Disability Insurance and the Dynamics of the Incentive-Insurance Trade-off

Online Appendix: Not For Publication

Hamish Low
University of Cambridge

Luigi Pistaferri
Stanford University

May 19, 2015

In this online appendix we discuss a number of issues that we left out of the main text for reasons of space. Section A describes further details of the model and solution method. Section B provides further information on the data, including comparability between our data and alternative sources, the quality of self-reported measures, and the consumption imputation. Section C describes the estimation method more fully. Section D provides extensive further robustness checks of the estimation results, particularly of the exclusion restrictions, and of the assumed ways that health enters the model. Section E provides more details on the robustness of our policy experiments to changing the distribution of the noisy signal that the government observes and to changes in risk aversion.

A Further Model Details

A.1 Preferences

We use a utility function of the form

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

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.

The complementarity or substitutability between consumption and poor health is governed by the sign of θ. To see this, note that from the first-order condition for consumption, $u_c = \lambda$, one can take the total derivative to yield:

$$u_{cc}dc + u_{cL}dL + u_{cP}dP = d\lambda.$$

The Frisch complementarity can be shown by setting $d\lambda = dP = 0$, and rearranging:

$$\left.\frac{dc}{dL}\right|_{d\lambda=0,dP=0} = -\frac{u_{cL}}{u_{cc}}$$

which implies (given concavity):

$$sign\left(\left.\frac{dc}{dL}\right|_{d\lambda=0,dE=0}\right) = sign\left(u_{cL}\right) = -sign\left(\theta\right),$$

and so consumption and (poor) health are Frisch complements only if $\theta < 0$ (i.e., the marginal utility of consumption is higher when suffering from a work limitation). Similarly, consumption and employment are Frisch complements if $\eta < 0$. In our empirical framework, we pre-set $\gamma = 1.5$ and we lack explicit additive terms in L and P. Hence we require $\theta < 0$ and $\eta < 0$ to be consistent with disability and work being "bads".

A.2 Unemployment Insurance

We assume that unemployment benefits are paid only for the quarter immediately following job destruction. Unemployment insurance is paid only to people who have worked in the previous period, and only to those who had their job destroyed (job quitters are ineligible for UI payments, and we assume this can be perfectly monitored). We assume $B_{it} = b \times w_{it-1}h$, subject to a cap, and we set the replacement ratio $b = 75\%$. This replacement ratio is set at this high value because the payment that is made is intended to be of a similar magnitude to the maximum available to someone becoming unemployed (i.e., a 75% replacement rate paid over a single quarter is comparable in expectation to the average 45% replacement rate over 2 quarters that we observe in practice).[1] This simplifying assumption means that, since the period of choice is one quarter, unemployment benefit is like a lump-sum payment to those who exogenously lose their job and so does not distort the choice about whether or not to accept a new job offer. Similarly, there is no insurance against the possibility of not receiving a job offer after job loss.

In the US, the UI program varies across states. It would be very complex (in our model, at least) to capture the myriad differences that exist across US states in terms of UI provision. Our modeling choices are instead intended to replicate the most salient features of the program.

[1] The advantage of this choice is primarily computational: We do not have to keep track of one's employment status and eliminate moral hazard in job search that is not the direct focus of the paper.

A.3 Universal Means-Tested Program

We assume that the universal means-tested program is an anti-poverty program providing a floor to income for all individuals, similar to the actual food stamps program but with three important differences. First, means-testing is on household income rather than on income and assets;[2] second, the program provides a cash benefit rather than a benefit in kind;[3] and third, we assume there is 100% take-up.

The value of the program is given by

$$W_{it} = \begin{cases} \overline{T} - 0.3 \times y_{it} & \text{if } Z_{it}^{W} = 1 \ \left(\text{i.e., if } y_{it} \leq \underline{y}\right) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

In reality the maximum value of the payment, $\overline{T}$, depends on family size and composition. In our model we assume that the individuals' family size changes over the life cycle as to mimic the average family size in the data, and we adjust the food stamps generosity variable $\overline{T}$ accordingly using actual program parameters. The term $\underline{y}$ is the poverty line and so only people with net earnings y_{it} below the poverty line are eligible. Net income y_{it} is defined as $y = (1 - \tau_w)\, y^{gross} - d$, where d is the standard deduction that people are entitled to when computing net income for the purpose of determining food stamp allowances. Finally, y^{gross} is given by

$$y_{it}^{gross} = w_{it} h P_{it} + \left(B_{it} Z_{it}^{UI} \left(1 - Z_{it}^{DI}\right) + D_{it} Z_{it}^{DI} + SSI_{it} Z_{it}^{DI} Z_{it}^{W}\right) (1 - P_{it}) \tag{3}$$

The amount of food stamp benefits that individuals receive in the model is determined by first computing y_{it}^{gross} (which depends on employment choices and whether the person is receiving other government payments if unemployed), then computing net income y_{it}, and finally using (2). As we discuss in the main text, this means-tested program interacts in complex ways with disability insurance: the Food Stamps program provides a consumption floor during application for DI, and an alternative mechanism for income support for those of low productivity.

A.4 Taxation on Earnings

The government's main source of revenue is the tax on labor. When people are in the working stage of the life cycle, the government's outlays are partly on social insurance payments to people

[2] The difficulty with allowing for an asset test in our model is that there is only one sort of asset which individuals use for retirement saving as well as for short-term smoothing. In reality, the asset test applies only to liquid wealth and thus excludes pension wealth (as well as real estate wealth and other durables).

[3] See Hoynes and Schanzenbach (2009) for evidence that families react to food stamps as if it were cash.

on UI, DI and SSI, and partly on the means-tested welfare program. The government also spends q for each DI beneficiary who is re-assessed (which happens with probability P_L^{Re}).[4]

When performing the policy experiments we use as a baseline a case in which the tax rate on labor is set to 0, and hence we let the government run a deficit $\bar{D}$. When varying parameters of the social insurance programs, we hold the government budget deficit $\bar{D}$ constant. This is achieved by adjusting the proportional payroll tax, τ_w, such that the present discounted value of net revenue flows is constant:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{R^t}\left[\left(B_{it}Z_{it}^{UI}\left(1-Z_{it}^{DI}\right)+D_{it}Z_{it}^{DI}+SSI_{it}Z_{it}^{DI}Z_{it}^{W}\right)\left(1-P_{it}\right)+W_{it}Z_{it}^{W}+qP^{\text{Re}}Z_{it}^{DI}\right] \qquad (4)$$
$$=\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{R^t}\tau_w w_{it}hP_{it}+\bar{D}$$

This is done iteratively because labor supply and DI application decisions change as a consequence of changes in government policy.

A.5 Disability Insurance Payments and the Award Process

The value of disability insurance received by DI beneficiaries is given by

$$D_{it}=\begin{cases}0.9\times\overline{w}_i & \text{if } \overline{w}_i\leq a_1\\ 0.9\times a_1+0.32\times(\overline{w}_i-a_1) & \text{if } a_1<\overline{w}_i\leq a_2\\ 0.9\times a_1+0.32\times(a_2-a_1)+0.15\times(\overline{w}_i-a_2) & \text{if } a_2<\overline{w}_i\leq a_3\\ 0.9\times a_1+0.32\times(a_2-a_1)+0.15\,(a_3-a_2) & \text{if } \overline{w}_i>a_3\end{cases} \qquad (5)$$

where $\overline{w}_i$ is average earnings computed before the time of the application and a_1, a_2, and a_3 are thresholds we take from the legislation.[5] We assume $\overline{w}_i$ can be approximated by the value of the permanent wage.

Figure 1 displays the five steps of the DI determination process. These steps are discussed in detail in many papers, including Bound and Burkhauser (1999), to which the interested reader is referred. At step 1 applicants who are working (or more precisely, earning above a limit known as Substantial Gainful Activity, SGA) are immediately excluded. At step 2, applicants who do not satisfy a medical requirement ("a determinable impairment that has lasted or can be expected to last 12 consecutive months or to result in death") are also excluded. At step 3, the SSA awards benefits without any further review to applicants who have a "listed impairment", i.e., an impairment that

[4] 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 use this figure deflating it back to 1992 prices.

[5] In reality what is capped is $\overline{w}_i$ (the AIME), because annual earnings above a certain threshold are not subject to payroll taxation. We translate a cap on AIME into a cap on DI payments.

is easily identifiable and recognizable (such as amputation of both hands, leukemia, etc.). At steps 4 and 5 the SSA awards benefits to individuals who, despite not having a listed impairment, have a vocational disability, i.e., are unable to perform their previous jobs and are unable to perform other jobs that "exist in significant numbers in the national economy" and that fit the applicant's age, education and past work experience.

In reality, people may appeal a negative decision by the SSA. We do not model appeals explicitly. However, since individuals in our model can re-apply at any point in the future following a rejection, the appeal option is implicitly available to them and we use this consideration to compute the award rates since time of first application that are reported in Section 5.6.1 in the main text.

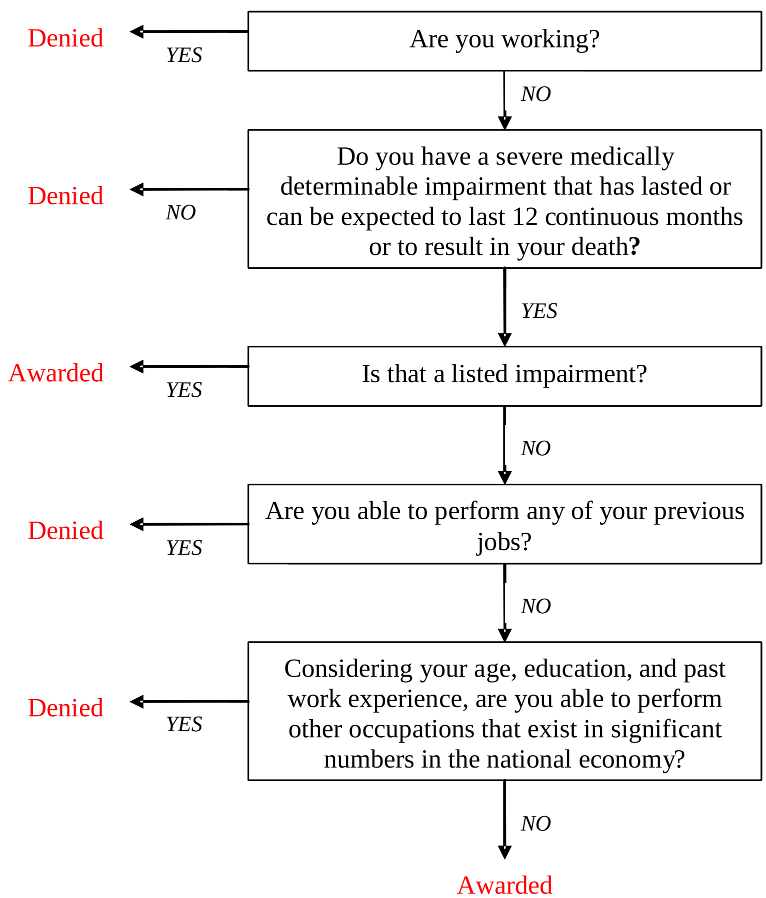

Figure 1: Disability Insurance Determination

A.6 Solution Method

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. We do this separately for each of the fixed types, f_i.

We start by constructing the value functions for the individual when employed and when out of work (conditional on "type"). When employed, the state variables are $\{A_{it}, \varepsilon_{it}, L_{it}\}$, corresponding to current assets, individual productivity and work limitation status. We denote the value function when employed as V^e. When unemployed, there are three alternative discrete states the individual can be: unemployed and not applying for disability (giving a value V^n), unemployed and applying for disability (giving a value V^{App}), and unemployed and already receiving disability insurance (giving a value V^{Succ}). We consider the specification of each of these value functions in turn.

The value function if working is written as:

$$
\begin{aligned}
&V_t^e\left(A_{it}, \varepsilon_{it}, L_{it}\right) = \\
&\max_c \left\{ \begin{array}{c} U\left(c_{it}, P_{it} = 1; L_{it}\right) + \\ \beta\delta E_t\left[V_{t+1}^n\left(A_{it+1}, \varepsilon_{it+1}, L_{it+1}, DI_{it+1}^{Elig} = 1\right)\right] \\ +\beta\left(1-\delta\right) E_t \max\left\{ \begin{array}{c} V_{t+1}^n\left(A_{it+1}, \varepsilon_{it+1}, L_{it+1}, DI_{it+1}^{Elig} = 1\right) \\ V_{t+1}^e\left(A_{it+1}, \varepsilon_{it+1}, L_{it+1}\right) \end{array} \right\} \end{array} \right.
\end{aligned}
$$

where DI_{it+1}^{Elig} is an indicator for whether the individual is eligible to apply for DI.

Consider now the value function for an unemployed individual who is not applying for disability insurance in period t. We need to define as a state variable whether or not an individual has already applied for disability in the current unemployment spell in order to distinguish between those who have the option of applying for disability and those who are ineligible to apply. The value function when eligible for disability is given by:

$$V_t^n\left(A_{it},\varepsilon_{it},L_{it},DI_{it}^{Elig}=1\right)=$$

$$\max_{c,DI^{App}}\left\{\begin{array}{c} u\left(c_{it},P_{it}=0;L_{it}\right) \\ \\ +\beta\left\{1.App=1\right\}E_tV_{t+1}^{App}\left(A_{it+1},\varepsilon_{it+1},L_{it+1}\right) \\ \\ +\beta\left\{1.App=0\right\} \\ \left[\begin{array}{c} \lambda^n E_t\max\left\{\begin{array}{c} V_{t+1}^n\left(A_{it+1},\varepsilon_{it+1},L_{it+1},DI_{it+1}^{Elig}=1\right) \\ V_{t+1}^e\left(A_{it+1},\varepsilon_{it+1},L_{it+1}\right)\end{array}\right\} \\ +\left(1-\lambda^n\right)E_t\left[V_{t+1}^n\left(A_{it+1},\varepsilon_{it+1},L_{it+1},DI_{it+1}^{Elig}=1\right)\right]\end{array}\right]\end{array}\right.$$

The value function when applying is given by

$$V_t^{App}\left(A_{it},\varepsilon_{it},L_{it}\right)=$$

$$\max_c\left\{\begin{array}{c} u\left(c_{it},P_{it}=0;L_{it}\right) \\ \\ +\beta\pi_L^tE_tV_{t+1}^{Succ}\left(A_{it+1},\varepsilon_{it+1},L_{it+1},D_t=0\right) \\ \\ +\beta(1-\pi_L^t)E_tV_{t+1}^n\left(A_{it+1},\varepsilon_{it+1},L_{it+1},DI_{t+1}^{Elig}=0\right)\end{array}\right.$$

where D_t is a state variable for the duration of the spell on disability insurance and

$$\pi_L^t=\Pr\left(DI_{it+1}=1|\,DI_{it}^{App}=1,L_{it},t\right)$$

is the probability of a successful application.

Finally, we have to define the value function if an application for disability has been successful.

$$V_t^{Succ}\left(A_{it},\varepsilon_{it},L_{it},DI_{it}\right)= \tag{6}$$

$$\max_c\left\{\begin{array}{c} u\left(c_{it},L_{it},P_{it}=0\right) \\ +\beta\left(1-\pi_t^R\right)E_t\left[\max\left\{\begin{array}{c} V_{t+1}^{Succ}\left(A_{it+1},\varepsilon_{it+1},L_{it+1},DI_{it+1}\right) \\ V_{t+1}^e\left(A_{it+1},\varepsilon_{it+1},L_{it+1}\right)\end{array}\right\}\right] \\ +\beta\pi_t^RE_t\left[V_{t+1}^n\left(A_{it+1},\varepsilon_{it+1},L_{it+1},DI_{it+1}^{Elig}=0\right)\right]\end{array}\right\} \tag{7}$$

where

$$\begin{aligned}\pi_t^R &= P_L^{\mathrm{Re}}\Pr\left(DI_{it+1}=0\,|L_{it+1},DI_{it}=1,t,\mathrm{Re}\right) \\ &= P_L^{\mathrm{Re}}\left(1-\pi_t^L\right)\end{aligned}$$

is the probability of being reassessed and removed from the program.

Our model has discrete state variables for: Wage productivity, Work limitation status, Participation, Eligibility to apply for DI (if not working), and Length of time on DI (over 1 year or less than 1 year). The only continuous state variable is assets. We use backward induction to obtain policy functions.

Value functions are increasing in assets A_{it} but they are not necessarily concave, even if we condition on labor market status in t. The non-concavity arises because of changes in labor market status in future periods: the slope of the value function is given by the marginal utility of consumption, but this is not monotonic in the asset stock because consumption can decline as assets increase and expected labor market status in future periods changes. This problem is also discussed in Lentz and Tranaes (2001), Attanasio et al. (2008) and in Low et al. (2010). By contrast, in Danforth (1979) employment is an absorbing state and so the conditional value function will be concave. Under certainty, the number of kinks in the conditional value function is given by the number of periods of life remaining. If there is enough uncertainty, then changes in work status in the future will be smoothed out leaving the expected value function concave: whether or not an individual will work in $t+1$ at a given A_{it} depends on the realization of shocks in $t+1$. Using uncertainty to avoid non-concavities is analogous to the use of lotteries elsewhere in the literature. In the value functions above, the choice of participation status in $t+1$ is determined by the maximum of the conditional value functions in $t+1$.

B Further Data Issues

B.1 Sample Selection

We focus on male heads because in the PSID, it is the head (almost always the male) who assesses the work limitations of both partners. Using "second-hand" measures of disability status may be misleading. There is now an established literature (see Kapteyn, Smith and van Soest, 2007) arguing that individuals use different response scales when they answer questions about work disability (or any other economic question that involve measurement on a subjective scale, such as assessing the probability of keeping one's job, etc.). If men are tough critics of disability and they are asked to evaluate someone else's disability status, their report of their partner's disability status may be biased downward. Similarly, if men have different subjective assessments of pain than women, their report about the work disability status of their partner may be biased towards their own "pain threshold" (see Mogil and Bailey, 2010, and the references therein).

We exclude those younger than 23 or older than 62, those with missing reports on education, the state of residence, the self-employed, those with less than 3 years of data, and some hourly wage outliers (those with an average hourly wage that is below half the minimum wage and those whose hourly wage declines by more than 85% or grows by more than 500%).[6] To identify whether an individual in the PSID is receiving DI, we use a question that asks whether the amount of social security payments received was due to disability.[7]

B.2 Comparison of PSID to Alternative Data Sources

Figures 2 reports a comparison of disability rates in different surveys (the source is Bound and Burkhauser, 1999). Figure 3 shows a comparison of trends in DI rates by age as estimated in the PSID and using aggregate statistics. These figures show that (a) disability rates as estimated in the PSID are not different from those estimated in other, larger data sets, and (b) that the PSID reproduces quite well the age-gradient of the fraction of people receiving DI, although the levels are slightly lower (most likely because we focus only on male heads rather than all males in this age group and because we miss the institutionalized population).

[6] The hourly wage is defined as annual earnings/annual hours.

[7] The survey first asks the amount of Social Security payments received in year t by the year $t+1$ head. Then, it asks *Was that disability, retirement, survivor's benefits, or what?*. Possible responses are: 1) Disability, 2) Retirement, 3) Survivor's benefits; dependent of deceased recipient, 4) Dependent of disabled recipient, 5) Dependent of retired recipient, 6) Other, 7) Any combination of the codes above.

Data	Year	Survey question	Population	% of population with disabilities
PSID	1989	Do you have any nervous or physical condition that limits the type or the amount of work you can do? (Must have responded yes in both 1988 and 1989)	Men 25-61 Women 25-61	9.2 10.6
CPS	1990	Do you have a health problem or disability which prevents you from working or which limits the kind or the amount of work you can do? Or, Main reason did not work in 1989 was ill or disabled; or Current reason not looking for work is ill or disabled (One period)	Men 25-61 Women 25-61	8.1 7.8
SIPP	1990	Do you have a physical, mental, or other health condition which limits the kind or amount of work you can do? (One period)	Men 21-64 Women 21-64	11.7 11.6
SIPP	1990	Do you have a physical, mental, or other health condition which limits the kind or amount of work you can do? (Must have responded yes in wave 3 and wave 6)	Men 25-61 Women 25-61	9.8 9.8
NHIS	1994	Are you limited in the kind or amount of work you can do because of any impairment or health problem? (One period)	Men 25-61 Women 25-61	10.8 11.4
HRS	1992	Do you have any impairment or health problem that limits the kind or amount of paid work you can do? (One period)	Men 51-61 Women 51-61	27.3 29.8

Figure 2: Survey questions on disability and estimated disability rates (Source: Bound and Burkhauser, 1999).

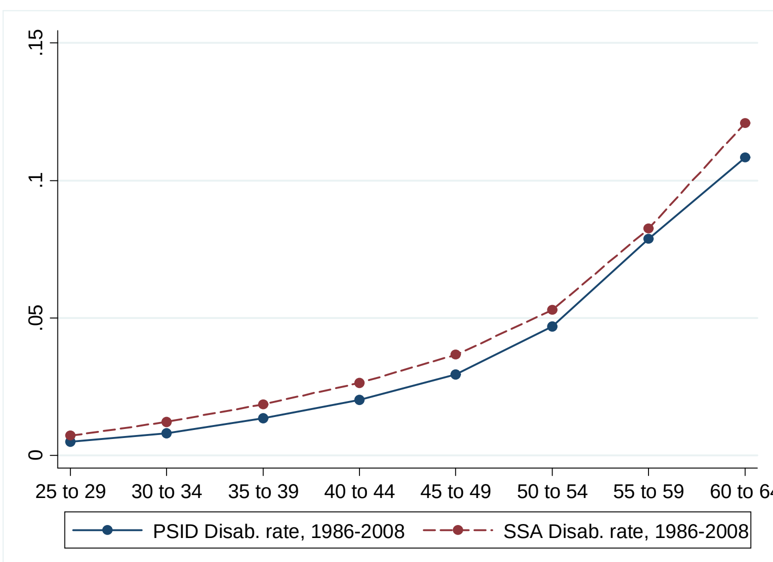

Figure 3: Fraction receiving DI by age, PSID (male heads) and aggregate data.

B.3 Heterogeneity by skills

The empirical analysis in the paper focuses only on the low educated group. Here we report summary statistics on disability status and DI receipt by education groups (which we take as proxies for skill levels). There are important differences by skill level both in terms of probability of disability shocks and disability insurance recipiency rates. In particular, we find that individuals with low education (at most high school degree) and high education (some college or more) have very similar DI recipiency rates until their mid 30s, but after that the difference increases dramatically (see Figure 4). By age 60, the low educated are 2.5 times more likely to be DI claimants than the high educated (17% vs. 7%). In part, this is due to the fact that low educated individuals are more likely to have some form of disability (moderate or severe) at all ages, as shown in Figure 5.

The proportions on DI have varied over time, and particularly there has been a marked increase in recipiency numbers from the trough in 1984, as documented in Duggan and Imberman (2009) among others. The age composition of those flowing onto DI is shown in figure 6, which highlights that capturing the behavior of those under 50 is an important part of our understanding of disability insurance incentive-insurance dynamic trade-off.

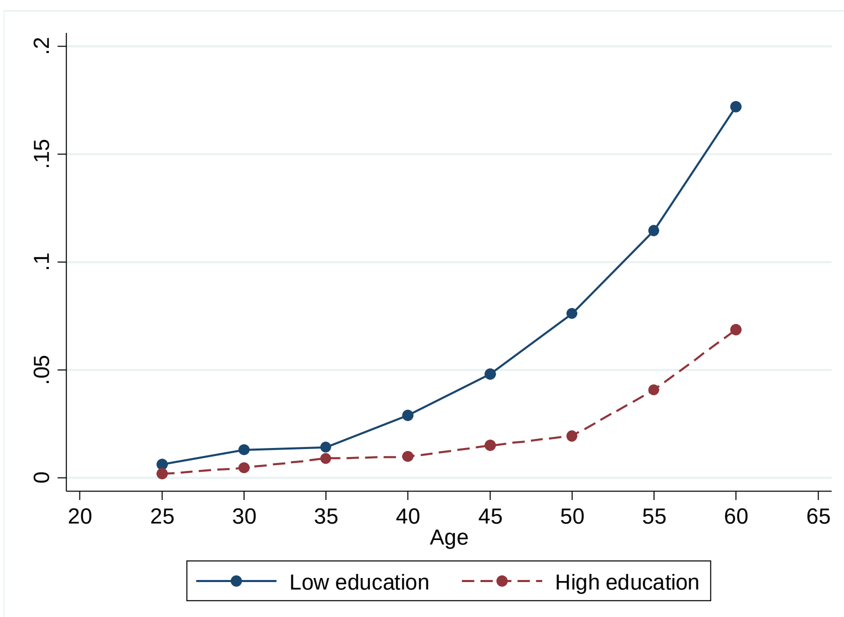

Figure 4: Fraction of individuals receiving disability insurance.

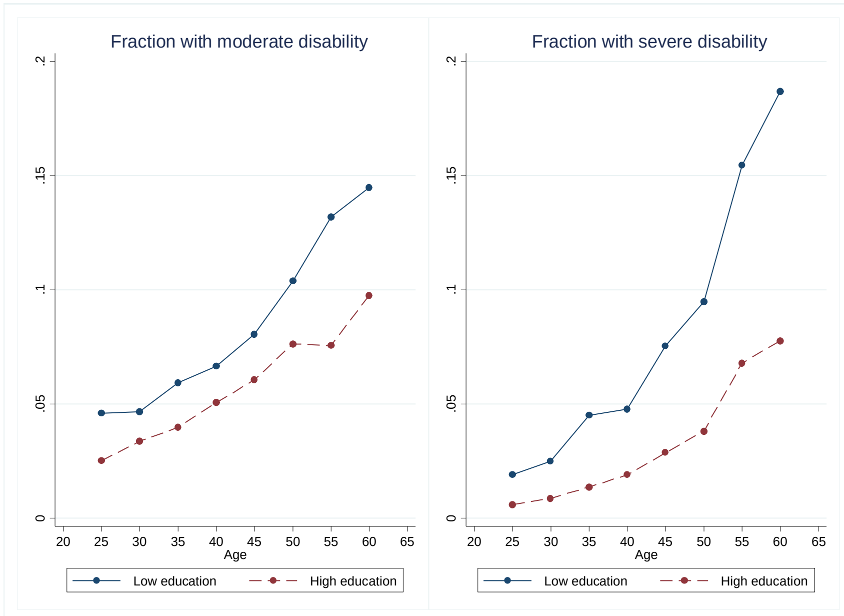

Figure 5: Fraction of individuals with a moderate or severe disability.

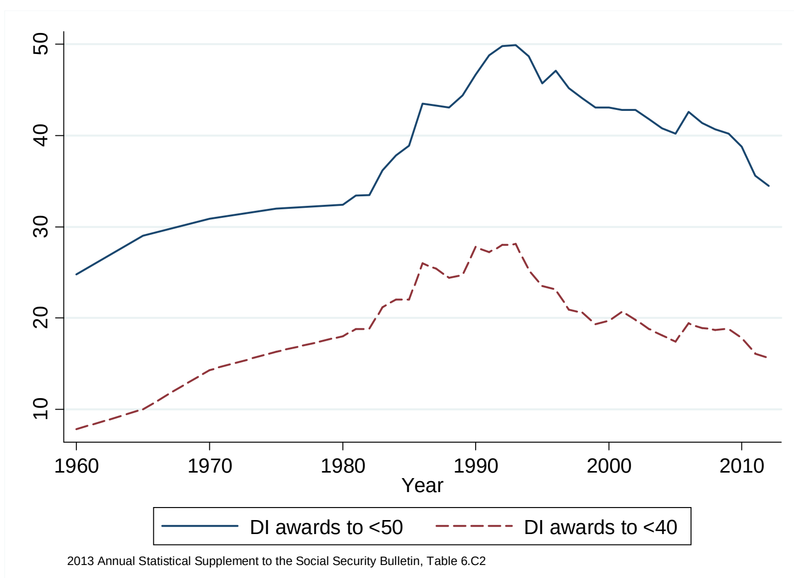

Figure 6: Proportion of DI Awards.

B.4 Self-Reported Disability Measures

Is our self-reported disability measure informative about objective impairment conditions? The 1986 wave of the PSID and most of the waves after 1997 include special modules containing detailed information about the respondents' health, including "objective" or "clinical" measures of health, as discussed in Burkhauser and Daly (1986). Table 1 shows that our distinction between a severe and moderate disability is correlated with a wide variety of objective/clinical measures, including Activities of Daily Living (ADLs), hospital stays, extreme BMI values (obesity and being underweight), subsequent mortality, and prevalence of specific medical conditions (morbidity). This confirms that our indicator provides a meaningful and informative classification of work limitations due to disabilities.

Table 1: **Validity of Self-Reported Disability Status**

Objective indicator	Survey Year	$L = 0$	$L = 1$	$L = 2$
Trouble walking	1986	0.0503	0.5376	0.7909
Trouble bending	1986	0.0733	0.6129	0.8273
Trouble driving	1986	0.0064	0.1183	0.3455
Trouble eyesight	1986	0.0244	0.0591	0.1545
Need travel assistance	1986	0.0015	0.043	0.2636
Need to stay inside	1986	0.0034	0.0538	0.3182
Stays in bed	1986	0	0.0699	0.2545
Limited physical activity	1986	0.0958	0.8172	0.9636
Underweight	1986	0.0059	0.0108	0.0636
Obese	1986	0.1413	0.2151	0.1818
Dies between 1986 and 2009	1986-2009	0.154	0.2957	0.4273
Spent time in hospital	1986, 2003-09	0.0544	0.1892	0.3142
Days in hospital	1986, 2003-09	0.3625	1.9099	5.1403
Had stroke	1999-2009	0.0089	0.0533	0.0919
Has high blood pressure	1999-2009	0.1995	0.4006	0.482
Has diabetes	1999-2009	0.0616	0.1319	0.205
Has cancer	1999-2009	0.0123	0.0378	0.0522
Has lung disease	1999-2009	0.0193	0.0834	0.1528
Had heart attack	1999-2009	0.0173	0.0883	0.1441
Has heart disease	1999-2009	0.0202	0.0892	0.159
Has emotional troubles	1999-2009	0.0282	0.1484	0.2696
Has arthrirtis	1999-2009	0.078	0.2852	0.4335
Has asthma	1999-2009	0.0613	0.1261	0.1478

Note: The sample is low education male heads of household, aged 23-62.

B.5 Consumption Data

Before 1999, PSID collected data on very few consumption items, such as food and (occasionally) rent and child care. However, starting in 1999 consumption expenditure data cover many other nondurable and services consumption categories, including utilities, gasoline, car maintenance, health expenditures, transportation, education and child care. Other consumption categories have been added starting in 2005 (such as clothing), but we do not use these other categories to keep the consumption series consistent over time. The main items that are missing are clothing, recreation, alcohol and tobacco.

While rent is reported whenever the household rents a house, it is not reported for home owners. To construct a series of housing services for home owners we impute the rent expenditures for home owners using the self reported house price.[8] We then aggregate all nondurable and services consumption categories to get the household consumption series.[9]

Blundell, Pistaferri and Saporta-Eksten (2014) report and discuss descriptive statistics on the various components of aggregate consumption. A comparison of the main aggregates (total consumption, nondurables, and services) against the NIPA series is offered in Table 2. As shown in Table 2, taking into account that the PSID consumption categories that we use are meant to cover 70% of consumption expenditure, the coverage rate is remarkably good.

C Further Details of the Estimation Process

C.1 Exclusion restrictions in the employment equation

Our reduced form model of employment is:

$$\begin{aligned} P_{it}^{*} &= X_{it}'\gamma + \sum_{j=1}^{2} \delta_j L_{it}^{j} + \theta G_{it} + \vartheta_{it} \\ &= h_{it} + \vartheta_{it} \end{aligned} \tag{8}$$

where P_{it}^{*} is the utility from working, ϑ_{it} the unobserved "taste for work", and we observe the indicator $P_{it} = \mathbf{1}\{P_{it}^{*} > 0\}$. Finally, the vector G_{it} includes exclusion restrictions: They affect the likelihood of observing an individual at work (through an income effect and through affecting the

[8] For our baseline measure we approximate the rent equivalent as 6% of the house price. See Flavin and Yamashita (2002).

[9] We treat missing values in the consumption subcategories as zeros.

Table 2: **A comparison of PSID consumption data and NIPA consumption data.**

	1998	2000	2002	2004	2006	2008
PSID Total	3,276	3,769	4,285	5,058	5,926	5,736
NIPA Total	5,139	5,915	6,447	7,224	8,190	9,021
ratio	0.64	0.64	0.66	0.70	0.72	0.64
PSID Nondurables	746	855	887	1,015	1,188	1,146
NIPA Nondurables	1,330	1,543	1,618	1,831	2,089	2,296
ratio	0.56	0.55	0.55	0.55	0.57	0.50
PSID Services	2,530	2,914	3,398	4,043	4,738	4,590
NIPA Services	3,809	4,371	4,829	5,393	6,101	6,725
ratio	0.66	0.67	0.70	0.75	0.78	0.68

Note: We use PSID weights (we have a total of 47,206 obs. for 1999-09). Total consumption is defined as Nondurables + Services. PSID consumption categories include food, gasoline, utilities, health, rent (or rent equivalent), transportation, child care, education and other insurance. NIPA figures are from NIPA table 2.3.5. Data are expressed in nominal terms.

expectation that the individual will apply for DI in the subsequent period), but they do not affect the wage, conditional on X_{it} and L_{it}.

We assume that "potential" government transfers and its interaction with work limitation 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 done 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 2004).

In general, benefits for a given program p are given by:

$$\widetilde{B_{it}^{p}} = m_{it} f\left(d_{it}, w_{it}, l_{it}\right)$$

where d_{it} are exogenous characteristics (state of residence, year, number of children, etc.), w_{it} are family earnings, l_{it} are other endogenous variables used in the computation of benefits (such as liquid assets, etc.), m_{it} is a take up indicator, and $f(.)$ a set of state-year specific rules mapping exogenous and endogenous household characteristics into benefits.

For the purpose of computing *potential* benefits, we assume $m_{it} = 1$. We also have to choose a "representative" level of w_{it} and l_{it}. We assume that family earnings are given by the wages earned by a household head working part-time at the minimum wage (w_t), and that assets, etc., are the average of the head's demographic group g (defined by state of residence, year, number of children, education, and age), or $\overline{l_{gt}}$. These assumptions give us a value of *potential* benefits for program p equal to $B_{it}^p = f\left(d_{it}, w_t, \overline{l_{gt}}\right)$. Total potential benefits, which we use as our exclusion restriction, are then $B_{it} = \sum_p B_{it}^p$. Hence, in the computation of potential benefits 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).

The various components of total potential benefits are defined next.

Food Stamps Following Wilde (2001), we use the following formula to compute potential food stamps benefits:

$$B^{FS} = \max\{M - 0.3 * N_2, 10\}$$

where B are benefits paid, M is the maximum benefit amount (which varies by year and family size and which we take from legislation), and N_2 is net income, defined as:

$$N_2 = \max\{w_t - (D_1 + D_2 + D_3 + D_4 + D_5 + D_6), 0\}$$

where w_t is the representative household's family income, D_1 to D_6 are allowed deductions (standard, earned income, dependent care, medical, child support payment, and excess shelter expense deductions, respectively). We set the dependent care and medical deductions to zero because we have no information on child care or health care expenses before 1999. We use information on w_t, child support paid, and rent paid (together with legislation parameters) to compute the other deductions. The shelter deduction is defined as:

$$D_6 = \min\{H^*, \max\{H - 0.5 * N_1, 0\}\}$$

where H are shelter expenses, H^* a cap (from legislation), and N_1 intermediate net income (defined as $N_1 = \max\{w_t - (D_1 + D_2 + D_3 + D_4 + D_5), 0\}$. Finally, we impose eligibility rules (B^{FS} is set to zero if gross income is above 130% of poverty line or if net income is above poverty line).

AFDC/TANF Before 1996, we follow Solomon (1995) and compute potential AFDC benefits using the formula:

$$B^{AFDC} = \min\left\{\max\left\{0, r^{NS}NS - r^{y}y\right\}, M\right\}$$

where B^{AFDC} are benefits paid, r^{NS}, r^{y} and M are parameters that vary by state and year, NS are need standards (also varying by state and year, as well as family size) and y is countable income, which we define as the representative's household earned income net of AFDC deductions allowed under the policy. Finally, we impose eligibility rules and set B^{AFDC} to zero for households with no children, or households with before-deductions countable income above 185% of the poverty line or after-deductions countable income above the poverty line.

After the 1996 welfare reform, households may be eligible to receive TANF. We assume that B^{TANF} is equal to the maximum amount available in the state of residence in a given year (which varies by household demographic composition). We correct this amount for the earnings disregard amounts, also varying by state and year. Then, we impose eligibility rules and set $B^{TANF} = 0$ for households earning above a certain threshold (coded in the legislation and also varying by state and year) and those with no children.

EITC We set EITC benefits to:

$$B^{EITC} = \tau_1 w \mathbf{1}\left\{0 < w \leq k_1\right\} + \tau_1 k_1 \mathbf{1}\left\{k_1 < w \leq k_2\right\} + \left(\tau_1 k_1 - \tau_2\left(w - k_2\right)\right)\mathbf{1}\left\{k_2 < w \leq k_3\right\}$$

where $\mathbf{1}\{.\}$ are indicator functions, τ_1, τ_2, k_1, k_2 and k_3 are policy parameters that vary with family size and time and w is the representative household's earned income. We add state-specific EITC supplements, and set $B^{EITC} = 0$ for those with 0 earnings or earnings above k_3. See Hotz and Scholtz (2003).

UI We assume that those who qualify for UI are fully covered and receive it for 26 weeks. We compute UI benefits as:

$$B^{UI} = \max\left\{\min\left\{r^{UI}y, M\right\}, m\right\}$$

where r^{UI}, M, and m vary by state and time and y is "reference" earnings (which also varies by state; some states take it to be annual earnings, others take it to be the average of the best two quarters, and so forth - this again is a function of the representative household's earnings w_t). We add a dependent allowance that is available in certain states and that varies with time and family size. Details come from US D.O.L. (various years).

C.2 Selection-corrected wage estimates

Our log wage process is

$$\begin{aligned}\ln w_{it} &= X_{it}'\beta + \sum_{j=1}^{2} \varphi_j L_{it}^j + f_i + \varepsilon_{it} + \omega_{it} \\ \varepsilon_{it} &= \varepsilon_{it-1} + \zeta_{it}\end{aligned}$$

where ω_{it} is an i.i.d. measurement error.

To eliminate the effect of unobserved heterogeneity we first transform the model in differences:

$$\begin{aligned}\Delta^s \ln w_{it} &= \Delta^s X_{it}'\beta + \sum_{j=1}^{2} \rho_j \Delta^s L_{it}^j + \Delta^s \varepsilon_{it} + \Delta^s \omega_{it} \\ &= \Delta^s X_{it}'\beta + \sum_{j=1}^{2} \rho_j \Delta^s L_{it}^j + \sum_{j=0}^{s-1} \zeta_{it-j} + \Delta^s \omega_{it} \qquad (9)\end{aligned}$$

We consider generic s-period wage growth ($s \geq 1$) to account for the fact that some of our data are biannual (after 1997) and because we do not necessarily observe individuals working in consecutive calendar years. The model for wage growth (9) removes the endogeneity bias induced by correlation between L_{it}^j and unobserved heterogeneity f_i. However, since wage growth is only observed for people working in both t and $t-s$, we still have selection:[10]

$$\begin{aligned}E\left(\Delta^s \ln w_{it} | P_{it} = P_{it-s} = 1\right) &= \Delta^s X_{it}'\beta + \sum_{j=1}^{2} \rho_j \Delta^s L_{it}^j + E\left(\sum_{j=0}^{s-1} \zeta_{it-j} | P_{it} = 1, P_{it-s} = 1\right) \\ &= \Delta^s X_{it}'\beta + \sum_{j=1}^{2} \rho_j \Delta^s L_{it}^j + E\left(\sum_{j=0}^{s-1} \zeta_{it-j} | \vartheta_{it} > -h_{it}, \vartheta_{it-s} > -h_{it-s}\right) \qquad (10)\end{aligned}$$

Assume, as in Low, Meghir and Pistaferri (2010) that $\begin{pmatrix} \vartheta_{it} & \vartheta_{it-s} \end{pmatrix}' \sim N(\mathbf{0}, \mathbf{I})$. Then we can write the conditional expectation reflecting selection as:

$$E\left(\sum_{j=0}^{s-1} \zeta_{it-j} | \vartheta_{it} > -h_{it}, \vartheta_{it-s} > -h_{it-s}\right) = \left(\sigma_\zeta \sum_{j=0}^{s-1} \rho_{\zeta_{t-j}\vartheta_t}\right) \lambda_{it} + \left(\sigma_\zeta \sum_{j=0}^{s-1} \rho_{\zeta_{t-j}\vartheta_{t-s}}\right) \lambda_{it-s}$$

[10] Assuming that there is no selection on the measurement error.

where $\lambda_{ik} = \frac{\phi(h_{ik})}{\Phi(h_{ik})}$ is the inverse Mills' ratio (which we estimate with $\widehat{\lambda}_{ik} = \frac{\phi\left(\widehat{h_{ik}}\right)}{\Phi\left(\widehat{h_{ik}}\right)}$, with $\widehat{h_{ik}}$ derived from the probit estimate of the first stage), $\sigma_\zeta^2 = var\left(\zeta_{it}\right)$, and $\rho_{\zeta_k \vartheta_l} = corr\left(\zeta_k, \vartheta_l\right)$. Hence (10) can be rewritten as:

$$E\left(\Delta^s \ln w_{it} | P_{it} = P_{it-s} = 1\right) = \Delta^s X'_{it}\beta + \sum_{j=1}^{2} \rho_j \Delta^s L^j_{it} + \left(\sigma_\zeta \sum_{j=0}^{s-1} \rho_{\zeta_{t-j}\vartheta_t}\right)\lambda_{it} + \left(\sigma_\zeta \sum_{j=0}^{s-1} \rho_{\zeta_{t-j}\vartheta_{t-s}}\right)\lambda_{it-s} \tag{11}$$

Consistent estimation of the wage growth parameters can be obtained by using only the sample of participants in both periods t and $t-s$ and regressing wage growth against $\Delta^s X'_{it}$, $\Delta^s L^1_{it}$, $\Delta^s L^2_{it}$, and interaction of the inverse Mills ratios for the two periods with dummies for s (the gap between observed annual wages). The full set of results are reported in Table 3.

It is useful to discuss the type of biases induced by failing to correct for selection into work and that of ignoring fixed unobserved heterogeneity correlated with disability. Assume that the estimates in column 5 of Table 3, which control for both sample selection and eliminate fixed effects, are unbiased. Consider the effect of L^2_t on wages. The difference between the estimates reported in column (4) and those reported in column (5) is, approximately, a measure of the "sample selection bias". This difference is 0.18, implying a positive correlation between unobserved preferences for work and wage shocks (which is actually confirmed by the formal estimates of $\rho_{\zeta\vartheta}$, reported below). The "endogeneity bias" (neglecting fixed effects correlated with poor health) can be recovered as (approximately) the difference between the estimates in column (3) and those in column (5), i.e., -0.11. As expected, people with high unobserved heterogeneity in wages (say, ability) are less likely to develop disabilities. The total bias should then be the sum of the two, i.e., -0.07. This is indeed very close to the difference between the estimates of column (2) (where both biases are present) and column (5) (where both biases have been eliminated under the maintained hypotheses), which is ≈ -0.04.

C.3 The variance of productivity shocks

The estimates of productivity risk are obtained using the fact that our model provides us with the following restrictions on first and second order conditional moments of wage growth, as well as first order (unconditional) autocovariance:

Table 3: **Estimates of the wage equation.**

	(1) Part.	(2) Levels	(3) Levels	(4) Diff.	(5) Diff.
L_t^2	-0.744***	-0.214***	-0.295***	0.010	-0.177**
	(0.106)	(0.036)	(0.106)	(0.026)	(0.080)
L_t^1	-0.270***	-0.146***	-0.169***	-0.002	-0.057**
	(0.118)	(0.028)	(0.041)	(0.013)	(0.025)
Age	0.010***	0.057***	0.059***	0.054***	0.052***
	(0.002)	(0.005)	(0.005)	(0.009)	(0.015)
Age sq.	-0.016***	-0.058***	-0.061***	-0.060***	-0.067***
	(0.002)	(0.006)	(0.007)	(0.007)	(0.008)
White	0.033***	0.222***	0.226***	-.-	-.-
	(0.006)	(0.017)	(0.018)		
Married	0.059***	0.122***	0.128***	0.001	0.011
	(0.009)	(0.017)	(0.019)	(0.012)	(0.013)
Year effects	YES	YES	YES	YES	YES
P-value excl. restr.	0.032				
P-value sel. corr.			0.433		0.000
Observations	22953	20611	20611	17771	17771

Note: Clustered standard errors in parenthesis. *,**,*** indicate statistical significance at 10, 5 and 1 percent level respectively

$$
\begin{aligned}
E\left(\Delta^{s}\left(\varepsilon_{it}+\omega_{it}\right)|\vartheta_{it}>-h_{it},\vartheta_{it-s}>-h_{it-s}\right) &= \sigma_{\zeta}\lambda_{it}\sum_{j=0}^{s-1}\rho_{\zeta_{t-j}\vartheta_{t}}+\sigma_{\zeta}\lambda_{it-s}\sum_{j=0}^{s-1}\rho_{\zeta_{t-j}\vartheta_{t-s}} \\
E\left(\Delta^{s}\left(\varepsilon_{it}+\omega_{it}\right)^{2}|\vartheta_{it}>-h_{it},\vartheta_{it-s}>-h_{it-s}\right) &= \sigma_{\zeta}^{2}\left(s-h_{it}\lambda_{it}\sum_{j=0}^{s-1}\rho_{\zeta_{t-j}\vartheta_{t}}^{2}-h_{it-s}\lambda_{it-s}\sum_{j=0}^{s-1}\rho_{\zeta_{t-j}\vartheta_{t-s}}^{2}\right) \\
&\quad +2\sigma_{\omega}^{2} \\
E\left(\Delta^{s}\left(\varepsilon_{it}+\omega_{it}\right)\Delta^{l}\left(\varepsilon_{it-s}+\omega_{it-s}\right)\right) &= -\sigma_{\omega}^{2}
\end{aligned}
$$

We use these three moment conditions in a GMM framework to identify the parameters of interest (primarily, the variance of productivity shocks σ_{ζ}^{2}).

The full set of results are reported in Table 4. We have already commented about the variances in the main text. The correlation coefficients suggest that permanent innovations to productivity are correlated with current and 1-period lagged unobservable shocks to propensity to work. There is no correlation with future or 2-period lagged unobservable propensities to work. The sign and pattern of the correlations are consistent with unobserved shock to the propensity to work being correlated with the current value of permanent productivity alone.

Table 4: **Estimates of the variance of wage innovations.**

Parameter	Wage Process
σ_{ζ}^{2}	0.027^{***} (0.002)
σ_{ω}^{2}	0.044^{***} (0.002)
$\rho_{\zeta_t\vartheta_t}$	0.901^{***} (0.120)
$\rho_{\zeta_t\vartheta_{t-1}}$	-0.936^{***} (0.152)
$\rho_{\zeta_{t-1}\vartheta_t}$	-0.030 (0.243)
$\rho_{\zeta_t\vartheta_{t-2}}$	-0.016 (0.271)

Note: Clustered standard errors in parenthesis. *,**,*** indicate statistical significance at 10, 5 and 1 percent level respectively

C.4 Health Transitions

In this section we present estimates of $\Pr(L_{it} = j|L_{it-1} = k, f_q)$, for $j, k = \{0, 1, 2\}$ and $q = \{L, M, H\}$. For reason of space, only selected transition probabilities were included in the main text. Figure 7 reports transitions out of the "no disability" state, $L_{it-1} = 0$, for the three heterogeneity types. In general disability risk increases over the life cycle (the probability of remaining healthy goes down, that of transitioning into moderate or severe disability goes up with age), and all the trends are more accentuated for the lower ability types.

Transitions out of the moderate disability state $L_{it-1} = 1$ also have interesting life cycle patterns, as shown in Figure 8. Persistence in that state increases for the high ability types, but declines for the other two types. "Recoveries" become less likely with age, and uniformly less so for the low types. Finally, transitions into severe disability are rather flat among the high/medium ability types, but more than double for the low ability types.

Figure 9 reports transitions out of severe disability. Here, the broad picture is very stark. The probability of complete or mild recovery declines strongly for the low ability types, but only slightly for the other ability types. All types face an increase in persistence in the severe disability state, but this is again higher in the low ability group.

C.5 Indirect Inference

The Indirect Inference statistical criterion that we use is:[11]

$$\hat{\phi} = \arg\min_{\phi} \left(\widehat{\alpha}^D - S^{-1} \sum_{s=1}^{S} \widehat{\alpha}^S(\phi) \right)' \Omega \left(\widehat{\alpha}^D - S^{-1} \sum_{s=1}^{S} \widehat{\alpha}^S(\phi) \right)$$

where $\widehat{\alpha}^D$ are the moments/parameters of the auxiliary model estimated in the data, $\widehat{\alpha}^S(\phi)$ are the corresponding simulated moments/parameters (which we average over S simulations) for given structural parameter values ϕ. The function $\alpha(\phi)$ is the binding function relating the structural parameters to the auxiliary parameters, and Ω is the weighting matrix. The optimal weighting matrix is the the inverse of the covariance matrix from the data, $\widehat{\Omega} = var\left(\widehat{\alpha}^D\right)^{-1}$.

Standard errors of the structural parameters can be computed using the formula provided in Gourieroux et al. (1993),

[11] Indirect Inference is a generalization of the more traditional method of simulated moments, MSM, or the Efficient Method of Moments, EMM. Indirect Inference is becoming a standard estimation method in analyses of the type we conduct in our papers. See for recent examples Alan and Browning (2009); Guvenen and Smith (2014); Altonji, Smith and Vidangos (2013).

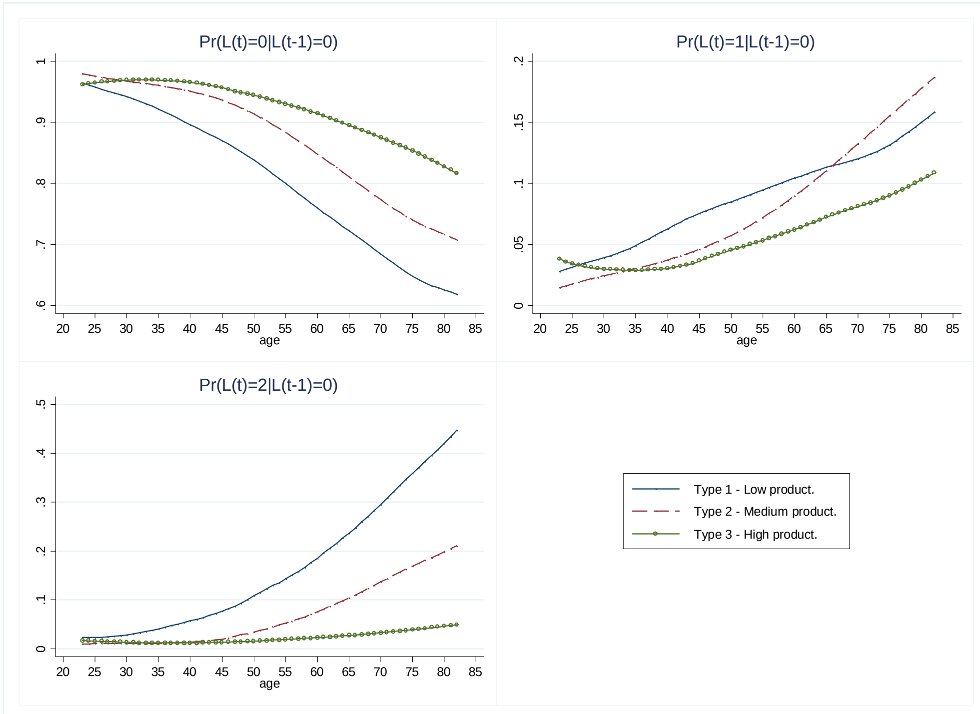

Figure 7: Transitions out of $L = 0$ (Sample: Low education males, aged 23-62).

$$var\left(\widehat{\phi}\right) = \left(J'\mathbf{\Omega}J\right)^{-1} J'\mathbf{\Omega}\mathbf{V}\mathbf{\Omega}J\left(J'\mathbf{\Omega}J\right)^{-1}$$

where $J = \frac{\partial\widehat{\alpha}^S(\phi)}{\partial\phi}$, and $V = var\left(\widehat{\alpha}^D - \widehat{\alpha}^S\left(\widehat{\phi}\right)\right)$. Asymptotically, V reduces to $\left(1 + \frac{1}{S}\right) var\left(\widehat{\alpha}^D\right)$, and we obtain $var\left(\widehat{\alpha}^D\right)$ using a block bootstrap procedure. We check that the average of the variance of the simulated moments across the S replications is well approximated by the variance of the moments in the data. We calculate J by finite difference.

Figure 10 gives a graphical account of the match between auxiliary moments/parameters in the data and in the simulations. Cases in which the simulated value is outside the 95% confidence interval are cases in which the gray (darker) bar is outside the vertical orange (lighter) line.

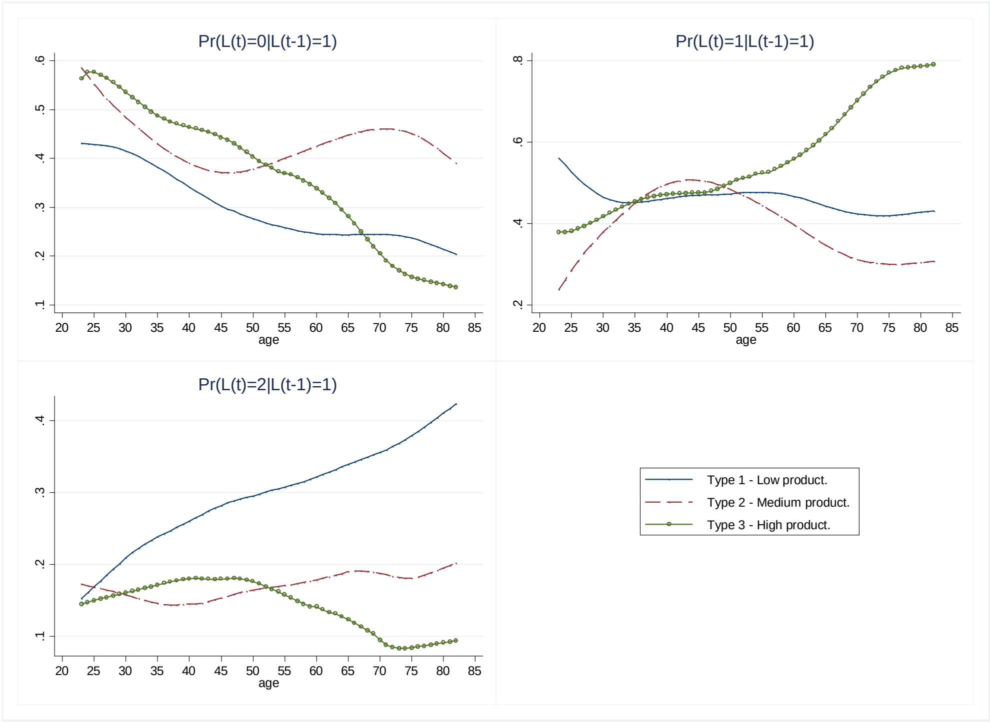

Figure 8: Transitions out of $L = 1$ (Sample: Low education males, aged 23-62).

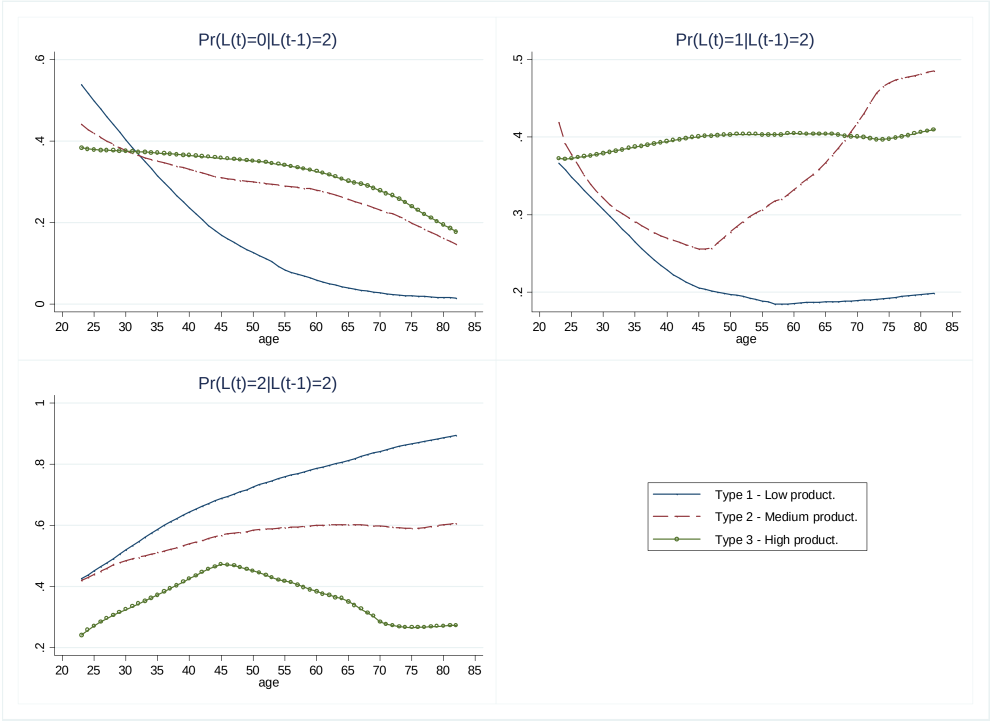

Figure 9: Transitions out of $L = 2$ (Sample: Low education males, aged 23-62).

Figure 10: The match between auxiliary moments/parameters in the data and in the simulations.

D Details of the Robustness of the Estimation

In this section we describe the results of a series of sensitivity analysis on the estimation process.

D.1 Alternative exclusion restrictions

Here we report results under a variety of different exclusion restrictions. See Table 5.

In the first row we report the results using the baseline set of exclusion restrictions (potential benefits from all programs: Food Stamps, EITC, AFDC/TANF, UI). Set 2 uses only potential welfare benefits (Food Stamps, EITC, AFDC/TANF). Set 3 uses only potential UI benefits. We find that the use of these different exclusion restrictions induce only small changes in the effect of disability variables on wages and virtually no effects on the estimated variances of the permanent shock (the only structural parameters in this exercise).

Table 5: **Sensitivity Analysis on Exclusion Restrictions**

Exclusion restrictions	$\{L = 2\}$ on wage	$\{L = 1\}$ on wage	σ^2_ζ
Baseline - Set 1	-0.177^{**}	-0.057^{**}	0.027^{***}
	(0.080)	(0.025)	(0.002)
Set 2	-0.147^{*}	-0.048^{**}	0.026^{***}
	(0.077)	(0.024)	(0.002)
Set 3	-0.165^{**}	-0.053^{**}	0.026^{***}
	(0.083)	(0.025)	(0.002)

Note. Set 1: Baseline exclusion restrictions (potential benefits from all programs: Food Stamps, EITC, AFDC/TANF, UI). Set 2: Potential welfare benefits (Food Stamps, EITC, AFDC/TANF). Set 3 uses only potential UI benefits. Clustered standard errors in parenthesis.

D.2 The normality assumption in the Selection Equation

We checked that our results do not depend on the normality assumption that we make when estimating the wage equation controlling for selection into work. We repeat our wage equation estimates using a non-parametric approach.

In the first step we estimate the univariate employment model

$$P^*_{it} = h_{it} + \vartheta_{it}$$

using the semi-nonparametric estimator of Gallant and Nychka (1987), which assumes that the unknown density of the latent regression error ϑ_{it} can be approximated by a Hermite polynomial expansion. The estimation procedure gives us the predicted value $\widetilde{h_{it}}$. For the second step, Newey (2009) suggests running the equivalent of (11) but replacing the inverse Mills ratio λ_{it} with a polynomial in functions of $\widetilde{h_{it}}$. He suggests using $\widetilde{h_{it}}$ itself, $\Phi\left(\widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{h_{it}}\right)$, and $\lambda\left(\widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{h_{it}}\right) = \frac{\phi\left(\widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{h_{it}}\right)}{\Phi\left(\widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{h_{it}}\right)}$, where $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ are probit estimates from a regression of participation on a constant and $\widetilde{h_{it}}$. We follow these suggestions and use separate 2^{nd} degree polynomials. We find that the results remain very similar (see Table 6).

Table 6: **Sensitivity Analysis on selection correction**

	(1) FD	(2) FD	(3) FD	(4) FD
Selection correction terms	$\lambda\left(\widehat{h_{it}}\right)$	$\Phi\left(\widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{h_{it}}\right)$	$\lambda\left(\widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{h_{it}}\right)$	$\widetilde{h_{it}}$
L=2	-0.177**	-0.206**	-0.220**	-0.193*
	(0.080)	(0.086)	(0.086)	(0.147)
L=1	-0.057**	-0.060*	-0.063***	-0.082
	(0.025)	(0.023)	(0.023)	(0.062)
Age	0.052***	0.045***	0.035	0.081*
	(0.015)	(0.011)	(0.090)	(0.047)
Age sq.	-0.067***	-0.073***	-0.072***	-0.072***
	(0.008)	(0.011)	(0.009)	(0.013)
P-value excl. restr.	0.032	0.015	0.015	0.015
P-value sel. corr.	0.000	0.073	0.043	0.448
Observations	17771	17771	17771	17771

Note: First we fit univariate participation model using the semi-nonparametric estimator of Gallant and Nychka (1987). Then use the appropriate selection correction term. Clustered standard errors in parenthesis.

D.3 Restricting the Sample

Our sample includes both the SRC and SEO components of the PSID. The SRC was the component of the PSID that was initially representative of the population, while the SEO oversampled individuals close to the poverty line in 1967. Given our focus on the low educated subsample, we have kept in our empirical analysis both the SRC and SEO components of the PSID to have a larger sample. In Table 7 we show results obtained using only the SRC sample. The results are qualitatively similar, but the estimates (especially the effect of severe disability on wages) are less precise as expected.

Table 7: **Sensitivity Analysis on sample selection**

Sample	$\{L=2\}$ on wage	$\{L=1\}$ on wage	σ^2_ζ	N
Baseline	-0.177^{**} (0.080)	-0.057^{**} (0.025)	0.027^{***} (0.002)	17771
Only SRC	-0.149 (0.102)	-0.042^{*} (0.025)	0.030^{***} (0.002)	10385

Note: Clustered standard errors in parenthesis.

D.4 The Selection Process in the Model

To reduce the computational burden, we estimate the wage process outside the model. In particular, we estimate the wage process controlling for selection into work and unobserved heterogeneity and then use this as an input into the simulated model. It is useful to check whether the model can replicate the estimates we get if we try to mimic in the simulated data the empirical strategy we adopt in the actual data. To do so, we run a probit regression for employment in the simulated data and construct the corresponding Mills' ratio. We then estimate the wage growth equation with and without the Mills' ratio as a control (in addition to age and the disability indicators). Columns (1)-(3) of Table 8 reproduce the estimates of the relevant parameters in the data, together with 95% confidence intervals, and the estimates obtained in the simulated data. In no case do we find that the estimates obtained using simulated data fall outside the 95% confidence interval of the data estimates. In fact, the estimates are numerically very similar. The effect of severe disability on wages is lower in the simulated data, but the procedures are not exactly identical because we do not have exclusion restrictions in the simulated data. In column (4) we report estimates from the data obtained omitting the exclusion restrictions, and this shows an even closer similarity between data and simulation estimates.

D.5 Assets and Life-cycle Profiles

In section 5.6.4 in the main text, we discuss the comparison between simulated life-cycle profiles and profiles in the data. Here we plot the evolution of average wealth over the life cycle for the

Table 8: **The Log Wage Equation in the Model and in the Data**

		Employment Probit (1)	Wage Growth (Observed) (2)	Wage Growth (Corrected) (3)	Wage Growth (Corr., no excl. restr.) (4)
$L=2$	Data	-0.744^{***}	0.010	-0.177^{**}	-0.141^{**}
	95% C.I.	$(-0.952, -0.536)$	$(-0.042, 0.061)$	$(-0.334, -0.021)$	$(-0.304, 0.023)$
	Simulations	-0.673	-0.027	-0.113	-0.113
$L=1$	Data	-0.270^{***}	-0.002	-0.057^{**}	-0.046^{**}
	95% C.I.	$(-0.500, -0.039)$	$(-0.028, 0.024)$	$(-0.105, -0.008)$	$(-0.096, 0.005)$
	Simulations	-0.336	-0.018	-0.055	-0.055

Note: Clustered 95 percent confidence intervals in parenthesis. *,**,*** = significant at 10, 5, and 1 percent, respectively. Probit estimates are marginal effects.

three health groups, although the composition of these groups changes with age.[12] We find that the model approximates well the asset profiles we observe in the data. For the group who are not disabled, the simulated 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.

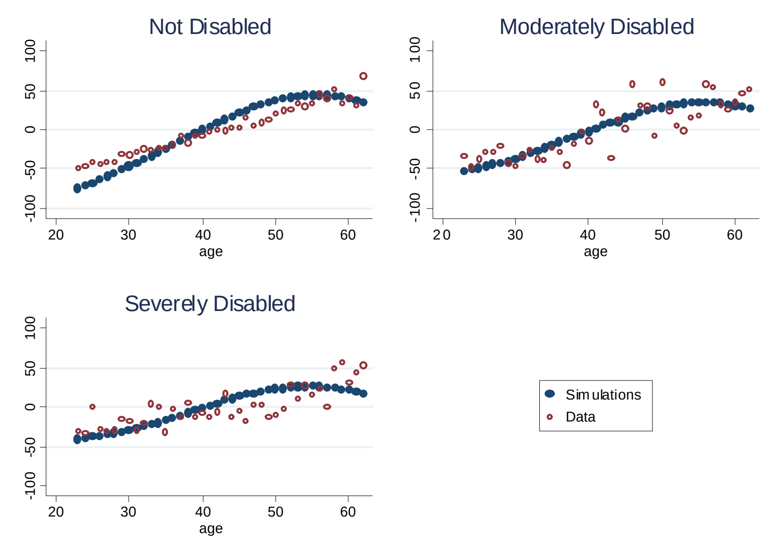

Figure 11: Weath over the life cycle.

[12]These pictures are obtained as deviation from life cycle means and winsorizing the data at the bottom and top 2.5% of the distribution to reduce the impact of outliers.

D.6 The Importance of Modelling Work Limitations

In section 5.6.5 in the main text, we discussed the importance of allowing the fixed cost of work to vary with work limitation status, and the importance of allowing for non-separability between consumption and health. Each of these components substantially improves our ability to match the moments of disability insurance take-up, participation and the consumption regressions. Table 5 in the text reports the standard errors on these parameter estimates: the statistical significance of θ and the significant difference in the estimates of L by work limitation status shows that these are statistically necessary to explain the moments. To show this more explicitly, we reestimate the model twice, first forcing the fixed cost to be the same regardless of work limitation status and secondly, switching off the non-separability and so imposing $\theta = 0$.

When we force the fixed cost to be independent of L, the estimation is not able to get anywhere near to the moments, particularly because of the differences in participation rates by work limitation status. When, instead, we set $\theta = 0$, we are able to get reasonably close except in the consumption regression, where work limitation status has a much larger effect on consumption when $\theta = 0$. This is because when $\theta < 0$, the non-separability increases the marginal utility of consumption in the work-limited state and shifts resources into this state. This means that simulated consumption when $\theta < 0$ is higher and closer to the data. Results of the moments when $\theta = 0$ are reported in Table 9, with corresponding parameter estimates in Table 10.

Table 9: **Targeted Moments**

Variable/Moment	Data	Simulations: Baseline	Simulations: $\theta = 0$	95% C.I.
Panel A: The Log Consumption Regression				
$\{L_{it} = 1\}$	-0.051 (0.035)	-0.091	-0.183	(0.06,0.20)
$\{L_{it} = 2\}$	-0.162 (0.070)	-0.172	-0.348	(0.05,0.33)
$\{L_{it} = 1\} \times DI_{it}$	0.177 (0.134)	0.154	0.178	(-0.25,0.25)*
$\{L_{it} = 2\} \times DI_{it}$	0.260 (0.148)	0.374	0.393	(-0.41,0.15)*
DI_{it}	-0.105 (0.123)	-0.336	-0.362	(0.02,0.50)
Employed	0.390 (0.055)	0.242	0.247	(0.04,0.25)
Panel B: Employment by Disability Status				
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 0,\ t \leq 45\}$	0.927 (0.0034)	0.917	0.914	(0.01,0.02)
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 0,\ t > 45\}$	0.868 (0.0074)	0.914	0.911	(-0.06,-0.03)
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 1,\ t \leq 45\}$	0.701 (0.0217)	0.683	0.664	(-0.01,0.08)*
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 1,\ t > 45\}$	0.499 (0.0277)	0.516	0.499	(-0.05,0.05)*
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 2,\ t \leq 45\}$	0.161 (0.0185)	0.169	0.170	(-0.04,0.02)*
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 2,\ t > 45\}$	0.077 (0.0125)	0.0872	0.089	(-0.04,0.01)*
Panel C: DI Coverage				
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 2,\ t \leq 45\}$	0.308 (0.032)	0.298	0.308	(-0.06,0.06)*
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 2,\ t > 45\}$	0.552 (0.030)	0.544	0.523	(-0.03,0.08)*
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 1,\ t \leq 45\}$	0.081 (0.014)	0.091	0.113	(-0.06,-0.00)
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 1,\ t > 45\}$	0.187 (0.021)	0.182	0.178	(-0.03,0.05)*
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 0,\ t \leq 45\}$	0.003 (0.001)	0.003	0.003	(-0.00,0.00)*
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 0,\ t > 45\}$	0.016 (0.003)	0.014	0.014	(-0.00,0.01)*

Variable/Moment	Data	Simulations: Baseline	Simulations: $\theta = 0$	95% C.I.
Panel D: Composition of DI Recipients				
$\text{Fr}\{L_{it} = 2 \mid DI_{it} = 1,\ t \leq 45\}$	0.638 (0.0390)	0.605	0.571	(-0.02,0.15)*
$\text{Fr}\{L_{it} = 2 \mid DI_{it} = 1,\ t > 45\}$	0.678 (0.0224)	0.691	0.687	(-0.05,0.03)*
$\text{Fr}\{L_{it} = 1 \mid DI_{it} = 1,\ t \leq 45\}$	0.243 (0.0314)	0.266	0.300	(-0.12,0.01)*
$\text{Fr}\{L_{it} = 1 \mid DI_{it} = 1,\ t > 45\}$	0.209 (0.0189)	0.220	0.223	(-0.05,0.02)*
$\text{Fr}\{L_{it} = 0 \mid DI_{it} = 1,\ t \leq 45\}$	0.120 (0.0228)	0.128	0.129	(-0.06,0.04)*
$\text{Fr}\{L_{it} = 0 \mid DI_{it} = 1,\ t > 45\}$	0.113 (0.0147)	0.090	0.090	(-0.01,0.05)*
Panel E: Flows into DI				
$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 2,\ t \leq 45\}$	0.168 (0.022)	0.158	0.161	(-0.03,0.05)*
$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 2,\ t > 45\}$	0.279 (0.024)	0.283	0.265	(-0.03,0.06)*
$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 1,\ t \leq 45\}$	0.039 (0.008)	0.030	0.041	(-0.02,0.01)*
$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 1,\ t > 45\}$	0.067 (0.011)	0.043	0.042	(0.00,0.05)
$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 0,\ t \leq 45\}$	0.001 (0.0003)	0.0005	0.001	(-0.00,0.00)*
$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 0,\ t > 45\}$	0.007 (0.001)	0.002	0.002	(0.00,0.01)

Note: The confidence interval is computed with the block bootstrap. An asterisks denotes that the difference is statistically insignificant (5% level).

Table 10: **Estimated Parameters when $\theta = 0.0$**

Frictions and Preferences			Disability Insurance Program		
Parameter	Estimate		Parameter	Estimate	
	Baseline	$\theta = 0$		Baseline	$\theta = 0$
θ	-0.448^{***} (0.126)	0.00	$\pi_{L=0}^{Young}$	0.006 (0.964)	0.002
η	-0.185^{*} (0.160)	-0.22	$\pi_{L=0}^{Old}$	0.075 (0.800)	0.028
δ	0.0624^{***} (0.002)	0.0647	$\pi_{L=1}^{Young}$	0.171^{***} (0.025)	0.207
			$\pi_{L=1}^{Old}$	0.180^{***} (0.032)	0.171
$F_{L=0}$	0.000 [$0] (0.371)	0.000	$\pi_{L=2}^{Young}$	0.331^{***} (0.031)	0.337
$F_{L=1}$	0.547^{***} [$2472] (0.111)	0.551	$\pi_{L=2}^{Old}$	0.626^{***} (0.046)	0.607
$F_{L=2}$	0.952^{***} [$4301] (0.109)	0.951			

Note: 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 parenthesis. *,**,*** = significant at 10, 5, and 1 percent, respectively.

E Robustness of the Policy Counterfactuals

E.1 Welfare Calculations

To define the welfare cost or benefit of reforming DI, we write the lifetime expected utility of an individual as

$$E_0 U(k) = E_0 \sum_{t=0}^{T} \beta^t \frac{\left(c_{it}(k) \exp(\theta L_{it}) \exp(\eta P_{it}(k))\right)^{1-\gamma}}{1-\gamma}$$

where k refers to the implied consumption and labor supply stream in the baseline economy ($k = b$) or an alternative economy with different parameters of the DI program ($k = b'$) and E_0 is the expectation at the beginning of working life. Now define π as the proportion of consumption an individual is willing to pay to be indifferent between environment $k = b'$ and $k = b$. This is implicitly defined by

$$\begin{aligned} E_0 U\left(b'\right)|_{\pi} &\equiv E_0 \sum_{t=0}^{T} \beta^t \frac{\left(\left(1-\pi\right) c_{it}(b') \exp(\theta L_{it}) \exp(\eta P_{it}(b'))\right)^{1-\gamma}}{1-\gamma} \\ &= \sum_{t=0}^{T} \beta^t \frac{\left(c_{it}(b) \exp(\theta L_{it}) \exp(\eta P_{it}(b))\right)^{1-\gamma}}{1-\gamma} \\ &\equiv E_0 U(b) \end{aligned}$$

We compute π in the different scenarios and use it as our measure of welfare in Section 6.[13]

[13] Kitao (2014) builds on our paper to show how allowing for Medicare affects the incentive effects and insurance value of disability insurance.

E.2 Reassessment of DI Recipients

The main text discusses that changing reassessment rates has limited effects. This section provides figure 12 to back up that result. The left-hand graph shows that an increase in the reassessment rate discourages false applications by those who are not severely disabled. The cost of this is the reduced 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 labour force participation and output, and increased asset accumulation as individuals have to self-insure further, as shown on the right hand side graph. The net effect on welfare of increasing reassessments is negative, but negligible.

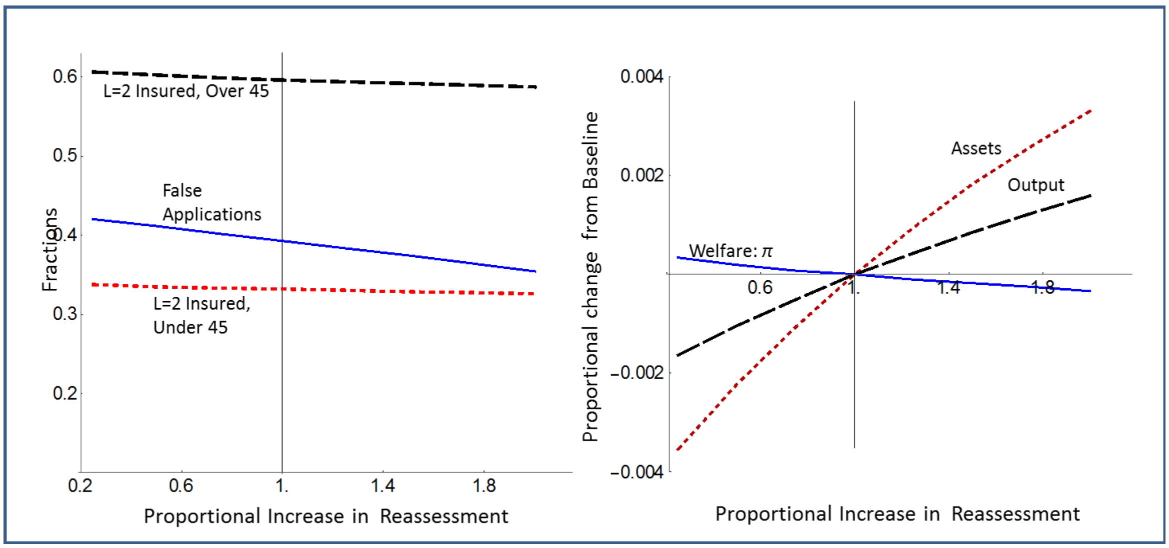

Figure 12: Increasing the Reassessment Rate

E.3 The Beta Distribution

In the baseline presented in the main paper, we assume that the signal that the government observes about an individual's health status has a Beta distribution. The Beta distribution is a continuous distribution defined over $x \in [0, 1]$, with the distribution defined by two parameters, q and r. The formula for the PDF is given by:

$$f(x) = \frac{x^{q-1}(1-x)^{r-1}}{B(q,r)}, \qquad q > 1, r > 1$$

where $B(q, r)$ is a normalization to keep the PDF between 0 and 1. This normalization is

$$\frac{\Gamma(q+r)}{\Gamma(q)\,\Gamma(r)},$$

where Γ is the gamma distribution.

We impose that r is held constant across all types, but that q varies with health and with age. When q and r are equal, the distribution is symmetric around $x = 0.5$. As q increases for a given value of r, the mean of the distribution shifts to the right, and the variance declines. The level of the variance is determined by the absolute size of q and r. We make the assumption that the signal follows a Beta distribution because this implies that the precision of the signal increases as the severity of the true health limitation increases.

To parameterize the Beta distribution and how it changes with health status and with age, we need 6 values of q and 1 value of r. To characterize the government's decision process, we also need the value of the threshold acceptance onto DI, $\bar{S}$. These 8 parameters are set by the 6 probabilities of acceptance onto DI estimated in the paper, as well as 2 normalizations. The normalizations are to the mean of the signal for the young, healthy individuals, and the mean of the signal for the old, severely disabled individuals (i.e., those with the lowest and highest probability of success in the data). The resulting distribution of the signal is shown in figure 4 in the paper.

E.4 Strictness with lognormally distributed signals

There are alternative assumptions that could be made about the distribution of the signal. For example, the shock could have lognormal distribution. We illustrate this case here. We assume that S lies on the real line and is distributed:

$$\ln S_{L,t} \sim N\left(\mu_{L,t}, \sigma\right)$$

where μ varies with age and work limitation status. The six values of μ, the value of σ and $\overline{S}$ are pinned down by the six structural probabilities (π_L^t) estimated.[14] The resulting values are mapped into PDFs of the conditional distribution as shown in Figure 13. The fraction of false positives and false negatives, as well as the probabilities of success at $\bar{S}$ are held constant with the alternative distribution. The shape of the distribution and the renormalized value of $\bar{S}$ changes.

As in Section 6, we then consider changing the strictness of the regime. We show the implications for coverage, false applications and welfare in figure 14. As strictness increases, the fraction insured

[14] We normalise the mean of the log of the signal, S, for the old who are severely disabled to be 1, and and the mean of the log of S for the young who are not at all disabled to being 0.1.

falls, the number of false applications falls, and assets and output both rise. The net effect of greater strictness on welfare is, however, negative. These are very similar results to the Beta distribution case in the direction of changes, magnitudes and in the overall effect on welfare.

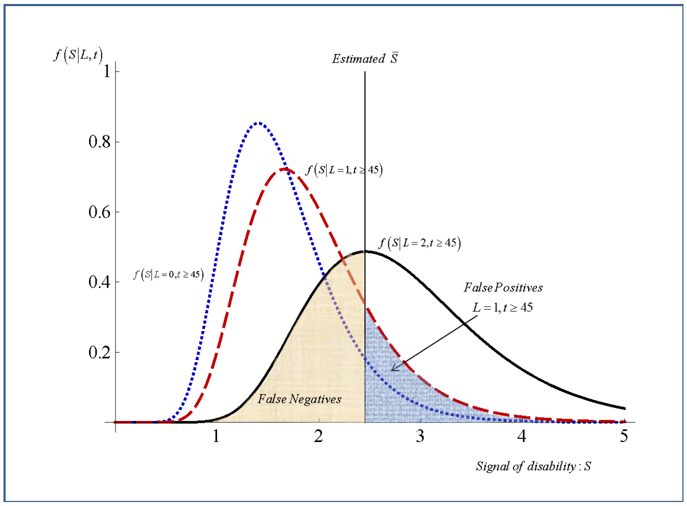

Figure 13: The Distribution of S with Log Normal Signals

E.5 Sensitivity Analysis: The Importance of Risk Aversion

The value for the coefficient of relative risk aversion is taken as the central estimate from various papers in the literature, $\gamma = 1.5$ (see footnote 23 in the main text). Clearly, how risk averse agents are will have an impact on how they value the insurance benefit of disability insurance. Risk aversion may also alter how individuals respond to policy changes. To explore this possibility and to show the sensitivity of our baseline results to the degree of risk aversion, we reestimated and resimulated all the results assuming a coefficient of relative risk aversion equal to 3.

We first reestimated the structural parameter, using the same set of moments from the data as in the main paper. This resulted in the match shown in Table 11, with corresponding parameter estimates in Table 12.

We then re-did the three core policy counterfactuals using this new set of parameter estimates and the higher value of risk aversion. First, we changed the generosity of DI payments. Figure 15 corresponds to Figure 3 in the main text. The left hand side shows how the fraction insured and the fraction of false applicants increases as generosity increases. The right hand side shows the effects on output, asset accumulation and welfare. The sensitivity of the behavior of those with

Table 11: **Targeted Moments when $\gamma = 3.0$**

Variable/Moment	Data	Simulations: Baseline	Simulations: $\gamma = 3$	95% C.I. diff.	Variable/Moment	Data	Simulations: Baseline	Simulations: $\gamma = 3$	95% C.I. diff.
Panel A: The Log Consumption Regression					**Panel D: Composition of DI Recipients**				
$\{L_{it} = 1\}$	-0.051 (0.035)	-0.091	-0.102	(-0.02,0.12)*	$\text{Fr}\{L_{it} = 2 \mid DI_{it} = 1,\ t \leq 45\}$	0.638 (0.0390)	0.605	0.564	(-0.01,0.16)*
$\{L_{it} = 2\}$	-0.162 (0.070)	-0.172	-0.164	(-0.14,0.14)*	$\text{Fr}\{L_{it} = 2 \mid DI_{it} = 1,\ t > 45\}$	0.678 (0.0224)	0.691	0.694	(-0.06,0.03)*
$\{L_{it} = 1\} \times DI_{it}$	0.177 (0.134)	0.154	0.151	(-0.22,0.28)*	$\text{Fr}\{L_{it} = 1 \mid DI_{it} = 1,\ t \leq 45\}$	0.243 (0.0314)	0.266	0.306	(-0.13,0.00)*
$\{L_{it} = 2\} \times DI_{it}$	0.260 (0.148)	0.374	0.334	(-0.36,0.21)*	$\text{Fr}\{L_{it} = 1 \mid DI_{it} = 1,\ t > 45\}$	0.209 (0.0189)	0.220	0.217	(-0.05,0.03)*
DI_{it}	-0.105 (0.123)	-0.336	-0.350	(0.00,0.49)	$\text{Fr}\{L_{it} = 0 \mid DI_{it} = 1,\ t \leq 45\}$	0.120 (0.0228)	0.128	0.130	(-0.06,0.04)*
Employed	0.390 (0.055)	0.242	0.223	(0.06,0.27)	$\text{Fr}\{L_{it} = 0 \mid DI_{it} = 1,\ t > 45\}$	0.113 (0.0147)	0.090	0.089	(-0.01,0.06)*
Panel B: Employment by Disability Status					**Panel E: Flows into DI**				
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 0,\ t \leq 45\}$	0.927 (0.0034)	0.917	0.914	(0.01,0.02)	$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 2,\ t \leq 45\}$	0.168 (0.022)	0.158	0.148	(-0.02,0.06)*
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 0,\ t > 45\}$	0.868 (0.0074)	0.914	0.911	(-0.06,-0.03)	$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 2,\ t > 45\}$	0.279 (0.024)	0.283	0.272	(-0.04,0.05)*
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 1,\ t \leq 45\}$	0.701 (0.0217)	0.683	0.677	(-0.02,0.07)*	$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 1,\ t \leq 45\}$	0.039 (0.008)	0.030	0.040	(-0.02,0.01)*
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 1,\ t > 45\}$	0.499 (0.0277)	0.516	0.528	(-0.08,0.02)*	$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 1,\ t > 45\}$	0.067 (0.011)	0.043	0.039	(0.00,0.05)
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 2,\ t \leq 45\}$	0.161 (0.0185)	0.169	0.167	(-0.04,0.03)*	$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 0,\ t \leq 45\}$	0.001 (0.0003)	0.0005	0.0006	(-0.00,0.00)*
$\text{Fr}\{P_{it} = 1 \mid L_{it} = 2,\ t > 45\}$	0.077 (0.0125)	0.0872	0.097	(-0.04,0.00)*	$\text{Fr}\{DI_{it} = 1 \mid DI_{it-2} = 0, L_{it} = 0,\ t > 45\}$	0.007 (0.001)	0.002	0.002	(0.00,0.01)
Panel C: DI Coverage									
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 2,\ t \leq 45\}$	0.308 (0.032)	0.298	0.291	(-0.05,0.08)*					
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 2,\ t > 45\}$	0.552 (0.030)	0.544	0.532	(-0.04,0.08)*					
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 1,\ t \leq 45\}$	0.081 (0.014)	0.091	0.110	(-0.06,-0.00)					
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 1,\ t > 45\}$	0.187 (0.021)	0.182	0.175	(-0.03,0.05)*					
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 0,\ t \leq 45\}$	0.003 (0.001)	0.003	0.003	(-0.00,0.00)*					
$\text{Fr}\{DI_{it} = 1 \mid L_{it} = 0,\ t > 45\}$	0.016 (0.003)	0.014	0.014	(-0.00,0.01)*					

Note: The confidence interval is computed with the block bootstrap. An asterisks denotes that the difference is statistically insignificant (5% level).

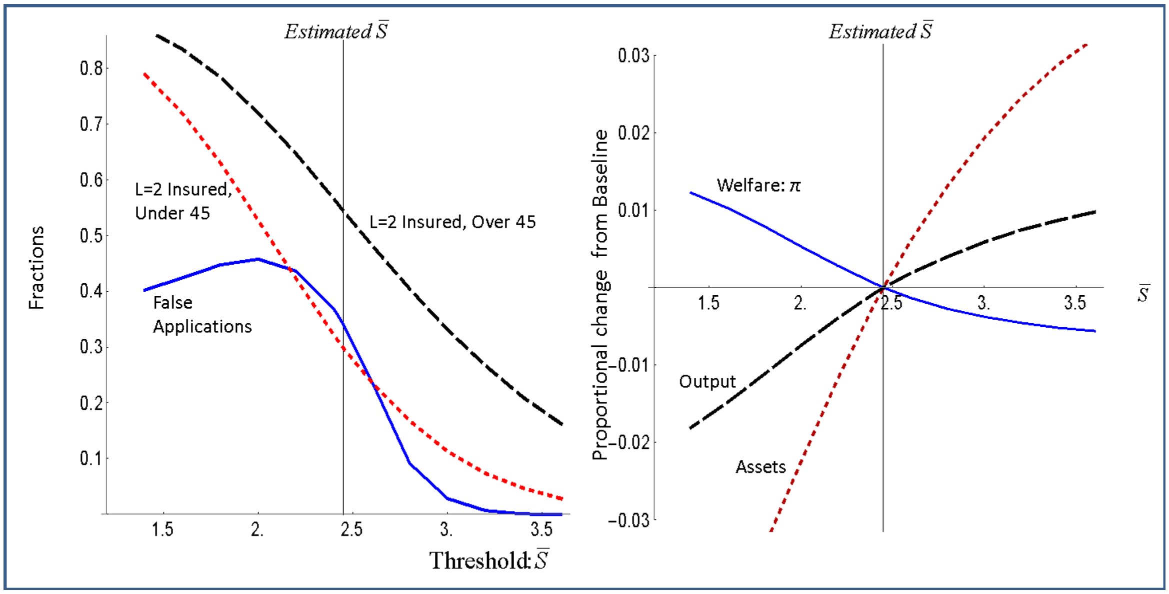

Figure 14: Increasing Strictness of the DI Regime with Lognormal Signals

Table 12: **Estimated Parameters when** $\gamma = 3.0$

Frictions and Preferences Parameter	Estimate		Disability Insurance Program Parameters	Estimate	
	Baseline	$\gamma = 3.0$		Baseline	$\gamma = 3.0$
θ	-0.448^{***} (0.126)	-0.165 (0.113)	$\pi^{Young}_{L=0}$	0.006 (0.964)	0.006 (598.8)
η	-0.185 (0.160)	-0.060 (0.096)	$\pi^{Old}_{L=0}$	0.075 (0.800)	0.100 (1.98)
δ	0.062^{***} (0.002)	0.065^{***} (0.002)	$\pi^{Young}_{L=1}$	0.171^{***} (0.025)	0.224^{***} (0.050)
			$\pi^{Old}_{L=1}$	0.180^{***} (0.032)	0.197^{***} (0.061)
$F_{L=0}$	0.000 [$0] (0.371)	0.000 [$0] (0.458)	$\pi^{Young}_{L=2}$	0.331^{***} (0.031)	0.354^{***} (0.041)
$F_{L=1}$	0.547^{***} [$2472] (0.111)	0.620^{***} [$2802] (0.070)	$\pi^{Old}_{L=2}$	0.626^{***} (0.046)	0.649^{***} (0.042)
$F_{L=2}$	0.952^{***} [$4301] (0.109)	1.052^{***} [$4754] (0.113)			

Note: 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 parenthesis (see the Appendix for definitions). *,**,*** = significant at 10, 5, and 1 percent, respectively.

only a moderate limitation is again evident from the False Application line. As in the baseline, welfare is increasing in the generosity of the program but more sharply now because the value of insurance is higher. Figure 16 shows the effect of changing the strictness threshold of DI, $\bar{S}$. The results are qualitatively the same as in the main paper. Again, the difference is in the magnitude of the welfare gains: reducing strictness has much larger welfare gains because of the higher value of the insurance. Figure 17 reports the details of increasing generosity of the food stamp program for the higher risk aversion. Welfare rises with the generosity of food stamps, and again this is much more marked because of the higher risk aversion.

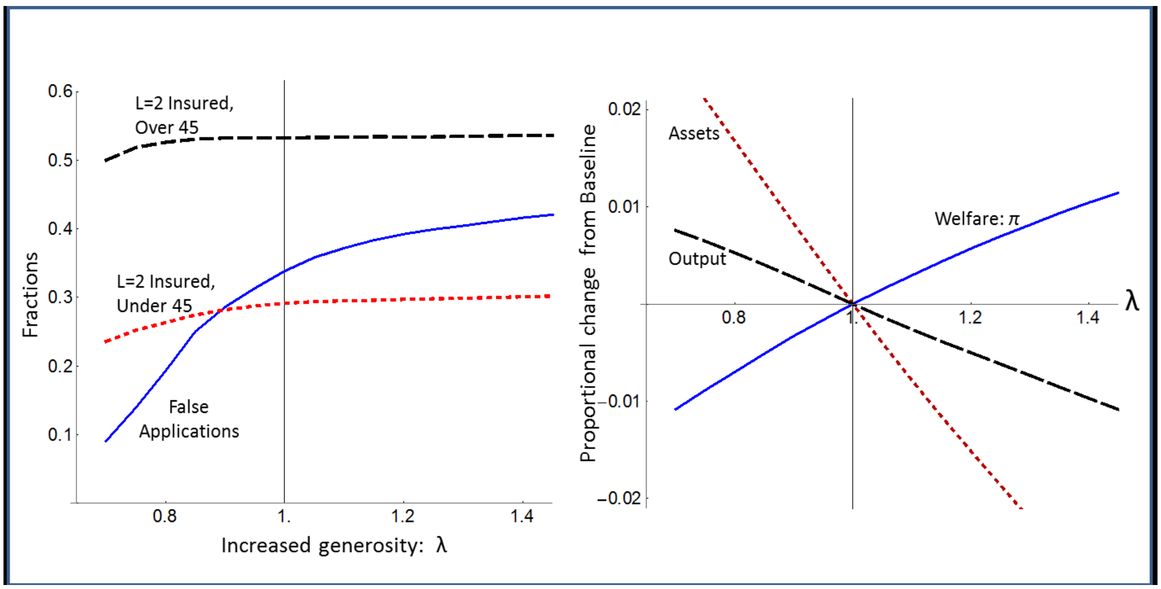

Figure 15: Increasing Generosity of DI when $\gamma = 3.0$

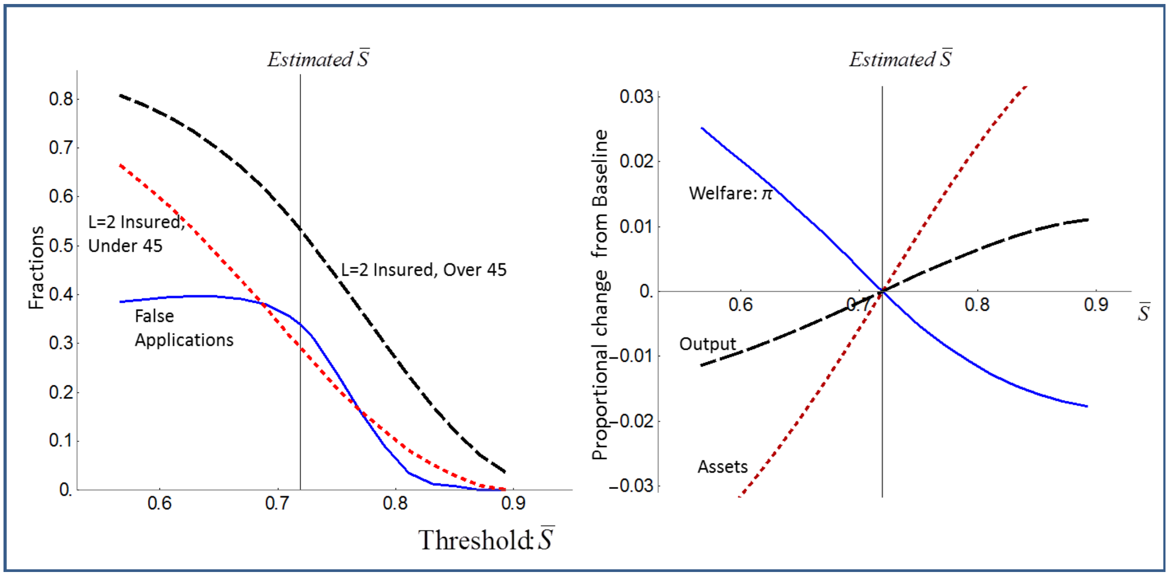

Figure 16: Increasing Strictness of the DI Regime when $\gamma = 3.0$

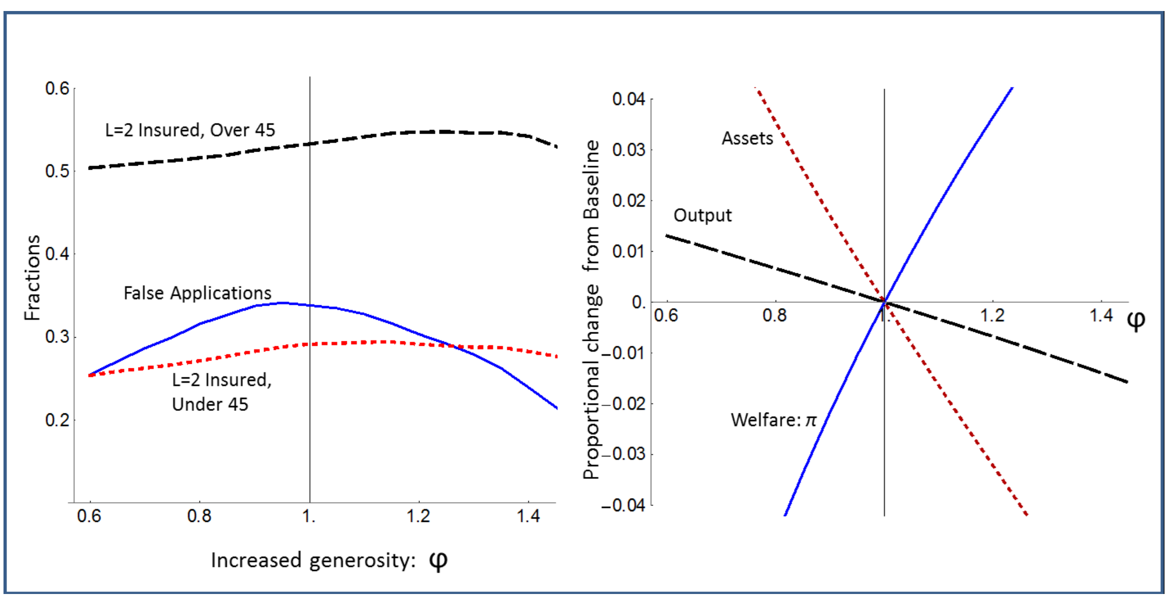

Figure 17: Increasing Generosity of Food Stamps when $\gamma = 3.0$

References

[1] Alan, S. and M. Browning (2010), “Estimating Intertemporal Allocation Parameters using Synthetic Residual Estimation,” Review of Economic Studies, 77(4), 1231-61.

[2] Altonji, J., A. Smith and I. Vidangos (2013), “Modeling earnings dynamics”, Econometrica, 81(4), 1395-1454.

[3] Attanasio, O., Low, H. and Sanchez-Marcos, V. (2008) “Explaining changes in female labour supply in a life-cycle model” *American Economic Review* 98(4): 1517–1552.

[4] Bound, J. and R. V. Burkhauser (1999), “Economic Analysis of Transfer Programs Targeted on People with Disabilities”, in Orley C. Ashenfelter and David Card (eds.), Handbook of Labor Economics. Volume 3C. Amsterdam: Elsevier Science, pp. 3417-3528.

[5] Burkhauser, R.V. and M. C. Daly. (1996) “Employment and Economic Well-Being Following the Onset of a Disability,” In Disability, Work, and Cash Benefits, edited by Jerry Mashaw, Virginia Reno, Richard Burkhauser, and Monroe Berkowitz, Upjohn Institute for Employment Research, Kalamazoo, MI.

[6] Danforth, J. P. (1979) “On the Role of Consumption and Decreasing Absolute Risk Aversion in the Theory of Job Search” in: Studies in the Economics of Search, ed. by S. A. Lippman, and J. J. McCall, pp. 109-131. North-Holland, New York.

[7] Gallant, A. R. and Nychka, D. (1987).“Semi-nonparametric Maximum Likelihood Estimation” Econometrica, 15, 363-390

[8] Gourieroux, C., A. Monfort, and E. Renault (1993), “Indirect Inference”, Journal of Applied Econometrics, Vol. 8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques, pp. S85-S118.

[9] Guvenen, F. and A. Smith (2010), “Inferring labor income risk from economic choices: an indirect inference approach,” Econometrica (forthcoming).

[10] Hotz, J., and J.K. Scholtz (2003), "The Earned Income Tax Credit," in R. Moffitt (Ed.), *Means-Tested Transfer Programs in the United States*, 141-198, National Bureau of Economic Research, Inc.

[11] Hoynes, H.W. and D. Schanzenbach, 2009. "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program," American Economic Journal: Applied Economics, vol. 1(4), pages 109-39

[12] Kapteyn, Arie, James. Smith, and Arthur van Soest. (2007). "Vignettes and Self-Reports of Work Disability in the United States and the Netherlands." American Economic Review, 97(1): 461-473.

[13] Kitao, S. (2014) "A life-cycle model of unemployment and disability insurance", Journal of Monetary Economics 68, 1-18.

[14] Lentz, R. and T. Tranaes (2005), "Job Search and Savings: Wealth Effects and Duration Dependence", Journal of Labor Economics, 23(3), 467-490.

[15] Low, Hamish, Costas Meghir and Luigi Pistaferri (2010), "Wage risk and employment risk over the life cycle", American Economic Review, 100(4): 1432-67.

[16] Mogil, Jeffrey and Andrea Bailey (2010), "Sex and gender differences in pain and analgesia", in I. Savic (Ed.), *Progress in brain research*, Elsevier North Holland.

[17] Newey, Whitney K. (2009), "Two-Step Series Estimation of Sample Selection Models", Econometrics Journal, 12(s1), S217-S229.

[18] Solomon, Carmen (1995), "Aid to Families With Dependent Children (AFDC): Need Standards, Payment Standards, and Maximum Benefits", Congressional Research Service.

[19] U.S. Department of Labor (various years), "Significant Provisions of State Unemployment Insurance Laws", July issue.

[20] Wilde, Parke (2001), "Understanding the Food Stamp Benefit Formula", Economic Research Service/USDA.