the baseline. Further, welfare increases as in the baseline, but much more markedly: the higher risk aversion makes individuals value the insurance provided by food stamps more highly.

VI. Conclusions

In this paper, we provide a life-cycle framework for estimating the extent of work-limiting health risk that individuals face and for analyzing the effectiveness of government disability insurance against that risk. Work limitations have substantial effects on wages, with wages falling by 18 percent for the severely work limited. Government insurance against these shocks is incomplete: There are substantial false rejections for those in need. We estimate that 37 percent of older workers with a severe work limitation who apply for benefits are rejected on their first application, with an even greater rejection rate for younger, severely work limited individuals. This is alongside other negative effects, with some workers discouraged from applying because of the uncertainty surrounding the application process. The limits on coverage implied by these estimates are more costly than the rates of false acceptances, with an estimate of the acceptance rate of 17 percent of applications from those who have a moderate work limitation.

We use the model to simulate various policy changes aimed at improving the insurance coverage and mitigating the incentive costs of DI. The simulations show that the number of moderately disabled individuals receiving DI is particularly sensitive to the policy parameters, whereas the number of severely disabled is less sensitive. Thus, reducing DI generosity leads to a fall off in false applications and misdirected insurance, without reducing applications from the severely disabled who are essentially inelastic with respect to benefit generosity. On the other hand, the severely disabled receive less insurance with the reduced generosity, and so the reduced generosity decreases welfare. This conclusion on welfare is at an aggregate level, which comprises both insurance and redistribution between types: the lowest productivity types benefit both from the insurance and also from redistribution from the high types.

Increasing the strictness of the DI screening process leads to a decline in welfare because the existing program already suffers from turning down large numbers of the severely disabled with little assets enabling them to self-insure. Increasing the generosity of food stamps leads to a fall off in false applications for DI and a reduction in misdirected insurance, leading to better targeting of DI and a substantial welfare improvement despite the extra cost of food stamps. More frequent reassessments of recipients addresses the problem of individuals recovering while on DI and directly reduces the number of false claimants, but it also reduces the number of recipients who are severely work limited, and this leads to (small) welfare losses.

In summary, welfare increases if the threshold for acceptance is lower, disability payments are higher, reassessment less frequent, and food stamp payments more generous. These conclusions arose because these reforms lead to better coverage for the severely work limited. This highlights the need to have disability classified into more than just a "yes" or "no" state, and raises the question of whether allowing for partial disability and partial DI payments (as in the Netherlands, for example) may be a way to reduce the incentive cost of DI. One limitation of these policy conclusions is the clear nonlinearities in behavior apparent from the simulations in Section V. This highlights the value of having careful structural models of behavior in analyzing disability shocks and the DI process.

One of the implications of our simulations is that changes to the DI process can have sizable effects on asset accumulation, both by changing the need for self-insurance and by changing the amount of time that individuals spend out of the labor force. Related to this, Golosov and Tsyvinski (2006) propose that an asset-test should be introduced to the DI award process to identify those applicants who accumulated assets explicitly to smooth consumption while falsely claiming DI. We could, in principle, explore in our framework whether an asset test discourages applicants among the moderately or severely disabled. However, the difficulty of performing such an exercise is that assets in our framework are fully fungible and serve multiple purposes, including retirement saving, general consumption smoothing, as well as self-insurance. An asset test for DI applicants would therefore have the unfortunate side effect of reducing retirement saving.

In terms of limitations and further extensions, our model of the disability insurance process is incomplete: Benítez-Silva et al. (2004) and French and Song (2014) have emphasized the importance of the appeal process, whereas we have allowed the Social Security Administration to make just one decision, albeit we assume that individuals in the model are able to reapply. In the context of capturing behavior over the life cycle this may be less problematic, but it means we cannot examine one dimension of reform, namely the strictness and length of the appeal judgment relative to the initial judgment. A second restriction is in terms of the stochastic process for work limitations, which we take to be exogenous. The probability of receiving a negative shock to the ability to work is likely to be partly under the individual's control, through occupation choice and other decisions on the job. These decisions may be affected by the properties of the disability insurance scheme. Finally, we have ignored the health insurance component of the program, which Kitao (2014) suggests is important. This means we estimate a lower bound of the insurance value provided by the program.

REFERENCES

Adda, Jerome, James Banks, and Hans-Martin von Gaudecker. 2009. "The Impact of Income Shocks on Health: Evidence from Cohort Data." *Journal of the European Economic Association* 7 (6): 1361–99.

Attanasio, Orazio P., James Banks, Meghir Costas, and Guglielmo Weber. 1999. "Humps and Bumps in Lifetime Consumption." *Journal of Business and Economic Statistics* 17 (1): 22–35.

Attanasio, Orazio, Hamish Low, and Virginia Sanchez-Marcos. 2008. "Explaining Changes in Female Labor Supply in a Life-Cycle Model." *American Economic Review* 98 (4): 1517–52.

Attanasio, Orazio P., and Guglielmo Weber. 1995. "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* 103 (6): 1121–57.

Autor, David H., and Mark G. Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118 (1): 157–205.

Ball, Steffan, and Hamish Low. 2014. "Do Self-Insurance and Disability Insurance Prevent Consumption Loss on Disability?" *Economica* 81 (323): 468–90.

Banks, James, Richard Blundell, and Agar Brugiavini. 2001. "Risk Pooling, Precautionary Saving and Consumption Growth." *Review of Economic Studies* 68 (4): 757–79.

Benítez-Silva, Hugo, Moshe Buchinsky, Hiu Man Chan, Sofia Cheidvasser, and John Rust. 2004. "How Large Is the Bias in Self-Reported Disability?" *Journal of Applied Econometrics* 19 (6): 649–70.

Benítez-Silva, Hugo, Moshe Buchinsky, and John Rust. 2006a. "How Large are the Classification Errors in the Social Security Disability Award Process?" National Bureau of Economic Research Working Paper 10219.

Benítez-Silva, Hugo, Moshe Buchinsky, and John Rust. 2006b. "Induced Entry Effects of a $1 for $2 Offset in SSDI Benefits." https://editorialexpress.com/jrust/crest_lectures/induced_entry.pdf (accessed August 13, 2015).

Benítez-Silva, Hugo, Moshe Buchinsky, Hiu Man Chan, John Rust, and Sofia Sheidvasser. 1999. "An Empirical Analysis of the Social Security Disability Application, Appeal, and Award Process." *Labour Economics* 6 (2): 147–78.

Black, Dan, Kermit Daniel, and Seth Sanders. 2002. "The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust." *American Economic Review* 92 (1): 27–50.

Blank, Rebecca M. 2002. "Evaluating Welfare Reform in the United States." *Journal of Economic Literature* 40 (4): 1105–1166.

Blundell, Richard, Martin Browning, and Costas Meghir. 1994. "Consumer Demand and the Life-Cycle Allocation of Household Expenditures." *Review of Economic Studies* 61 (1): 57–80.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2014. "Consumption Inequality and Family Labor Supply." European Central Bank Working Paper 1656.

Bound, John. 1989. "The Health and Earnings of Rejected Disability Insurance Applicants." *American Economic Review* 79 (3): 482–503.

Bound, John. 1991. "Self-Reported versus Objective Measures of Health in Retirement Models." *Journal of Human Resources* 26 (1): 106–38.

Bound, John, and Richard V. Burkhauser. 1999. "Economic Analysis of Transfer Programs Targeted on People with Disabilities." In *Handbook of Labor Economics*, Vol. 3C, edited by Orley Ashenfelter and David Card, 3417–3528. Amsterdam: Elsevier Science.

Bound, John, Julie Berry Cullen, Austin Nichols, and Lucie Schmidt. 2004. "The Welfare Implications of Increasing Disability Insurance Benefit Generosity." *Journal of Public Economics* 88 (12): 2487–2514.

Bound, John, Todd Stinebrickner, and Timothy Waidmann. 2010. "Health, Economic Resources and the Work Decisions of Older Men." *Journal of Econometrics* 156 (1): 106–29.

Burkhauser, Richard V., and Mary C. Daly. 1996. "Employment and Economic Well-Being Following the Onset of a Disability: The Role for Public Policy." In *Disability, Work, and Cash Benefits*, edited by Jerry L. Mashaw, Virginia P. Reno, Richard V. Burkhauser, and Monroe Berkowitz, 59–101. Kalamazoo, MI: Upjohn Institute for Employment Research.

Cagetti, Marco. 2003. "Wealth Accumulation over the Life Cycle and Precautionary Savings." *Journal of Business and Economic Statistics* 21 (3): 339–53.

Chandra, Amitabh, and Andrew A. Samwick. 2006. "Disability Risk and the Value of Disability Insurance." In *Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly*, edited by D. Cutler and D. Wise. Chicago: University of Chicago Press.

Chen, Susan, and Wilbert van der Klaauw. 2008. "The Effect of Disability Insurance on Labor Supply of Older Individuals in the 1990s." *Journal of Econometrics* 142 (2): 757–84.

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

Daly, Mary C., and Richard V. Burkhauser. 2003. "The Supplemental Security Income Program." In *Means-Tested Transfer Programs in the United States*, edited by Robert A. Moffitt, 79–139. Chicago: University of Chicago Press.

Denk, Oliver, and Jean-Baptiste Michau. 2013. "Optimal Social Security with Imperfect Tagging." HAL Cahier de Recherche 2013–04.

Diamond, Peter, and Eytan Sheshinski. 1995. "Economic Aspects of Optimal Disability Benefits." *Journal of Public Economics* 57 (1): 1–23.

Duggan, Mark, and Scott A. Imberman. 2009. "Why Are the Disability Rolls Skyrocketing? The Contribution of Population Characteristics, Economic Conditions, and Program Generosity." In *Health at Older Ages: The Causes and Consequences of Declining Disability among the Elderly*, edited by David M. Cutler and David A. Wise. Chicago: University of Chicago Press.

Finkelstein, Amy, Erzo F. P. Luttmer, and Matthew J. Notowidigdo. 2013. "What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption." *Journal of the European Economic Association* 11: 221–58.

Flavin, Marjorie, and Takashi Yamashita. 2002. "Owner-Occupied Housing and the Composition of the Household Portfolio." *American Economic Review* 92 (1): 345–62.

French, Eric, and Jae Song. 2014. "The Effect of Disability Insurance Receipt on Labor Supply." *American Economic Journal: Economic Policy* 6 (2): 291–337.

Golosov, Mikhail, and Aleh Tsyvinski. 2006. "Designing Optimal Disability Insurance: A Case for Asset Testing." *Journal of Political Economy* 114 (2): 257–79.

Gourieroux, Christian, Alain Monfort, and Eric Renault. 1993. "Indirect Inference." *Journal of Applied Econometrics* 8 (Supplement): S85–S118.

Gourinchas, Pierre-Oliver, and Jonathan A. Parker. 2002. "Consumption over the Life Cycle." *Econometrica* 70 (1): 47–89.

Halpern, Janice H. 1979. "The Social Security Disability Insurance: Reasons for Its Growth and Prospects for the Future." *New England Economic Review* May/June: 30–48.

Halpern, Janice, and Jerry A. Hausman. 1986. "Choice under Uncertainty: A Model of Applications for the Social Security Disability Insurance Program." *Journal of Public Economics* 31 (2): 131–61.

Haveman, Robert, and Barbara Wolfe. 2000. "The Economics of Disability and Disability Policy." In *Handbook of Health Economics*, Vol. 1, edited by Anthony J. Culyer and Joseph P. Newhouse, 995–1051. Amsterdam: Elsevier Science.

Hoynes, Hilary Williamson, and Robert Moffitt. 1999. "Tax Rates and Work Incentives in the Social Security Disability Insurance Program: Current Law and Alternative Reforms." *National Tax Journal* 52 (4): 623–54.

Kitao, Sagiri. 2014. "A Life-Cycle Model of Unemployment and Disability Insurance." *Journal of Monetary Economics* 68 (C): 1–18.

Kreider, Brent. 1998. "Workers' Applications to Social Insurance Programs When Earnings and Eligibility Are Uncertain." *Journal of Labor Economics* 16 (4): 848–77.

Kreider, Brent. 1999. "Latent Work Disability and Reporting Bias." *Journal of Human Resources* 34 (4): 734–69.

Lando, Mordechai E., Malcolm B. Coate, and Ruth Kraus. 1979. "Disability Benefit Applications and the Economy." *Social Security Bulletin* 42 (10): 3–10.

Lillard, Lee A., and Yoram Weiss. 1997. "Uncertain Health and Survival: Effects on End-of-Life Consumption." *Journal of Business and Economic Statistics* 15 (2): 254–68.

Low, Hamish, Costas Meghir, and Luigi Pistaferri. 2010. "Wage Risk and Employment Risk over the Life Cycle." *American Economic Review* 100 (4): 1432–67.

Low, Hamish, and Luigi Pistaferri. 2015. "Disability Insurance and the Dynamics of the Incentive Insurance Trade-off: Dataset." *American Economic Review*. http://dx.doi.org/10.1257/aer.20110108.

MaCurdy, Thomas E. 1982. "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics* 18 (1): 83–114.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review* 103 (5): 1797–1829.

Meyer, Bruce D., and Wallace K. C. Mok. 2013. "Disability, Earnings, Income and Consumption." National Bureau of Economic Research Working Paper 18869.

Nagi, Saad Z. 1969. *Disability and Rehabilitation*. Columbus, OH: Ohio State University Press.

Newey, Whitney K. 2009. "Two-Step Series Estimation of Sample Selection Models." *Econometrics Journal* 12 (1): S217–29.

Parsons, Donald O. 1980. "The Decline in Male Labor Force Participation." *Journal of Political Economy* 88 (1): 117–34.

Smith, James P. 2004. "Unravelling the SES Health Connection." Institute for Fiscal Studies, IFS Working Paper W04/02.

Social Security Administration. 2014. *Annual Statistical Report on the Social Security Disability Insurance Program*. SSA Publication No. 13–11826.

Stephens, Melvin, Jr. 2001. "The Long-Run Consumption Effects of Earnings Shocks." *Review of Economics and Statistics* 83 (1): 28–36.

Stern, Steven. 1989. "Measuring the Effect of Disability on Labor Force Participation." *Journal of Human Resources* 24 (3): 361–95.

von Wachter, Till, Jae Song, and Joyce Manchester. 2011. "Trends in Employment and Earnings of Allowed and Rejected Applicants to the Social Security Disability Insurance Program." *American Economic Review* 101 (7): 3308–29.

Waidmann, Timothy A., John Bound, and Austin Nichols. 2003. "Disability Benefits as Social Insurance: Tradeoffs Between Screening Stringency and Benefit Generosity in Optimal Program Design." University of Michigan, Michigan Retirement Research Center Working Paper 2003-042.

This article has been cited by:

1. Alexander Gelber, Timothy J. Moore, Alexander Strand. 2017. The Effect of Disability Insurance Payments on Beneficiaries' Earnings. *American Economic Journal: Economic Policy* **9**:3, 229-261. [Abstract] [View PDF article] [PDF with links]
2. Hamish Low, Costas Meghir. 2017. The Use of Structural Models in Econometrics. *Journal of Economic Perspectives* **31**:2, 33-58. [Abstract] [View PDF article] [PDF with links]
3. Richard Blundell. 2017. What Have We Learned from Structural Models?. *American Economic Review* **107**:5, 287-292. [Abstract] [View PDF article] [PDF with links]
4. Manasi Deshpande. 2016. Does Welfare Inhibit Success? The Long-Term Effects of Removing Low-Income Youth from the Disability Rolls. *American Economic Review* **106**:11, 3300-3330. [Abstract] [View PDF article] [PDF with links]