Online Appendix to ‘How Important Is Health Inequality for Lifetime Earnings Inequality?’

Roozbeh Hosseini
University of Georgia
roozbeh@uga.edu

Karen A. Kopecky
Federal Reserve Bank of Cleveland
Emory University
karen.kopecky@clev.frb.org

Kai Zhao
University of Connecticut
kai.zhao@uconn.edu

January 2025

1 Data Description

1.1 Panel Study of Income Dynamics

We use waves 1999–2017 of the PSID (covering years 1998–2016). The PSID is biennial over this period. The PSID expanded its disability and health-related questions in the 2003 wave to include questions on specific medical conditions, activities of daily living (ADLs) and instrumental activities of daily living (IADLs) which we rely on to construct individuals' frailty indices. Thus, we only construct frailty values for waves 2003 to 2017. PSID only collects detailed health information for household heads and spouses. Our base sample consists of individuals who are household heads or the spouses of a household head, are at least 25 years of age, and for which we have enough health information to construct their frailty value in at least one wave. A good description of the PSID household head definition is in Heathcote et al. (2010). The base sample consists of 19,972 individuals (9,534 men, 10,438 women).

Table 1 lists the 28 variables we use to construct frailty indices for PSID respondents. Each index is constructed by summing the variables in the first column of the table using their values which are assigned according to the rules in the second column. Then dividing this sum by the total number of variables observed for the individual in the year. The construction of the frailty index mostly follows the guidelines laid out in Searle et al. (2008), and uses a set of PSID variables similar to the those used by Yang and Lee (2009). Table 2 provides summary statistics on frailty for individuals ages 25–64 in our PSID sample.

Table 3 and 4 report summary statistics on the PSID samples used for the dynamic panel estimations. Table 3 reports summary statistics for the dynamic panel PSID sample used for the employment estimation and Table 4 reports these statistics for the worker subsample used for the earnings, hours, and wage estimation. Each sample consists of household heads and spouses aged 25 to 64. All individuals in the sample are in at least 3 consecutive waves of the PSID. Annual earnings are total annual labor earnings (including wages and salaries, bonuses, overtime, tips, commissions, professional practice or trade, any additional job income, the labor part of self-employment income and any miscellaneous labor income). The PSID splits self-employment income into asset and labor components using a 50-50 rule. Annual hours are the total annual work hours for all jobs, including overtime. Hourly wage is a PSID constructed variable that is constructed using annual earnings and annual hours. It is adjusted by PSID for outliers. Education was cleaned and reassigned so that education is constant across all waves for each individual. Labor force status is considered not employed if annual hours is between 0 and 519 and employed (workers) if annual hours are 520 or more. Observations where labor force status is deemed employed based on hours but wages are reported as less than \$4 an hour are dropped.

Table 1: Health Variables used to construct frailty indices for PSID respondents

Variable	Value
Some difficulty with ADL/IADLs:	
Eating	Yes=1, No=0
Dressing	Yes=1, No=0
Getting in/out of bed or chair	Yes=1, No=0
Using the toilet	Yes=1, No=0
Bathing/showering	Yes=1, No=0
Walking	Yes=1, No=0
Using the telephone	Yes=1, No=0
Managing money	Yes=1, No=0
Shopping for personal items	Yes=1, No=0
Preparing meals	Yes=1, No=0
Heavy housework	Yes=1, No=0
Light housework	Yes=1, No=0
Getting outside	Yes=1, No=0
Ever had one of following conditions:	
High Blood Pressure	Yes=1, No=0
Diabetes	Yes=1, No=0
Cancer	Yes=1, No=0
Lung disease	Yes=1, No=0
Heart disease	Yes=1, No=0
Heart attack	Yes=1, No=0
Stroke	Yes=1, No=0
Arthritis	Yes=1, No=0
Asthma	Yes=1, No=0
Loss of memory or mental ability	Yes=1, No=0
Psychological problems	Yes=1, No=0
Other serious, chronic condition	Yes=1, No=0
BMI ≥ 30	Yes=1, No=0
Has ever smoked	Yes=1, No=0
Currently smokes	Yes=1, No=0

Figure 1 shows the raw correlations between frailty and earnings, employment, hours worked (if employed) and hourly wages by 5-year age groups for individuals aged 25 to 64. The figure shows that frailty and earnings are negatively correlated at all ages. This negative correlation is due to a negative correlation between frailty and all three components of earnings (employment, hours worked, and wages). However, the correlation between frailty and employment is the largest and follows a similar life cycle pattern to that of earnings.

1.2 Health and Retirement Survey

The HRS is a biennial longitudinal survey of Americans over age 50. Aside from spouses of respondents, the HRS does not survey individuals under the age of 51. We use the HRS waves spanning the period 1998 to 2014. Our sample consists of 205,711 observations of 36,032 individuals (15,860 men and 20,172 women). We construct a frailty index for HRS respondents in the same way as we do for PSID respondents. The lifecycle dynamics of frailty in the HRS and PSID samples are very similar even though the HRS contains a larger number of deficit variables (36 versus 28). See Hosseini et al. (2022) for additional details.

1.3 Medical Expenditure Panel Survey

The MEPS consists of a collection of rotating two-year panels. We use MEPS data from the 2000 to 2016 period. Our sample consists of respondents aged 25 to 84 years. We do not include individuals aged 85

Table 2: Frailty summary statistics for 25- to 64-year-olds in the PSID

Mean	0.11	Median	0.07
by age:		Standard Deviation	0.10
25-34	0.07		
35-44	0.09	*Wave-to-wave:*	
45-54	0.11	$+\Delta$ Frailty	0.27
55-64	0.15	$-\Delta$ Frailty	0.12
by education:			
Less than high school	0.13	Effect of 1 additional deficit	+0.036
High school	0.11		
College	0.08		

Table 3: Summary statistics on our dynamic panel PSID sample

	1998	2000	2002	2004	2006	2008	2010	2012	2014	2016	Pooled 2002–2016
Panel A: Mean (median) [standard deviation] of sample characteristics											
Age	40.64	41.54	42.31	42.51	42.7	42.88	43.02	42.86	43.49	44.16	42.66
	(40)	(42)	(42)	(43)	(43)	(43)	(43)	(42)	(42)	(43)	(42)
	[8.86]	[9.49]	[10.17]	[10.58]	[10.90]	[11.20]	[11.49]	[11.61]	[11.50]	[11.21]	[10.84]
Frailty	-	-	0.08	0.09	0.10	0.10	0.11	0.11	0.11	0.12	0.10
	-	-	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
	-	-	[0.09]	[0.09]	[0.10]	[0.10]	[0.10]	[0.11]	[0.11]	[0.11]	[0.10]
Employed	0.88	0.87	0.85	0.84	0.84	0.83	0.79	0.80	0.80	0.81	0.83
	[0.33]	[0.34]	[0.36]	[0.37]	[0.37]	[0.38]	[0.41]	[0.40]	[0.40]	[0.39]	[0.38]
Panel B: Fraction of sample by characteristics											
Male	0.46	0.46	0.46	0.46	0.46	0.46	0.45	0.45	0.45	0.44	0.45
High School Dropouts (HSD)	0.14	0.14	0.13	0.13	0.12	0.11	0.11	0.11	0.11	0.11	0.12
High School Graduates (HS)	0.51	0.52	0.51	0.51	0.50	0.50	0.51	0.51	0.51	0.51	0.51
College Graduates (CL)	0.35	0.35	0.35	0.37	0.38	0.38	0.38	0.38	0.38	0.38	0.37
$+\Delta$ Frailty	-	-	-	0.27	0.30	0.29	0.27	0.26	0.25	0.25	0.27
$-\Delta$ Frailty	-	-	-	0.11	0.11	0.11	0.12	0.12	0.12	0.12	0.11
Observations (N)	7,541	8,350	9,124	9,349	9,703	10,074	10,281	10,310	9,549	8,840	93,121
# of Individuals (n)											14,468
Average # of Years Observed (T)											6.44

Note: Means are reported; median values are reported in parentheses; standard deviations are reported in brackets.

years or older because, starting in 2001, MEPS top codes age at 85. The base sample contains 345,022 observations on 191,165 individuals (88,389 men and 102,776 women). The frailty index is constructed in the same way as for PSID and HRS respondents and has similar lifecycle dynamics. See Hosseini et al. (2022) for additional details.

2 Dynamic Panel Analysis: More Results

In Section 2 of the paper we use a dynamic GMM panel estimator to estimate the impact of frailty on employment, earnings, hours and wages. In this section of the appendix we describe the dynamic panel estimator in more detail and present additional results regarding validity of instruments, causality, and further diagnostics. We also report, in Table 5, the mean employment rates used to calculate the percent changes in the probability of employment reported in Table 3 in the paper.

To obtain a consistent and unbiased estimate of the effect of frailty on earnings we use a dynamic GMM panel estimator. The basic estimation procedure is as follows. Let's write equation (1) in the

Table 4: Summary statistics on our dynamic panel PSID sample of workers

	1998	2000	2002	2004	2006	2008	2010	2012	2014	2016	Pooled 2002–2016
Panel A: Mean (median) [standard deviation] of sample characteristics											
Age	38.25	39.14	39.89	40.28	40.69	41.02	41.26	41.2	42	42.97	40.82
	(38)	(39)	(40)	(40)	(41)	(40)	(40)	(40)	(40)	(41)	(40)
	[7.61]	[8.23]	[8.87]	[9.40]	[9.85]	[10.31]	[10.79]	[11.11]	[11.11]	[11.03]	[10.13]
Frailty	-	-	0.06	0.07	0.07	0.08	0.08	0.08	0.08	0.09	0.08
	-	-	(0.04)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
	-	-	[0.06]	[0.06]	[0.06]	[0.07]	[0.07]	[0.07]	[0.07]	[0.08]	[0.07]
Annual Earnings	$50,271.05	$53,932.73	$53,966.43	$55,839.21	$56,779.83	$57,730.84	$55,439.87	$56,955.61	$56,538.02	$58,660.52	$55,862.43
	(39,885.13)	(41,797.15)	(42,084.94)	(42,648.47)	(43,846.86)	(44,595.46)	(42,919.39)	(42,000.00)	(43,761.55)	(46,113.50)	(43,176.04)
	[50,848.39]	[56,984.74]	[87,328.86]	[68,068.02]	[62,083.35]	[66,494.52]	[64,696.26]	[122,776.17]	[89,417.44]	[63,326.55]	[77,796.60]
Annual Hours	2,155.06	2,118.07	2,145.05	2,163.69	2,152.99	2,062.04	2,069.14	2,104.42	2,117.51	2,120.13	2,118.28
	(2,080)	(2,056)	(2,080)	(2,080)	(2,080)	(2,016)	(2,040)	(2,040)	(2,056)	(2,065)	(2,056)
	[632.05]	[615.30]	[648.55]	[657.68]	[637.93]	[579.79]	[620.47]	[628.91]	[620.65]	[630.52]	[627.73]
Hourly Wage	$23.48	$25.45	$25.01	$26.09	$26.31	$27.95	$26.93	$26.24	$26.49	$27.87	$26.32
	(18.91)	(20.08)	(19.91)	(20.40)	(20.57)	(21.50)	(20.90)	(20.18)	(20.42)	(21.71)	(20.52)
	[19.48]	[23.73]	[24.04]	[28.89]	[24.95]	[29.15]	[26.38]	[26.94]	[24.63]	[25.13]	[25.72]
Panel B: Fraction of sample by characteristics											
Male	0.53	0.54	0.54	0.54	0.55	0.54	0.54	0.54	0.53	0.53	0.54
High School Dropouts (HSD)	0.09	0.09	0.09	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.08
High School Graduates (HS)	0.50	0.50	0.50	0.49	0.49	0.48	0.48	0.48	0.48	0.49	0.49
College Graduates (CL)	0.41	0.40	0.41	0.43	0.44	0.45	0.45	0.45	0.45	0.44	0.44
$+\Delta$ Frailty	-	-	-	0.23	0.27	0.25	0.22	0.22	0.22	0.22	0.23
$-\Delta$ Frailty	-	-	-	0.09	0.09	0.09	0.09	0.10	0.09	0.10	0.09
Observations (N)	3,514	3,996	4,493	4,579	4,835	5,107	5,384	5,676	5,447	5,104	48,135
# of Individuals (n)											7,598
Average # of Years Observed (T)											6.34

Note: Means are reported; median values are reported in parentheses; standard deviations are reported in brackets.

paper in first difference form:

$$\Delta y_{i,t} = \alpha_1 \Delta y_{i,t-1} + \alpha_2 \Delta y_{i,t-2} + \gamma \Delta f_{i,t} + \delta \Delta \boldsymbol{Z}_{i,t} + \Delta \varepsilon_{i,t}, \tag{1}$$

which eliminates time-invariant unobserved heterogeneity (the fixed effect term b_i). Note that the assumption of lack of serial correlation in $\varepsilon_{i,t}$ implies that the following orthogonality conditions must hold

$$E\left(y_{i,t-s} \Delta \varepsilon_{i,t}\right) = E\left(f_{i,t-s} \Delta \varepsilon_{i,t}\right) = 0, \quad \text{for } \forall s > 2. \tag{2}$$

In other words, the vectors $(y_{i,1}, y_{i,2}, \ldots, y_{i,t-3})$ and $(f_{i,1}, f_{i,2}, \ldots, f_{i,t-3})$ are uncorrelated with $\Delta\varepsilon_{i,t}$, and can be used as 'internal' instruments for estimating equation (1) using GMM. As we argued above, lagged values of frailty and earnings are predictors of current levels of earnings and frailty. Therefore, they provide sources of variations for current values.

The key identifying assumption for validity of this procedure is the assumption of no serial correlation in $\varepsilon_{i,t}$. The assumption should hold if we have included enough lags of earnings in equation (1) in the paper to capture the dynamics of earnings over time. By construction, the first-differences $\Delta\varepsilon_{i,t}$ should possess first-order serial correction. However, if the assumption of no serial correlation in $\varepsilon_{i,t}$ is correct, $\Delta\varepsilon_{i,t}$ should not exhibit second-order serial correlation. We can test this assumption directly. We can also test directly the validity of the orthogonality conditions (2) for our instrument set by using the Hansen-Sargan test, the standard GMM test of overidentifying restrictions (see Hansen (1982) and Sargan (1958)).

Although, in principle we can use the procedure above to estimate equation (1), there may still be a few shortcomings. For example, differencing can reduce variation in explanatory variables and therefore reduce accuracy of estimates (see Beck et al. (2000)). Moreover, as Arellano and Bover (1995) point out, variables in levels may be weak instruments for first-differences. This is especially true for highly persistent variables. As a stark example, imagine a random walk process. In that case, past levels are uncorrelated with first differences. To mitigate these shortcomings, we follow Blundell and Bond (1998) and Blundell and Bond (2000) and improve the GMM estimator by introducing an additional set of

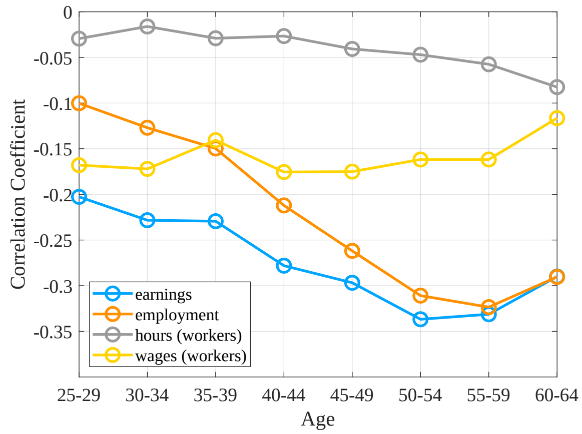


Figure 1: Raw correlations of earnings, employment, hours conditional on working and observed wages with frailty for 25- to 64-year-olds by five-year age groups. Data source: 2003–2017 PSID.

Table 5: Mean employment rate by education, health, and age

	All	By Education			By Health		By Age	
		HSD	HSG	CL	Good	Poor	Young	Old
Employment rate	0.825	0.676	0.815	0.890	0.863	0.542	0.863	0.774
	(0.002)	(0.008)	(0.003)	(0.003)	(0.002)	(0.008)	(0.002)	(0.004)

Note: 'HSD' is high school dropout, 'HS' is high school graduate, and 'CL' is college graduate. 'Good/Poor Health' is frailty below/above the 85th percentile. 'Young/Old' are individuals younger/older than 45 years of age. Standard errors are in parenthesis.

orthogonality conditions. Specifically,

$$E\left(\Delta y_{i,t-s}\left(b_i+\varepsilon_{i,t}\right)\right)=E\left(\Delta f_{i,t-s}\left(b_i+\varepsilon_{i,t}\right)\right)=0, \quad \text{for } \forall s>2. \tag{3}$$

These orthogonality conditions imply that changes in earnings and frailty over time are uncorrelated with the fixed effects in equation (1) in the paper. This is a strong overidentifying assumption which we can test using a Difference Hansen-Sargan test (see Bond (2002a)).

Under the assumption that the orthogonality conditions in equation (2) and equation (3) hold, we jointly estimate equation (1) in the paper and equation (1) above via GMM using lagged first differences as 'internal' instruments for levels and lagged levels as 'internal' instruments for first differences. Given the concerns about instrument proliferation and overfitting discussed in Roodman (2009) we do not use all the available lags as instruments. Instead, we use only the fourth and fifth lags for the estimations that include everyone in the sample. Some of the estimations are done on a subsample of workers. For these estimations we use the fifth and sixth lags in order to obtain valid instruments. In all the estimations, we also restrict the coefficients on the lags to be the same at each time t by collapsing the instrument matrix. As argued in Roodman (2009) this increases the power of the Hansen-Sargan test for over-identification.

2.1 Comparison with OLS and fixed effect estimators

2.1.1 The impact of frailty on employment

For purpose of comparison, we start by reporting results from OLS and fixed effect (within group) estimations of Equation (1) in Section 2 of the paper for employment. It is well known that the OLS estimates of the coefficients on lagged values of the left-hand-side variable have an upward bias. Moreover, as Nickell (1981), Arellano and Bond (1991), and Bond (2002b) have shown, the estimates acquired via a fixed effects estimator have a downward bias. Therefore, an unbiased estimate should lie between the OLS and FE estimates. As Bond (2002b) argues, if multiple lags of the left-hand-side variable are

Table 6: Effect of Frailty on Employment

	Panel A. OLS linear probability				**Panel B.** Fixed effect linear probability				**Panel C.** Logit			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
frailty_t	-0.018***				-0.031***				-0.014***			
	(0.001)				(0.001)				(0.000)			
$\text{frailty}_t \times$ HSD		-0.020***				-0.030***				-0.018***		
		(0.001)				(0.002)				(0.001)		
$\text{frailty}_t \times$ HS		-0.020***				-0.034***				-0.015***		
		(0.001)				(0.001)				(0.001)		
$\text{frailty}_t \times$ CL		-0.012***				-0.025***				-0.010***		
		(0.001)				(0.002)				(0.001)		
$\text{frailty}_t \times$ Good Health			-0.007***				-0.017***				-0.008***	
			(0.001)				(0.002)				(0.001)	
$\text{frailty}_t \times$ Poor Health			-0.023***				-0.033***				-0.015***	
			(0.001)				(0.001)				(0.001)	
$\text{frailty}_t \times$ Young				-0.015***				-0.026***				-0.013***
				(0.001)				(0.001)				(0.001)
$\text{frailty}_t \times$ Old				-0.020***				-0.034***				-0.015***
				(0.001)				(0.001)				(0.001)
employed_{t-1}	0.508***	0.507***	0.505***	0.508***	0.155***	0.155***	0.154***	0.154***	0.420***	0.420***	0.419***	0.420***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.007)
employed_{t-2}	0.187***	0.186***	0.184***	0.187***	-0.036***	-0.037***	-0.038***	-0.037***	0.132***	0.132***	0.131***	0.132***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Gender FE	YES	YES	YES	YES					YES	YES	YES	YES
Education FE	YES	YES	YES	YES					YES	YES	YES	YES
Observations	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576
Groups					14,468	14,468	14,468	14,468				
Adjusted R^2	0.500	0.501	0.502	0.500								
Within R^2					0.096	0.097	0.098	0.097				
Between R^2					0.154	0.154	0.146	0.154				
Overall R^2					0.167	0.168	0.164	0.168				
Pseudo R^2									0.453	0.453	0.454	0.453
Log likelihood									-17510.16	-17497.60	-17478.22	-17506.65

Note: Frailty effects are the effect of incurring one additional deficit. Logit estimates are the average marginal effects. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'HSD' is high school dropout, 'HS' is high school graduate, and 'CL' is college graduate. 'Good/Poor Health' is frailty below/above the 85th percentile. 'Young/Old' are individuals younger/older than 45 years of age. Standard errors clustered at the individual level are in parenthesis. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

included on the right-hand-side, the sum of the coefficients on these variables must satisfy this condition. In other words, a necessary condition for the dynamic panel estimates to be unbiased is that the sum of the estimated values of α_1 and α_2 in Equation (1) in the paper are smaller than the corresponding sum of OLS estimates, but larger than those from the fixed effect estimation.

Panels A and B of Table 6 show the OLS and fixed effect linear probability estimation results for employment. We report a similar set of results for earnings, hours, and wages in the next subsection. The numbers reported in Panels A and B, together with those reported in Table 2 of the main text, pass the consistency test described above. In other words, in all the specifications, we cannot reject that the sum of the coefficients on lagged employment in Table 2 lies between those in panel A and B of Table 6.

Results in the paper are from the estimation of a linear probability model with fixed effects using a dynamic panel GMM estimator. For employment, ideally, we would like to estimate a non-linear probability model. However, we are not aware of a standard procedure for estimating non-linear dynamic panel models with fixed effects. As discussed in Arellano and Honoré (2001), among others, the currently available methods for performing such estimations, rely on insights that are model-specific and are not generally applicable (unlike the system GMM procedure for linear models). We can, however, estimate a non-linear (logit) model of the effect of frailty on employment. We report the results of this estimation in Table 6. For purpose of comparison we also repeat the linear probability estimation using OLS and fixed effect estimators. We do not estimate a fixed effect logit model because computing average marginal effects and their standard errors in such models is not straightforward and coefficient values are not comparable across models.

Panel C in Table 6 shows the estimation results from the logit model. Here, we report the average marginal effects on probability of employment. Note that the estimated effects of frailty on employment are similar to those reported in Table 2 of the main text (from system GMM estimation of the linear probability model). However, the magnitudes are all slightly smaller relative to linear probability model. Also notice that all effects are significant. For example, unlike the system GMM estimator of the linear probability model, the logit model absent fixed effects shows a significant effect of frailty on the college educated, those in good health, and those younger than 45 years old. Despite these differences, the estimated effects of frailty on the probability of being employed are similar in magnitude and pattern across all the specifications considered.

2.1.2 The impact of frailty on earnings, hours, and wages

For purpose of comparison (in the same spirit as the previous section), we also estimate Equation (1) in Section 2 of the paper using OLS and fixed effect (within group) estimators for earnings, hours, and wages. The results are reported in Tables 7, 8, 9 and 10 for the overall effect, the effect by education, by health, and by age group, respectively. The three panels in the table show results for earnings, hours, and wages, respectively. All of these regressions are estimated on the sub-sample of workers used in the dynamic panel estimation. Therefore, the last column in each panel corresponds to the respective estimation in Table 3 of the main text.

As in the previous section we would like to see that the (sums of) the coefficients on lagged earnings, hours, and wages from the system GMM estimation (the last column in each panel) lie in between the OLS estimator (the first column) and fixed effect estimator (the middle column). The idea being that the OLS estimator is biased upward and fixed effect estimator is biased downward. We cannot statistically reject this condition in any of our estimations. Therefore, our regressions pass this consistency test.

Note, also that in Tables 7 through 10 the OLS estimation shows a significant effect of frailty on hours even for those who continuously work. This is likely due to the fact that, in these OLS estimations, individuals' fixed effects are ignored. The fixed effects are included in the FE estimation but this estimator is biased.

Table 7: Comparison with OLS and Fixed Effect Estimator, Average Frailty Effect

	Panel A. Earnings regression			**Panel B.** Hours regression			**Panel C.** Wage regression		
	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM
frailty_t	-0.016***	-0.016***	-0.029**	-0.005***	-0.014***	0.005	-0.011***	-0.002	-0.020**
	(0.002)	(0.003)	(0.014)	(0.001)	(0.002)	(0.009)	(0.001)	(0.003)	(0.009)
$\log(\text{earnings}_{t-1})$	0.532***	0.097***	1.259***						
	(0.010)	(0.006)	(0.455)						
$\log(\text{earnings}_{t-2})$	0.227***	-0.032***	-0.427						
	(0.008)	(0.006)	(0.411)						
$\log(\text{hours}_{t-1})$				0.331***	-0.028***	-0.086			
				(0.008)	(0.006)	(0.366)			
$\log(\text{hours}_{t-2})$				0.157***	-0.091***	0.336			
				(0.007)	(0.006)	(0.229)			
$\log(\text{wage}_{t-1})$							0.499***	0.066***	0.205
							(0.008)	(0.006)	(0.505)
$\log(\text{wage}_{t-2})$							0.268***	-0.029***	0.543
							(0.007)	(0.006)	(0.454)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34,505	34,505	34,505	34,505	34,505	34,505	34,401	34,401	34,401
R^2	0.611	0.077		0.234	0.001		0.604	0.056	
AR(1) test (p-value)			0.048			0.495			0.629
AR(2) test (p-value)			0.230			0.218			0.408
Hansen-Sargan test (p-value)			0.396			0.206			0.079
Diff. Hansen-Sargan test (p-value)			0.227			0.404			0.046

Note: Panel A shows regression results for the effect of frailty on earnings of workers. Panel B shows regression results for the effect of frailty on hours of workers. Panel C shows regression results for the effect of frailty on workers' wages. Frailty effects are the effect of incurring one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'FE' is fixed effect (within groups) estimation. Standard errors are in parenthesis. R^2 is adjusted R-squared for OLS, and overall R-squared for FE. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

2.2 Additional instrument diagnostic tests

The Difference Hansen-Sargan test results reported in Tables 2, 3, and Table 4 in the paper test jointly the exogeneity of all the lagged differences used as instruments in the dynamic panel estimation including those instrumenting the lagged values of the left-hand-side-variable (the y-lag set) and those instrumenting frailty. Following the recommendation of Roodman (2009), we also test separately the exogeneity of the y-lag set. Table 11 presents p-values from these Difference Hansen-Sargan tests. Notice that in all of the regressions we fail to reject the null that the instruments for the y-lag variables are valid.

2.3 Robustness to instrument set

Following the recommendation of Roodman (2009), we explore the robustness of our estimates to the set of lagged levels and first differences used as instruments. For the results reported in Section 2 in the paper we use the 4th and 5th lags as instruments for the employment regressions and the 5th and 6th lags as instruments for the earnings, hours, and hourly wage regressions. Here, we report the results of each estimation using either lags 4 and 5, 4 through 6, or 5 and 6 for the employment regressions. For the earnings, hours, and hourly wage regressions, we report the results of each estimation using either lags 4 and 5, 5 and 6, or 5 through 7.

We find that the results are robust to variation in the set of instruments used when valid instruments are chosen. Table 12 reports the results for the effect of frailty on employment. Tables 13, 14, and 15 report results for the effect on earnings, hours worked, and hourly wages of workers, respectively. Looking at Table 12, the estimations that use the 4th through 6th lags yield very similar results. Many of the lag combinations that only include two lags fail to pass the AR(2) and instrument validity tests and yield similar results to those using the 4th and 5th lag. Tables 13, 14, and 15 show a similar pattern. Except that in these cases, the estimation results using the 5th and 6th lags are very similar to those that use the

Table 8: Comparison with OLS and Fixed Effect Estimator, Frailty Effects by Education

	Panel A. Earnings regression			**Panel B.** Hours regression			**Panel C.** Wage regression		
	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM
$\text{frailty}_t \times$ HSD	-0.021***	-0.025***	-0.063**	-0.014***	-0.030***	-0.020*	-0.010*	0.006	-0.065***
	(0.005)	(0.009)	(0.028)	(0.004)	(0.006)	(0.012)	(0.005)	(0.008)	(0.021)
$\text{frailty}_t \times$ HS	-0.013***	-0.026***	-0.038***	-0.006***	-0.014***	0.001	-0.009***	-0.012***	-0.029***
	(0.002)	(0.004)	(0.015)	(0.001)	(0.003)	(0.010)	(0.002)	(0.003)	(0.010)
$\text{frailty}_t \times$ CL	-0.017***	0.002	-0.015	-0.003*	-0.010***	0.011	-0.015***	0.012***	-0.005
	(0.002)	(0.005)	(0.016)	(0.002)	(0.003)	(0.009)	(0.002)	(0.004)	(0.010)
$\log(\text{earnings}_{t-1})$	0.532***	0.096***	1.197***						
	(0.010)	(0.006)	(0.358)						
$\log(\text{earnings}_{t-2})$	0.227***	-0.033***	-0.394						
	(0.008)	(0.006)	(0.323)						
$\log(\text{hours}_{t-1})$				0.331***	-0.028***	0.013			
				(0.008)	(0.006)	(0.326)			
$\log(\text{hours}_{t-2})$				0.157***	-0.091***	0.191			
				(0.007)	(0.006)	(0.227)			
$\log(\text{wage}_{t-1})$							0.499***	0.066***	0.160
							(0.008)	(0.006)	(0.344)
$\log(\text{wage}_{t-2})$							0.268***	-0.030***	0.570*
							(0.007)	(0.006)	(0.307)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34,505	34,505	34,505	34,505	34,505	34,505	34,401	34,401	34,401
R^2	0.611	0.087		0.234	0.001		0.604	0.066	
AR(1) test (p-value)			0.016			0.354			0.486
AR(2) test (p-value)			0.154			0.489			0.190
Hansen-Sargan test (p-value)			0.770			0.375			0.292
Diff. Hansen-Sargan test (p-value)			0.411			0.333			0.160

Note: Panel A shows regression results for the effect of frailty on earnings of workers by education. Panel B shows regression results for the effect of frailty on hours of workers by education. Panel C shows regression results for the effect of frailty on workers' wages by education. Frailty effects are the effect of incurring one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'HSD' is high school dropout, 'HS' is high school graduate, and 'CL' is college graduate. 'FE' is fixed effect (within groups) estimation. Standard errors are in parenthesis. R^2 is adjusted R-squared for OLS, and overall R-squared for FE. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

5th through 7th lags. However, the estimations that use the 4th and 5th lags are smaller, not significant, and fail the Difference Hansen-Sargan instrument validity tests. To summarize, among the specifications that we consider, those that pass the instrument validity test yield very similar results.

2.4 Instrument power tests

The Hansen-Sargan test and Difference Hansen-Sargan test only test for instrument exogeneity. They do not test instrument power. To test instrument power, we use the methodology of Staiger and Stock (1997) and Stock and Yogo (2005) following Wintoki et al. (2012). That is, we look at the strength of the F-statistics in the first stage regressions. Under the system-GMM procedure that we use, there are two "first stage" equations, one where levels are instruments for first differences, and one where first differences are instruments for levels. We regress the endogenous variables from each equation on their corresponding instrument set, which yields an F-statistic that we evaluate for instrument strength.

To run the instrument power tests, we estimate OLS regressions of the endogenous variables in the level and first-difference equations on their appropriate instrument set and the vector of exogenous controls (marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth-degree polynomial in age). Each endogenous variable is regressed on it's instruments starting with the 4th lag alone and then adding the 5th lag. Each equation generates an F-statistic which is used to test whether the parameters estimated in the equation are jointly equal to zero. This tells us the statistical power of the instruments in explaining the variation in the endogenous

Table 9: Comparison with OLS and Fixed Effect Estimator, Frailty Effects by Health Status

	Panel A. Earnings regression			**Panel B.** Hours regression			**Panel C.** Wage regression		
	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM
frailty_t × Good Health	-0.013***	-0.004	-0.032	-0.003*	-0.008***	0.013	-0.012***	0.005	-0.009
	(0.002)	(0.004)	(0.033)	(0.001)	(0.003)	(0.017)	(0.002)	(0.004)	(0.023)
frailty_t × Poor Health	-0.016***	-0.016***	-0.027**	-0.006***	-0.015***	0.000	-0.011***	-0.002	-0.021**
	(0.002)	(0.003)	(0.012)	(0.001)	(0.002)	(0.008)	(0.001)	(0.003)	(0.008)
$\log(\text{earnings}_{t-1})$	0.532***	0.097***	1.215***						
	(0.010)	(0.006)	(0.400)						
$\log(\text{earnings}_{t-2})$	0.227***	-0.032***	-0.392						
	(0.008)	(0.006)	(0.362)						
$\log(\text{hours}_{t-1})$				0.331***	-0.028***	0.097			
				(0.008)	(0.006)	(0.361)			
$\log(\text{hours}_{t-2})$				0.157***	-0.091***	0.248			
				(0.007)	(0.006)	(0.215)			
$\log(\text{wage}_{t-1})$							0.499***	0.066***	0.381
							(0.008)	(0.006)	(0.411)
$\log(\text{wage}_{t-2})$							0.268***	-0.029***	0.380
							(0.007)	(0.006)	(0.368)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34,505	34,505	34,505	34,505	34,505	34,505	34,401	34,401	34,401
R^2	0.611	0.076		0.234	0.001		0.604	0.055	
AR(1) test (p-value)			0.030			0.353			0.463
AR(2) test (p-value)			0.202			0.429			0.543
Hansen-Sargan test (p-value)			0.604			0.325			0.140
Diff. Hansen-Sargan test (p-value)			0.366			0.380			0.156

Note: Panel A shows regression results for the effect of frailty on earnings of workers by health status. Panel B shows regression results for the effect of frailty on hours of workers by health status. Panel C shows regression results for the effect of frailty on workers' wages by health status. Frailty effects are the effect of incurring one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'Good/Poor Health' is frailty below/above the 85th percentile. 'FE' is fixed effect (within groups) estimation. Standard errors are in parenthesis. R^2 is adjusted R-squared for OLS, and overall R-squared for FE. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

variable. Following Wintoki et al. (2012) we adopt a critical value for sufficient instrument strength of 10. This value is proposed by Staiger and Stock (1997).

We conduct the instrument power tests for the system GMM estimation of the overall impact of frailty on employment (Table 1 in the main text). The results of the tests are reported in Table 16. Notice that, across each result, the F-statistic is decreasing with the specification of further lags of the instrument set. Excluding the last two values in the bottom row of the table, all the F-statistics are greater than 10 suggesting that the instruments have sufficient power in explaining the variation in the endogenous variables. Panel B of the table indicates that the lagged levels are relatively weak instruments for first-differences as compared to the strength of lagged differences as instruments for the levels. As we mentioned in Section 2 of the paper, this is not surprising given that both frailty and log earnings are highly persistent variables. It is also the reason we use the system GMM estimator as opposed to working only with the difference equation.

Table 17 shows similar results for earnings regression. Lagged differences seem to be powerful instruments for levels. However, lagged levels are not very powerful instruments for differences. This is due to the fact that both frailty and earnings are highly persistent variables. It also provides justification for using a more elaborate system-GMM estimator as opposed to a simpler estimator that only uses lagged values of levels as instruments for differences.

2.5 More discussion on instrument validity

As we discussed in the previous subsection, lagged values of variables are not always powerful instruments of first differences. This is true even when they pass the test for instrument validity (Equation (2) in the

Table 10: Comparison with OLS and Fixed Effect Estimator, Frailty Effects by Age Group

	Panel A. Earnings regression			**Panel B.** Hours regression			**Panel C.** Wage regression		
	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM	OLS	FE	SYS-GMM
$\text{frailty}_t \times$ Young	-0.018***	-0.015***	-0.035	-0.005***	-0.014***	0.014	-0.014***	-0.001	-0.015
	(0.002)	(0.004)	(0.026)	(0.001)	(0.002)	(0.016)	(0.002)	(0.003)	(0.020)
$\text{frailty}_t \times$ Old	-0.013***	-0.016***	-0.016	-0.005***	-0.014***	-0.003	-0.009***	-0.002	-0.022*
	(0.002)	(0.003)	(0.013)	(0.001)	(0.002)	(0.009)	(0.002)	(0.003)	(0.012)
$\log(\text{earnings}_{t-1})$	0.532***	0.097***	1.050***						
	(0.010)	(0.006)	(0.307)						
$\log(\text{earnings}_{t-2})$	0.227***	-0.032***	-0.250						
	(0.008)	(0.006)	(0.282)						
$\log(\text{hours}_{t-1})$				0.331***	-0.028***	0.181			
				(0.008)	(0.006)	(0.342)			
$\log(\text{hours}_{t-2})$				0.157***	-0.091***	0.244			
				(0.007)	(0.006)	(0.239)			
$\log(\text{wage}_{t-1})$							0.499***	0.066***	0.399
							(0.008)	(0.006)	(0.390)
$\log(\text{wage}_{t-2})$							0.268***	-0.029***	0.365
							(0.007)	(0.006)	(0.351)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34,505	34,505	34,505	34,505	34,505	34,505	34,401	34,401	34,401
R^2	0.611	0.077		0.234	0.001		0.604	0.056	
AR(1) test (p-value)			0.016			0.300			0.429
AR(2) test (p-value)			0.243			0.531			0.549
Hansen-Sargan test (p-value)			0.358			0.163			0.106
Diff. Hansen-Sargan test (p-value)			0.140			0.391			0.257

Note: Panel A shows regression results for the effect of frailty on earnings of workers by age. Panel B shows regression results for the effect of frailty on hours of workers by age. Panel C shows regression results for the effect of frailty on workers' wages by age. Frailty effects are the effect of incurring one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'Young/Old' are individuals younger/older than 45 years of age. 'FE' is fixed effect (within groups) estimation. Standard errors are in parenthesis. R^2 is adjusted R-squared for OLS, and overall R-squared for FE. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 11: Difference Hansen-Sargan test, y-lag set only (p-value) for regressions in Tables 1, 2 and 4 in the paper

	Overall (1)	By educ (2)	By health (3)	By age (4)
Effect of frailty on employment	0.201	0.078	0.051	0.394
Effect of frailty on earnings	0.396	0.330	0.383	0.223
Effect of frailty on hours	0.206	0.077	0.131	0.098
Effect of frailty on wages	0.079	0.066	0.061	0.060
Effect of employment on frailty	0.475	0.669	0.652	0.019
Effect of earnings on frailty	0.182	0.198	0.335	0.225

main body). To mitigate this issue Blundell and Bond (1998) and Blundell and Bond (2000) improve the GMM estimator by introducing the following additional set of orthogonality assumptions (Equation (3) in the main body):

$$E\left(\Delta y_{i,t-s}\left(b_i + \varepsilon_{i,t}\right)\right) = E\left(\Delta f_{i,t-s+2}\left(b_i + \varepsilon_{i,t}\right)\right) = 0, \qquad \text{for } \forall s > 3. \tag{4}$$

If these conditions hold, we can use lagged first differences as 'internal' instruments for levels. As we reported in the previous subsection these are powerful instruments. However, these additional orthogonality conditions require that changes in frailty be independent of initial fixed unobserved heterogeneity. This is a strong overidentifying assumption. One way to test for these assumptions is to use the Dif-

Table 12: Comparison With Different IV Lags, Employment Regressions

IV Lags	4-6	4-5	5-6	4-6	4-5	5-6	4-6	4-5	5-6	4-6	4-5	5-6
Number of Instruments	24	22	22	32	28	28	28	25	25	28	25	25
frailty_t	-0.015***	-0.016**	-0.010									
	(0.005)	(0.006)	(0.008)									
$\text{frailty}_t \times$ HSD				-0.020***	-0.025***	-0.015						
				(0.006)	(0.007)	(0.010)						
$\text{frailty}_t \times$ HS				-0.016***	-0.020***	-0.011						
				(0.005)	(0.007)	(0.008)						
$\text{frailty}_t \times$ CL				-0.007	-0.008	-0.004						
				(0.005)	(0.005)	(0.006)						
$\text{frailty}_t \times$ Good Health							-0.015	-0.015	-0.018			
							(0.011)	(0.011)	(0.015)			
$\text{frailty}_t \times$ Poor Health							-0.015***	-0.018***	-0.011			
							(0.005)	(0.005)	(0.008)			
$\text{frailty}_t \times$ Young										-0.015**	-0.014*	-0.010
										(0.007)	(0.007)	(0.010)
$\text{frailty}_t \times$ Old										-0.017***	-0.019***	-0.014*
										(0.005)	(0.006)	(0.007)
employed_{t-1}	0.551*	0.520	0.550	0.702**	0.327	0.714**	0.508*	0.255	0.600	0.520*	0.451	0.492
	(0.314)	(0.403)	(0.381)	(0.308)	(0.447)	(0.356)	(0.307)	(0.363)	(0.414)	(0.304)	(0.380)	(0.372)
employed_{t-2}	0.149	0.172	0.255	0.017	0.314	0.115	0.186	0.382	0.230	0.172	0.224	0.286
	(0.247)	(0.317)	(0.290)	(0.242)	(0.352)	(0.266)	(0.239)	(0.284)	(0.309)	(0.239)	(0.299)	(0.285)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES			
Observations	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576	66,576
AR(1) test (p-value)	0.154	0.289	0.278	0.070	0.478	0.135	0.171	0.441	0.285	0.159	0.311	0.303
AR(2) test (p-value)	0.933	0.890	0.727	0.661	0.598	0.904	0.814	0.379	0.810	0.856	0.751	0.637
Hansen-Sargan test (p-value)	0.337	0.201	0.091	0.067	0.158	0.081	0.201	0.148	0.102	0.526	0.332	0.160
Diff. Hansen-Sargan test (p-value)	0.147	0.125	0.063	0.130	0.087	0.085	0.168	0.074	0.205	0.146	0.163	0.057
Diff. Hansen-Sargan test (p-value), Y-lag set	0.252	0.201	0.091	0.010	0.078	0.043	0.465	0.051	0.317	0.403	0.394	0.059

Note: Panel A (B) [C] shows regression results for the effect of frailty on earnings (hours) [wages]. Frailty effects are the effect of incurring one additional deficit. All regressions include controls (marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, time dummies, and a quadratic in age). Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 13: Comparison With Different IV Lags, Workers' Earnings Regressions

IV Lags	4-5	5-6	5-7	4-5	5-6	5-7	4-5	5-6	5-7	4-5	5-6	5-7
Number of Instruments	22	22	24	28	28	32	25	25	28	25	25	28
frailty_t	-0.002	-0.029**	-0.030**									
	(0.013)	(0.014)	(0.014)									
$\text{frailty}_t \times$ HSD				-0.042*	-0.063**	-0.067**						
				(0.024)	(0.028)	(0.027)						
$\text{frailty}_t \times$ HS				-0.014	-0.038***	-0.038***						
				(0.012)	(0.015)	(0.014)						
$\text{frailty}_t \times$ CL				0.025*	-0.015	-0.012						
				(0.014)	(0.016)	(0.015)						
$\text{frailty}_t \times$ Good Health							0.022	-0.032	-0.027			
							(0.026)	(0.033)	(0.029)			
$\text{frailty}_t \times$ Poor Health							-0.014	-0.027**	-0.028**			
							(0.012)	(0.012)	(0.011)			
$\text{frailty}_t \times$ Young										0.023	-0.035	-0.037
										(0.019)	(0.026)	(0.026)
$\text{frailty}_t \times$ Old										-0.020**	-0.016	-0.017
										(0.010)	(0.013)	(0.012)
$\log(\text{earnings}_{t-1})$	-0.250	1.259***	1.180***	0.081	1.197***	0.989***	-0.107	1.215***	0.968***	0.191	1.050***	1.019***
	(0.434)	(0.455)	(0.409)	(0.278)	(0.358)	(0.290)	(0.393)	(0.400)	(0.302)	(0.230)	(0.307)	(0.272)
$\log(\text{earnings}_{t-2})$	0.864***	-0.427	-0.357	0.597***	-0.394	-0.213	0.760**	-0.392	-0.174	0.537***	-0.250	-0.227
	(0.334)	(0.411)	(0.363)	(0.218)	(0.323)	(0.259)	(0.304)	(0.362)	(0.270)	(0.183)	(0.282)	(0.249)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505
AR(1) test (p-value)	0.225	0.048	0.039	0.390	0.016	0.015	0.403	0.030	0.025	0.289	0.016	0.008
AR(2) test (p-value)	0.037	0.230	0.240	0.037	0.154	0.262	0.048	0.202	0.343	0.030	0.243	0.219
Hansen-Sargan test (p-value)	0.127	0.396	0.554	0.083	0.770	0.813	0.109	0.604	0.387	0.008	0.358	0.571
Diff. Hansen-Sargan test (p-value)	0.058	0.227	0.151	0.015	0.411	0.177	0.034	0.366	0.088	0.003	0.140	0.145
Diff. Hansen test (p-value), Y-lag set	0.127	0.396	0.429	0.009	0.330	0.232	0.040	0.383	0.170	0.002	0.223	0.281

Note: Panel A (B) [C] shows regression results for the effect of frailty on earnings (hours) [wages]. Frailty effects are the effect of incurring one additional deficit. All regressions include controls (marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, time dummies, and a quadratic in age). Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 14: Comparison With Different IV Lags, Workers' Hours Regressions

IV Lags	4-5	5-6	5-7	4-5	5-6	5-7	4-5	5-6	5-7	4-5	5-6	5-7
Number of Instruments	22	22	24	28	28	32	25	25	28	25	25	28
frailty_t	0.002	0.005	0.003									
	(0.008)	(0.009)	(0.009)									
$\text{frailty}_t \times$ HSD				0.001	-0.020*	-0.020*						
				(0.012)	(0.012)	(0.012)						
$\text{frailty}_t \times$ HS				0.002	0.001	-0.000						
				(0.008)	(0.010)	(0.010)						
$\text{frailty}_t \times$ CL				0.004	0.011	0.011						
				(0.008)	(0.009)	(0.009)						
$\text{frailty}_t \times$ Good Health							-0.014	0.013	0.010			
							(0.016)	(0.017)	(0.017)			
$\text{frailty}_t \times$ Poor Health							0.001	0.000	-0.001			
							(0.007)	(0.008)	(0.008)			
$\text{frailty}_t \times$ Young										0.014	0.014	0.013
										(0.013)	(0.016)	(0.016)
$\text{frailty}_t \times$ Old										-0.008	-0.003	-0.003
										(0.007)	(0.009)	(0.009)
$\log(\text{hours}_{t-1})$	-0.193	-0.086	0.114	-0.140	0.013	-0.006	-0.214	0.097	0.201	0.148	0.181	0.238
	(0.288)	(0.366)	(0.349)	(0.281)	(0.326)	(0.242)	(0.279)	(0.361)	(0.327)	(0.339)	(0.342)	(0.266)
$\log(\text{hours}_{t-2})$	0.440**	0.336	0.167	0.405**	0.191	0.071	0.444**	0.248	0.107	0.264	0.244	0.139
	(0.180)	(0.229)	(0.218)	(0.176)	(0.227)	(0.196)	(0.174)	(0.215)	(0.189)	(0.216)	(0.239)	(0.202)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505	34,505
AR(1) test (p-value)	0.432	0.495	0.298	0.414	0.354	0.172	0.419	0.353	0.175	0.347	0.300	0.108
AR(2) test (p-value)	0.056	0.218	0.615	0.078	0.489	0.659	0.042	0.429	0.814	0.478	0.531	0.751
Hansen-Sargan test (p-value)	0.136	0.206	0.148	0.372	0.375	0.344	0.363	0.325	0.274	0.014	0.163	0.179
Diff. Hansen-Sargan test (p-value)	0.377	0.404	0.050	0.841	0.333	0.098	0.143	0.380	0.089	0.006	0.391	0.222
Diff. Hansen-Sargan test (p-value), Y-lag set	0.136	0.206	0.096	0.168	0.077	0.033	0.197	0.131	0.069	0.007	0.098	0.068

Note: Panel A (B) [C] shows regression results for the effect of frailty on earnings (hours) [wages]. Frailty effects are the effect of incurring one additional deficit. All regressions include controls (marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, time dummies, and a quadratic in age). Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 15: Comparison With Different IV Lags, Workers' Wage Regressions

IV Lags	4-5	5-6	5-7	4-5	5-6	5-7	4-5	5-6	5-7	4-5	5-6	5-7
Number of Instruments	22	22	24	28	28	32	25	25	28	25	25	28
frailty_t	-0.009	-0.020**	-0.020**									
	(0.009)	(0.009)	(0.009)									
$\text{frailty}_t \times$ HSD				-0.037*	-0.065***	-0.062***						
				(0.021)	(0.021)	(0.020)						
$\text{frailty}_t \times$ HS				-0.017**	-0.029***	-0.028***						
				(0.008)	(0.010)	(0.010)						
$\text{frailty}_t \times$ CL				0.009	-0.005	-0.007						
				(0.012)	(0.010)	(0.009)						
$\text{frailty}_t \times$ Good Health							0.029	-0.009	-0.009			
							(0.019)	(0.023)	(0.023)			
$\text{frailty}_t \times$ Poor Health							-0.012	-0.021**	-0.022**			
							(0.008)	(0.008)	(0.008)			
$\text{frailty}_t \times$ Young										0.011	-0.015	-0.014
										(0.016)	(0.020)	(0.020)
$\text{frailty}_t \times$ Old										-0.016**	-0.022*	-0.021*
										(0.008)	(0.012)	(0.012)
$\log(\text{wage}_{t-1})$	0.662	0.205	0.252	0.587	0.160	0.427*	0.235	0.381	0.320	0.246	0.399	0.391
	(0.980)	(0.505)	(0.490)	(0.507)	(0.344)	(0.259)	(0.592)	(0.411)	(0.406)	(0.350)	(0.390)	(0.393)
$\log(\text{wage}_{t-2})$	0.142	0.543	0.501	0.194	0.570*	0.334	0.501	0.380	0.434	0.492*	0.365	0.374
	(0.831)	(0.454)	(0.442)	(0.424)	(0.307)	(0.238)	(0.499)	(0.368)	(0.365)	(0.294)	(0.351)	(0.354)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	34,401	34,401	34,401	34,401	34,401	34,401	34,401	34,401	34,401	34,401	34,401	34,401
AR(1) test (p-value)	0.636	0.629	0.605	0.403	0.486	0.212	0.664	0.463	0.496	0.460	0.429	0.438
AR(2) test (p-value)	0.998	0.408	0.446	0.908	0.190	0.438	0.504	0.543	0.451	0.267	0.549	0.537
Hansen-Sargan test (p-value)	0.072	0.079	0.230	0.169	0.292	0.403	0.388	0.140	0.269	0.090	0.106	0.239
Diff. Hansen-Sargan test (p-value)	0.031	0.046	0.087	0.061	0.160	0.334	0.211	0.156	0.234	0.079	0.257	0.348
Diff. Hansen-Sargan test (p-value), Y-lag set	0.072	0.079	0.143	0.045	0.066	0.082	0.164	0.061	0.133	0.061	0.060	0.130

Note: Panel A (B) [C] shows regression results for the effect of frailty on earnings (hours) [wages]. Frailty effects are the effect of incurring one additional deficit. All regressions include controls (marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, time dummies, and a quadratic in age). Standard errors are in parenthesis. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 16: Results of instrument power tests for system GMM estimation of employment on frailty

Panel A. Employment				
	employed_t	employed_t	Δ employed_t	Δ employed_t
Δ employed_{t-4}	0.117***	0.172***		
	(0.011)	(0.017)		
Δ employed_{t-5}		0.143***		
		(0.016)		
employed_{t-4}			-0.041***	-0.020*
			(0.006)	(0.011)
employed_{t-5}				-0.023*
				(0.012)
R^2	0.094	0.117	0.013	0.013
F-statistic	103.49	82.34	45.41	23.26
Panel B. Frailty				
	frailty_t	frailty_t	Δ frailty_t	Δ frailty_t
Δ frailty_{t-4}	0.591***	0.734***		
	(0.031)	(0.043)		
Δ frailty_{t-5}		0.794***		
		(0.047)		
frailty_{t-4}			0.013***	0.006
			(0.004)	(0.015)
frailty_{t-5}				0.001
				(0.016)
R^2	0.121	0.186	0.002	0.001
F-statistic	100.16	84.64	7.70	2.85

Note: F-statistics and R^2 values from OLS regressions of the endogenous variables in the main system GMM estimation (column (2) of Table 2 in the paper) on their instrument sets.

ference Hansen-Sargan test as suggested in Bond (2002a). The null hypothesis that changes in frailty are independent of initial unobserved heterogeneity cannot be rejected in any of our specifications at conventional confidence levels. We now provide more (visual) evidence in support of the validity of these assumptions.[1]

Figure 2 shows average frailty by education for the entire dynamic panel sample (Figure 2a) and the worker subsample (Figure 2b). Here, we think of education as a proxy for permanent income. Notice that in both panels the profiles of average frailty appear to be parallel up to age 55. In other words, the figures show little correlation between changes in frailty and education until very late in one's working life. Since permanent income is highly correlated with education, we take this a suggestive evidence that changes in frailty are not highly correlated with permanent income before age 55. While the frailty profiles are less parallel after age 55 this is not a problem for our dynamic panel estimations. Recall that our dynamic panel estimation results are based on a sample of individuals ages 25–64 meaning that the oldest individuals in our sample are 64 years old. We use the 4th and 5th lags of differenced frailty to instrument frailty in the employment regressions and the 5th and 6th lags of differenced frailty in the earnings, hours, and wage regressions. For this reason, all of the lagged first differences of frailty used as instruments are for individuals ages 56 or younger.

Next we use past employment, earnings, hours worked and hourly wages as a proxy for permanent income. The idea here is that average past earnings is highly correlated with individuals' fixed unobserved effects. The results are presented in Figure 3. Figure 3a sorts individuals by their employment in the past 10 years. We divide individuals into four groups: those who have been employed in each of the past 5 waves (10 years); those who have been employed more than 67 percent of time, but not always; those

[1] We thank an anonymous referee for suggesting these plots.

Table 17: Results of instrument power tests for system GMM estimation of earnings on frailty

Panel A. Employment				
	log(earnings$_t$)	log(earnings$_t$)	Δ log(earnings$_t$)	Δ log(earnings$_t$)
Δ log(earnings$_{t-5}$)	0.104***	0.172***		
	(0.023)	(0.036)		
Δ log(earnings$_{t-6}$)		0.202***		
		(0.032)		
log(earnings$_{t-5}$)			-0.015*	-0.025
			(0.008)	(0.024)
log(earnings$_{t-6}$)				0.010
				(0.021)
R^2	0.102	0.121	0.009	0.012
F-statistic	37.09	29.41	11.40	6.24
Panel B. Frailty				
	frailty$_t$	frailty$_t$	Δ frailty$_t$	Δ frailty$_t$
Δ frailty$_{t-5}$	0.520***	0.754***		
	(0.064)	(0.096)		
Δ frailty$_{t-6}$		0.679***		
		(0.097)		
frailty$_{t-5}$			0.015**	0.038
			(0.007)	(0.027)
frailty$_{t-6}$				-0.023
				(0.030)
R^2	0.096	0.150	0.003	0.004
F-statistic	31.90	24.49	4.45	2.53

Note: F-statistics and R^2 values from OLS regressions of the endogenous variables in the main system GMM estimation (column (2) of Table 2 in the paper) on their instrument sets.

who were employed more than 33 percent of time but less than 67 percent of time; and those who were employed less than 33 percent of time. The figure shows that up to age 55 the average frailty between these groups follows the same trend. However, for those older than 55 it appears that there is a decline in average frailty (or its growth) among those who were employed less than 67 percent of time. This decline is likely caused by selection due to mortality (evidence provided in Hosseini et al. (2022)) and, as we explain above, is not a problem for our dynamic panel estimation.

Figures 3b, 3c, and 3d shows profiles of average frailty in the worker subsample by past earnings, hours worked, and wages respectively. For example, in Figure 3b workers are divided into four groups according to the averages of their earnings over the past 10 years. The groups are simply the quartiles of the distribution of these average past earnings. We observe no differences in trends of frailty among these earnings groups. Again, this indicates that changes in frailty are not correlated with fixed factors that determine earnings. Figures 3c and 3d are constructed similarly and show a similar pattern for the relation between changes in frailty and quartiles of hours worked and hourly wages over the previous 10 years.

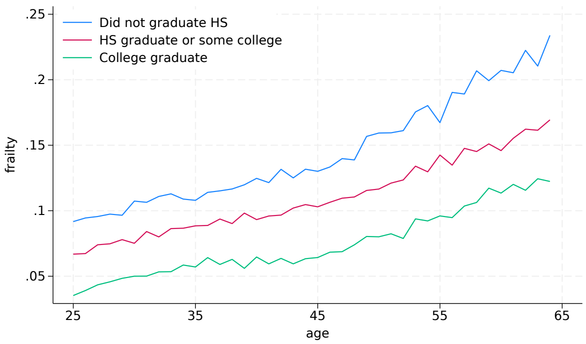
In summary, these plots, together with the results of the Difference Hansen-Sargan test, support the overidentifying assumption described above (Equation (3) in the main body).

2.6 Additional results on dynamic panel estimation

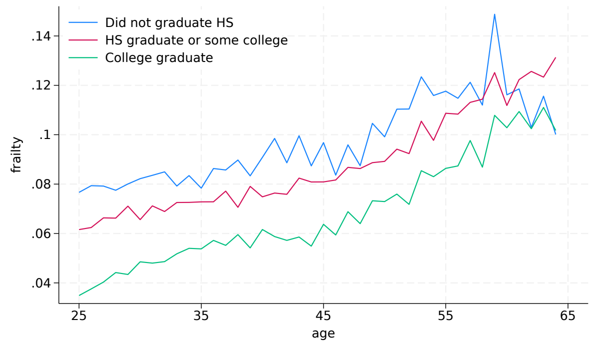
In this section we report additional results from the dynamic panel estimation.

2.6.1 Estimation results by gender

In Section 2 of the paper we estimate the impact of frailty on earnings for a sample of men and women and include controls for unobserved fixed heterogeneity and gender interacted with marital status and number

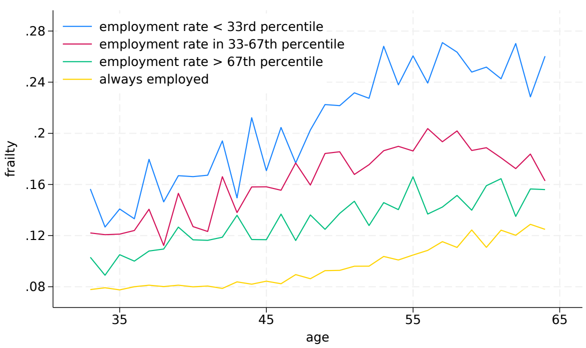


(a) Average frailty by age and education: all.

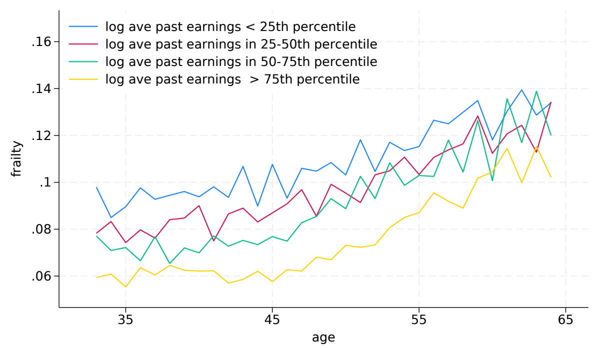


(b) Average frailty by age and education: workers only.

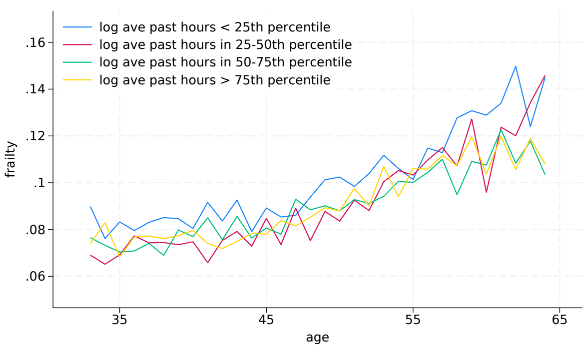
Figure 2: Average frailty by age and education for everyone in the dynamic panel sample (a) and for workers only (b).



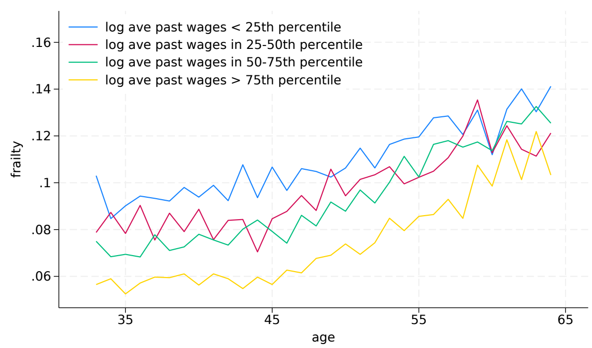
(a) Average frailty by age and past employment status: all.



(b) Average frailty by age and log average past earnings: workers only.



(c) Average frailty by age and log average past hours: workers only.



(d) Average frailty by age and log average past wages: workers only.

Figure 3: Average frailty by age and past employment, earnings, hours, and wages. Panel (a) is average frailty by employment in the past 10 years for everyone in the dynamic panel sample. 'Always employed' means the individual was employed in each of the previous 5 waves. Employment rates less than one are sorted into 3 percentile groups. Panels (b), (c), and (d) report average frailty by log average earnings, hours, and wages over the past 5 waves (10 years) for individuals in the worker sample only.

of kids. Here we report the results from estimating the impacts on separate subsamples by gender.

Table 18: Effect of Frailty on Employment, Men

	(1)	(2)	(3)	(4)
frailty_t	-0.016**			
	(0.008)			
$\text{frailty}_t \times$ HSD		-0.024***		
		(0.008)		
$\text{frailty}_t \times$ HS		-0.021**		
		(0.009)		
$\text{frailty}_t \times$ CL		-0.004		
		(0.008)		
$\text{frailty}_t \times$ Good Health			-0.028	
			(0.019)	
$\text{frailty}_t \times$ Poor Health			-0.019**	
			(0.008)	
$\text{frailty}_t \times$ Young				-0.020*
				(0.011)
$\text{frailty}_t \times$ Old				-0.015**
				(0.006)
employed_{t-1}	0.442	0.469	0.067	0.409
	(0.506)	(0.484)	(0.476)	(0.419)
employed_{t-2}	0.313	0.262	0.641	0.345
	(0.427)	(0.405)	(0.400)	(0.359)
Controls	YES	YES	YES	YES
Observations	29,877	29,877	29,877	29,877
AR(1) test (p-value)	0.478	0.426	0.650	0.419
AR(2) test (p-value)	0.697	0.761	0.226	0.582
Hansen-Sargan test (p-value)	0.397	0.268	0.308	0.737
Diff. Hansen-Sargan test (p-value)	0.237	0.198	0.367	0.439

Note: Frailty effects are the effect of incurring one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'HSD' is high school dropout, 'HS' is high school graduate, and 'CL' is college graduate. 'Good/Poor Health' is frailty below/above the 85th percentile. 'Young/Old' are individuals younger/older than 45 years of age. Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

First, Tables 18 and 19 report the estimation results for the effect of frailty on employment of men and women, respectively. Notice that for all the specifications, the estimated effects of frailty on employment are very similar for men and women (and very close to the estimations reported in Table 1 of the paper).

Second, Figures 20 and 21 report the estimation results for the effect of frailty on earnings, hours worked, and hourly wages of continuously working men and women, respectively. First, note that due to the smaller sample sizes the gender-specific estimation results are much noisier than those for all workers. This is especially true for men as most of the point estimates from these estimations are not statistically significant. Despite that fact, the results are generally consistent with those for the overall sample. The effects of one additional frailty deficit are larger for less educated individuals, and when significant, larger for those in poor health.

The effect of one additional deficit on workers' earnings tends to be larger for women as compared to men. Panel B and C show that this difference is largely driven by the impact of frailty on hours. While the estimated effects of frailty on hours are for the most part not significant, interestingly, all of the effects on hours are positive for men and vice versa for women. For example, one more deficit causes earnings of low educated working women to fall by more than 11 percent, compared to (an insignificant) 6 percent for men. This difference can be explained by differential responses in the hours of low educated men relative to low educated women. Accumulating one more deficit leads to a 2 percent (insignificant) rise in hours worked of men without a high school degree, conditional on employment. However, it causes

Table 19: Effect of Frailty on Employment, Women

	(1)	(2)	(3)	(4)
frailty_t	-0.019**			
	(0.008)			
$\text{frailty}_t \times$ HSD		-0.023**		
		(0.009)		
$\text{frailty}_t \times$ HS		-0.019**		
		(0.008)		
$\text{frailty}_t \times$ CL		-0.010		
		(0.007)		
$\text{frailty}_t \times$ Good Health			-0.002	
			(0.016)	
$\text{frailty}_t \times$ Poor Health			-0.021***	
			(0.008)	
$\text{frailty}_t \times$ Young				-0.010
				(0.009)
$\text{frailty}_t \times$ Old				-0.022***
				(0.007)
employed_{t-1}	0.277	0.378	-0.028	0.297
	(0.549)	(0.531)	(0.478)	(0.368)
employed_{t-2}	0.321	0.240	0.548	0.304
	(0.421)	(0.408)	(0.364)	(0.282)
Controls	YES	YES	YES	YES
Observations	36,699	36,699	36,699	36,699
AR(1) test (p-value)	0.588	0.507	0.671	0.402
AR(2) test (p-value)	0.637	0.769	0.243	0.518
Hansen-Sargan test (p-value)	0.419	0.028	0.285	0.555
Diff. Hansen-Sargan test (p-value)	0.300	0.005	0.282	0.349

Note: Frailty effects are the effect of incurring one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'HSD' is high school dropout, 'HS' is high school graduate, and 'CL' is college graduate. 'Good/Poor Health' is frailty below/above the 85th percentile. 'Young/Old' are individuals younger/older than 45 years of age. Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

a (significant) 4 percent reduction in hours worked by women in that education group. The results also indicate that declines in health may have a larger effect on the wages of female workers as compared to males. While the magnitudes of the wage effects within education are relatively similar by gender (and also similar to those in Table 2 in the paper), the gender-specific results suggest larger negative wage effects for older women and women in poor health relative to their male counterparts.

We find these results interesting and worthy of further analysis as they suggest another dimension by which men and women's labor market behavior may differ. However, it is our view that such an analysis falls outside the scope of this paper and we leave it for future projects.

2.6.2 The effect of spouses' frailty on employment and earnings

Table 22 shows the estimation results for the effect of one's own frailty and one's spouse's frailty on employment for married individuals. These results are obtained using our dynamic panel estimator and the same set of controls as used in the employment estimation reported in Table 1 of the paper. The first column shows the effect of incurring one additional deficit on the probability of being employed. Incurring one additional deficit reduces the probability of employment for a married individual by 2.0 percentage points. This is slightly larger than the 1.6 percentage point decline from incurring one additional deficit estimated for all individuals in the sample (reported in Table 1 of the paper).

Column 2 shows that when individuals' own frailty is not included in the regression equation, declines

Table 20: Effect of Frailty on Earnings, Hours and Wages of Male Workers

	Panel A. Earnings regression				**Panel B.** Hours regression				**Panel C.** Wage regression			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
frailty_t	-0.025				0.012				-0.012			
	(0.026)				(0.021)				(0.015)			
$\text{frailty}_t \times$ HSD		-0.060				0.022				-0.059**		
		(0.042)				(0.023)				(0.028)		
$\text{frailty}_t \times$ HS		-0.042*				0.010				-0.024*		
		(0.023)				(0.014)				(0.014)		
$\text{frailty}_t \times$ CL		-0.010				0.013				0.002		
		(0.028)				(0.014)				(0.018)		
$\text{frailty}_t \times$ Good Health			-0.040				0.014				-0.001	
			(0.076)				(0.052)				(0.052)	
$\text{frailty}_t \times$ Poor Health			-0.018				0.007				-0.014	
			(0.017)				(0.016)				(0.011)	
$\text{frailty}_t \times$ Young				-0.034				0.026				-0.007
				(0.047)				(0.035)				(0.035)
$\text{frailty}_t \times$ Old				0.008				0.020*				-0.010
				(0.020)				(0.012)				(0.016)
$\log(\text{earnings}_{t-1})$	1.174***	1.158***	1.110***	0.993***								
	(0.295)	(0.228)	(0.242)	(0.224)								
$\log(\text{earnings}_{t-2})$	-0.393	-0.408*	-0.325	-0.195								
	(0.314)	(0.241)	(0.260)	(0.228)								
$\log(\text{hours}_{t-1})$					0.989	0.753**	1.082*	0.569				
					(0.718)	(0.300)	(0.631)	(0.525)				
$\log(\text{hours}_{t-2})$					0.442	0.143	0.308	0.388				
					(0.495)	(0.321)	(0.484)	(0.389)				
$\log(\text{wage}_{t-1})$									0.855**	0.743***	0.934***	0.826***
									(0.372)	(0.277)	(0.259)	(0.275)
$\log(\text{wage}_{t-2})$									-0.051	0.024	-0.126	0.004
									(0.334)	(0.266)	(0.236)	(0.249)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	18,226	18,226	18,226	18,226	18,226	18,226	18,226	18,226	18,212	18,212	18,212	18,212
AR(1) test (p-value)	0.006	0.000	0.002	0.003	0.411	0.045	0.275	0.352	0.107	0.063	0.011	0.039
AR(2) test (p-value)	0.158	0.059	0.142	0.258	0.701	0.934	0.882	0.564	0.627	0.739	0.321	0.644
Hansen-Sargan test (p-value)	0.875	0.896	0.877	0.776	0.676	0.536	0.865	0.713	0.050	0.024	0.165	0.189
Diff. Hansen-Sargan test (p-value)	0.736	0.819	0.830	0.652	0.335	0.332	0.665	0.511	0.531	0.154	0.685	0.756
Diff. Hansen-Sargan test (p-value), Y-lag set	0.875	0.669	0.750	0.513	0.637	0.444	0.695	0.476	0.048	0.010	0.078	0.082

Note: All the frailty effects are reported as the effect of one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'HSD' is high school dropout, 'HS' is high school graduate, and 'CL' is college graduate. 'Good/Poor Health' is frailty below/above the 85th percentile. 'Young/Old' are individuals younger/older than 45 years of age. Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 21: Effect of Frailty on Earnings, Hours and Wages of Female Workers

	Panel A. Earnings regression				**Panel B.** Hours regression				**Panel C.** Wage regression			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
frailty_t	-0.045**				-0.011				-0.027**			
	(0.018)				(0.012)				(0.012)			
$\text{frailty}_t \times$ HSD		-0.112***				-0.041**				-0.063**		
		(0.040)				(0.016)				(0.027)		
$\text{frailty}_t \times$ HS		-0.054***				-0.011				-0.030**		
		(0.020)				(0.014)				(0.014)		
$\text{frailty}_t \times$ CL		-0.022				-0.002				-0.016		
		(0.015)				(0.012)				(0.010)		
$\text{frailty}_t \times$ Good Health			-0.059				-0.016				-0.050*	
			(0.046)				(0.031)				(0.029)	
$\text{frailty}_t \times$ Poor Health			-0.048***				-0.012				-0.026**	
			(0.018)				(0.013)				(0.012)	
$\text{frailty}_t \times$ Young				-0.044				-0.001				-0.024
				(0.031)				(0.021)				(0.021)
$\text{frailty}_t \times$ Old				-0.049***				-0.019				-0.033**
				(0.018)				(0.013)				(0.015)
$\log(\text{earnings}_{t-1})$	0.014	0.212	0.111	0.152								
	(0.493)	(0.429)	(0.461)	(0.338)								
$\log(\text{earnings}_{t-2})$	0.661	0.481	0.564	0.506*								
	(0.431)	(0.385)	(0.408)	(0.305)								
$\log(\text{hours}_{t-1})$					-0.258	-0.103	-0.326	0.161				
					(0.357)	(0.281)	(0.383)	(0.358)				
$\log(\text{hours}_{t-2})$					0.316	0.140	0.406	0.140				
					(0.281)	(0.289)	(0.264)	(0.302)				
$\log(\text{wage}_{t-1})$									0.388	0.321	0.444	0.167
									(0.303)	(0.270)	(0.291)	(0.310)
$\log(\text{wage}_{t-2})$									0.358	0.408	0.306	0.554*
									(0.279)	(0.250)	(0.262)	(0.285)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	16,279	16,279	16,279	16,279	16,279	16,279	16,279	16,279	16,189	16,189	16,189	16,189
AR(1) test (p-value)	0.571	0.549	0.584	0.459	0.574	0.351	0.597	0.310	0.298	0.283	0.233	0.416
AR(2) test (p-value)	0.223	0.358	0.284	0.203	0.231	0.576	0.120	0.828	0.511	0.351	0.606	0.195
Hansen-Sargan test (p-value)	0.129	0.309	0.248	0.130	0.625	0.428	0.785	0.330	0.148	0.394	0.249	0.134
Diff. Hansen-Sargan test (p-value)	0.102	0.156	0.222	0.060	0.644	0.424	0.777	0.693	0.196	0.269	0.286	0.040
Diff. Hansen-Sargan test (p-value), Y-lag set	0.129	0.068	0.134	0.052	0.625	0.278	0.586	0.141	0.148	0.128	0.157	0.071

Note: All the frailty effects are reported as the effect of one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'HSD' is high school dropout, 'HS' is high school graduate, and 'CL' is college graduate. 'Good/Poor Health' is frailty below/above the 85th percentile. 'Young/Old' are individuals younger/older than 45 years of age. Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

in an individual's spouse's frailty by one additional deficit reduce the probability of employment by 0.6 percentage points. Column 3 shows that when both individual's own frailty and their spouse's frailty are included, the effect of one's spouse's frailty on employment is no longer significant and the point estimate flips from positive to negative. At the same time, the effect of one's own frailty remains significant and becomes larger in magnitude. This result is interesting and deserves further investigation because it suggests that for some couples, increases in the frailty of one spouse may lead to an increase in the employment rate of the other. However, it is our view that it fall outside the scope of this paper and we leave it for future projects.

Table 23 shows the estimation results for the effect of one's own frailty and one's spouse's frailty on earnings, hours worked and hourly wages for the subsample of married workers. These estimations lack power and the estimation coefficients on frailty are not significant. This is likely due to the smaller sample size. However, looking at the magnitude of the estimated coefficients, the patterns are generally consistent with those found for employment in Table 22 and the main results in Section 2 of the paper. For example, in column 3 of Panel A, we see that when both own frailty and spouses frailty are included, the coefficient on own frailty is large (4.2 percent) and has a negative sign, while the coefficient on spouse's frailty is small and positive. Panel B, shows the effect of hours. Here, we don't expect to see any effect (consistent with the no-effect results we found using the full sample, presented in Table 2 of the main text). Finally, Panel C shows the effect on wages. Here, in column (3) we do not pass the instrument validity tests and thus the results are not informative.

Table 22: Effect of Spouse's Frailty on Employment

	(1)	(2)	(3)
frailty$_t$	-0.020*		-0.033*
	(0.010)		(0.020)
spouse's frailty$_t$		-0.006*	0.016
		(0.003)	(0.013)
employed$_{t-1}$	0.297	0.474	0.454
	(0.523)	(0.389)	(0.373)
employed$_{t-2}$	0.367	0.249	0.236
	(0.405)	(0.307)	(0.284)
Controls	YES	YES	YES
Observations	47,814	47,577	47,577
AR(1) test (p-value)	0.585	0.339	0.313
AR(2) test (p-value)	0.547	0.675	0.672
Hansen-Sargan test (p-value)	0.853	0.581	0.856
Diff. Hansen-Sargan test (p-value)	0.831	0.480	0.572

Note: Frailty effects are the effect of incurring one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'frailty sp' is the spouse's own frailty. Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

2.6.3 Effect of 'medical diagnosis' vs 'difficulties with Activities of Daily Living'

When we estimate the effect of frailty on employment and earnings, we use a frailty index that is constructed by taking averages of the 0/1 deficit variables listed in Table 1 in PSID. This list contains three groups of variables: those that indicate difficulties with Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs); medical diagnosis; and a third group with includes obesity (BMI of 30 and higher) and smoking. All these deficit variables have equal weight in our calculation.

One may wonder if equal weighting of deficits makes sense given that some deficits may, on average, be much more severe than others. In our earlier work, Hosseini et al. (2022), we show that properties of the frailty index are robust to various weighting schemes. In particular an index that is constructed using

Table 23: Effect of Spouse's Frailty on Earnings, Hours and Wages of Workers

	Panel A. Earnings regression			**Panel B.** Hours regression			**Panel C.** Wage regression		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
log(earnings$_{t-1}$)	0.899***	1.122**	0.909***						
	(0.325)	(0.444)	(0.315)						
log(earnings$_{t-2}$)	-0.159	-0.406	-0.163						
	(0.313)	(0.427)	(0.298)						
frailty$_t$	-0.013		-0.042	-0.017		0.035	-0.009		0.016
	(0.028)		(0.085)	(0.039)		(0.062)	(0.026)		(0.075)
frailty sp$_t$		-0.016	0.009		0.008	-0.014		-0.014*	-0.024
		(0.010)	(0.052)		(0.007)	(0.038)		(0.008)	(0.045)
log(hours$_{t-1}$)				1.532	-0.073	0.004			
				(1.269)	(0.376)	(0.483)			
log(hours$_{t-2}$)				-0.290	0.302	0.237			
				(0.565)	(0.261)	(0.296)			
log(wage$_{t-1}$)							0.570	0.076	0.168
							(0.611)	(0.504)	(0.420)
log(wage$_{t-2}$)							0.201	0.634	0.560
							(0.552)	(0.461)	(0.374)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	25,510	25,386	25,386	25,510	25,386	25,386	25,426	25,302	25,302
AR(1) test (p-value)	0.061	0.073	0.050	0.264	0.519	0.566	0.515	0.652	0.586
AR(2) test (p-value)	0.479	0.306	0.466	0.430	0.342	0.572	0.914	0.306	0.287
Hansen-Sargan test (p-value)	0.411	0.474	0.250	0.988	0.515	0.394	0.214	0.137	0.047
Diff. Hansen-Sargan (p-value)	0.238	0.353	0.094	0.989	0.381	0.422	0.136	0.079	0.019
Diff. Hansen-Sargan test (p-value), Y-lag set	0.411	0.474	0.161	0.988	0.515	0.260	0.214	0.137	0.039

Note: All the frailty effects are reported as the effect of one additional deficit. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. 'frailty sp' is the spouse's own frailty. Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

first principal component weighting of deficits or an index constructed using weights that are informed by subjective (self-reported) health status have similar power in predicting health-related outcomes and yield similar stochastic processes.

The reason for this is that incurring a more severe health deficit is positively correlated with incurring more health deficits and thus tends to lead to a larger increase in one's frailty. In particular, individuals who are newly diagnosed with more severe diseases/conditions tend to also incur more difficulties with (I)ADLs. At the same time, individuals who incur more severe (I)ADLs difficulties tend to also become newly diagnosed with more diseases/conditions. For example, an individual who is newly diagnosed with a stroke incurs, on average, an additional 2.2 (I)ADLs difficulties, more than five times the amount incurred by an individual who is newly diagnosed with high blood pressure (an additional 0.4 (I)ADLs difficulties). Likewise, an individual who becomes unable to eat without assistance is diagnosed with an additional 3.6 diseases/conditions while an individual who starts to have difficulty managing money is diagnosed with an additional 2.9 diseases/conditions.

Consistent with this logic, the effects of additional medical diagnoses on employment are larger for individuals who also incur (I)ADLs and likewise the effects of incurring (I)ADLs are larger for individuals who also incur relatively more medical diagnoses as illustrated in Table 24. The first column shows that, on average, one additional medical diagnosis reduces the probability of employment by 0.7 percentage points while one additional (I)ADL difficulty reduces it by 3.3 percentage points. Column (2) shows that the effect of an additional medical diagnosis is more than five times larger if it coincides with a change in (I)ADL difficulties versus if it does not. Column (3) shows that the effect of an additional (I)ADL difficulty is three times larger if is coincides with an above median change in the number of medical diagnoses. These claims are based on the point estimates which are robust to variations in the instrument set despite not all being significant.

To summarize, severity of health deficits is correlated with the number of health deficits incurred. So, even though each health deficit has the same weight in the frailty index, more severe health events tend to lead to a larger increase in frailty and thus a larger impact on earnings and employment.

3 Additional Information about the Structural Model

3.1 Recursive competitive equilibrium

In this section we provide the definition of equilibrium we employ in the structural model economy. Let $\{\mu^E(x,i_s), \mu^N(x,n_a), \mu^D(x,n_d), \mu^R(x)\}$ represent the time-invariant measures of individuals. We assume that these are the population measures *after* the labor participation decisions and DI application decisions are made. The concept of a stationary recursive competitive equilibrium can be defined as follows.

Given a fiscal policy $\{Tr^E(\cdot), Tr^{OE}(\cdot), Tr^N(\cdot), Tr^D(\cdot), Tr^R(\cdot), SS(\cdot), T(\cdot), \tau_K\}$, a *stationary recursive competitive equilibrium* is a set of value functions $\left\{V^E(x,i_s), V^N(x,n_a),\ V^D(x,n_d), V^R(x)\right\}$, households' consumption decisions $\{c^E(x,i_s), c^N(x,n_a), c^D(x,n_d), c^R(x)\}$, saving decisions $\{a^E(x,i_s), a^N(x,n_a), a^D(x,n_d), a^R(x)\}$, labor force participation decisions $I_E(x,i_s)$ and $I_N(x,n_a)$, DI application decisions $I_D(x,n_a)$; prices of labor and capital $\{w,r\}$; government expenditures G; and time-invariant measures of households $\left\{\mu^E(x,i_s), \mu^N(x,n_a),\ \mu^D(x,n_d), \mu^R(x)\right\}$ such that:

1. Given the fiscal policy and prices, households' decision rules solve the households' decision problems in Section 3.5 in the paper.

2. Rental rate r is exogenously given by $r = (1-\tau_k)\left(\alpha A(K/N)^{\alpha-1} - \delta\right)$, and the wage is given by equation $w = (1-\alpha)A(K/N)^{\alpha}$.

Table 24: Effect of Medical Diagnoses and Difficulties with Activities of Daily Living on Employment

	(1)	(2)	(3)
medical diagnoses$_t$	-0.007*		
	(0.004)		
(I)ADL difficulties$_t$	-0.033**		
	(0.014)		
medical diagnoses$_t$ $\times$ No Δ(I)ADL difficulties		-0.008	
		(0.013)	
medical diagnoses$_t$ $\times$ Δ(I)ADL difficulties		-0.046	
		(0.041)	
(I)ADL difficulties$_t$ $\times$ Below median Δmedical diagnoses			-0.015
			(0.033)
(I)ADL difficulties$_t$ $\times$ Above median Δmedical diagnoses			-0.045**
			(0.022)
employed$_{t-1}$	0.573**	0.879***	0.575
	(0.252)	(0.284)	(0.369)
employed$_{t-2}$	0.115	-0.116	0.135
	(0.198)	(0.236)	(0.289)
Controls	YES	YES	YES
Observations	66,584	58,881	58,852
AR(1) test (p-value)	0.061	0.019	0.204
AR(2) test (p-value)	0.962	0.312	0.986
Hansen test (p-value)	0.262	0.492	0.882
Diff-in-Hansen test (p-value)	0.144	0.319	0.921
Diff-in-Hansen test (p-value), Y-lag set	0.122	0.827	0.776

Note: 'medical diagnoses$_t$' is the number of medical diagnoses an individual has at time t. '(I)ADL difficulties$_t$' is the number of activities of daily living and instrumental activities of daily living, (I)ADLs, a person has difficulty with at time t. All the medical diagnoses and (I)ADL difficulties effects are reported as the effect of one additional deficit. 'No Δ(I)ADL difficulties' equals 1 if there is no change in the number of (I)ADLs a person has difficulty with between time $t-1$ and t. 'Δ(I)ADL difficulties' equals 1 if there is a change in the number of (I)ADL difficulties a person has difficulty with between time $t-1$ and time t. 'Below/Above median Δmedical diagnoses' equals 1 if the age-adjusted change in medical diagnosis from t to $t+1$ is above/below the median level. Controls included are marital status, marital status interacted with gender, number of kids, number of kids interacted with gender, year dummies, and a fourth degree polynomial in age. Standard errors are in parenthesis. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

3. Aggregate labor and capital input satisfy:

$$N = \sum_{\{x,i_s\}} \eta(x)\, \mu^E(x, i_s) \text{ and } r = (1-\tau_K)\left(\alpha A (K/N)^{\alpha-1} - \delta\right).$$

4. Government expenditures G are such that the government's budget constraint holds:

$$\begin{aligned} \sum_{\{x,i_s\}} T(w\eta(x))\, \mu^E(x,i_s) + \tau_K\left(\alpha A (K/N)^{\alpha-1} - \delta\right) = G \\ + \sum_{\{x,n_d\}} \left(\mu^D(x,n_d) + \mu^R(x)\right) SS(\bar{e}) \\ + \sum_{\{x,i_s\}} (\mathbf{1}_{j<R} Tr^E(x,i_s) + \mathbf{1}_{j\geq R} Tr^{OE}(x,i_s))\mu^E(x,i_s) \\ + \sum_{\{x,n_a\}} Tr^N(x)\, \mu^N(x,n_d) \\ + \sum_{\{x,n_d\}} \left(Tr^D(x,n_d)\, \mu^D(x,n_d) + Tr^R(x)\, \mu^R(x)\right). \end{aligned}$$

5. The measures $\{\mu^E(x,i_s),\mu^N(x,n_a),\mu^D(x,n_d),\mu^R(x)\}$ are stationary.

   (a) Employed:

$$\mu^E(x',0)=\frac{I_E(x',0)}{1+\nu}\sum_{\{x,i_s\}}(1-\sigma)\,p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^E(x,i_s)}\mu^E(x,i_s),$$

   and

$$\mu^E(x',1)=\frac{I_E(x',1)}{1+\nu}\sum_{\{x,i_s\}}\sigma p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^E(x,i_s)}\mu^E(x,i_s)$$
$$+\frac{\varphi}{1+\nu}\sum_{\{x,n_a\}}I_D(x,n_a)(1-\theta(x))\,p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^N(x,n_a)}I_N(x',n_a+1)\,\mu^N(x,n_a)$$
$$+\frac{\varphi}{1+\nu}\sum_{\{x,n_a\}}(1-I_D(x,n_a))p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^N(x,n_a)}I_N(x',0)\,\mu^N(x,n_a).$$

   (b) Non-employed:

$$\mu^N(x',0)=\frac{1-I_E(x',0)}{1+\nu}\sum_{\{x,i_s\}}(1-\sigma)\,p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^E(x,i_s)}\mu^E(x,i_s)$$
$$+\frac{1-I_E(x',1)}{1+\nu}\sum_{\{x,i_s\}}\sigma p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)$$
$$+\frac{\varphi}{1+\nu}\sum_{\{x,n_a\}}(1-I_D(x,n_a))p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^N(x,n_a)}(1-I_N(x',0))\mu^N(x,n_a)$$
$$+\frac{1-\varphi}{1+\nu}\sum_{\{x,n_a\}}(1-I_D(x,n_a))p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^N(x,n_a)}\mu^N(x,n_a)\,\mathbf{1}_{a'=a^E(x,i_s)}\mu^E(x,i_s),$$

   and

$$\mu^N(x',n_a+1)=\frac{\varphi}{1+\nu}\sum_{\{x,n_a\}}I_D(x,n_a)(1-\theta(x))\,p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)$$
$$\times\mathbf{1}_{a'=a^N(x,n_a)}(1-I_N(x',n_a+1))\mu^N(x,n_a)$$
$$+\frac{1-\varphi}{1+\nu}\sum_{\{x,n_a\}}I_D(x,n_a)(1-\theta(x))\,p(x)\,\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^N(x,n_a)}\mu^N(x,n_a).$$

   (c) DI beneficiary:

$$\mu^D(x',0)=\frac{1}{1+\nu}\sum_{\{x,n_a\}}I_D(x,n_a)\,\theta(x)\,p(x)\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^N(x,i_s)}\mu^N(x,n_a),$$

   and

$$\mu^D(x',n_d+1)=\frac{1}{1+\nu}\sum_{\{x\}}p(x)\pi^e(\epsilon'|x)\,\pi^f(f'|x)\,\mathbf{1}_{a'=a^D(x,n_d)}\mu^D(x,n_d).$$

   (d) Retiree:

When $j = R - 1$,

$$
\begin{aligned}
\mu^R(x') = &\frac{1}{1+\nu} \sum_{\{x,n_d\}} p(x)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^D(x,i_s)}\mu^D\left(x,n_d\right) \\
&+ \frac{1-I_E(x',1)}{1+\nu} \sum_{\{x,i_s\}} (1-\sigma)\,p\left(x\right)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^E(x,i_s)}\mu^E\left(x,i_s\right) \\
&+ \frac{1-I_E(x',0)}{1+\nu} \sum_{\{x,i_s\}} \sigma p\left(x\right)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^E(x,i_s)}\mu^E\left(x,i_s\right) \\
&+ \frac{\varphi}{1+\nu} \sum_{\{x,n_a\}} p\left(x\right)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^N(x,i_s)}\left(1-I_N\left(x',n_a+1\right)\right)\mu^N\left(x,n_a\right) \\
&+ \frac{1-\varphi}{1+\nu} \sum_{\{x,n_a\}} p\left(x\right)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^N(x,i_s)}\mu^N\left(x,n_a\right).
\end{aligned}
$$

When $j > R - 1$,

$$
\begin{aligned}
\mu^R(x') = &\frac{1}{1+\nu} \sum_{\{x\}} p\left(x\right)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^R(x)}\mu^R\left(x\right) \\
&+ \frac{1-I_E(x',1)}{1+\nu} \sum_{\{x,i_s\}} (1-\sigma)\,p\left(x\right)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^E(x,i_s)}\mu^E\left(x,i_s\right) \\
&+ \frac{1-I_E(x',0)}{1+\nu} \sum_{\{x,i_s\}} \sigma p\left(x\right)\pi^e\left(\epsilon'|x\right)\pi^f\left(f'|x\right)\mathbf{1}_{a'=a^E(x,i_s)}\mu^E\left(x,i_s\right).
\end{aligned}
$$

4 Additional Calibration Details

This section includes additional information, moments, and estimation results related to the calibration of the structural model as described in Section 4 of the paper. Figure 1 in the paper (the left-panel of Figure 3 shows the same data moments) is contructed using data from MEPS. MEPS contains information on SSI beneficiary status but not SSDI. However, it does include data on whether an individual receives Medicare benefits. SSDI and some SSI beneficiaries are the only groups younger than 65 years of age who qualify for Medicare (after being on SSDI for two years). We compute the fraction in each frailty and age group who receive SSI or Medicare benefits in MEPS. We then adjust the fractions by the Social Security Administration population data on SSDI and SSI recipient by age so that the numbers are consistent on aggregate.

Table 25 shows the distribution of 24- to 26-year-old males across employment states by education and frailty percentile group in the data. This distribution is used as the initial distribution of individuals across employment states in the model. Individuals who start out as DI beneficiaries are assumed to be severely disabled and have permanent zero productivity. To calibrate the survival probability rates in the model we predict conditional survival rates for men using the estimation results in Table 26. The table shows results from the estimation of mortality probits using HRS data. We adjust the value of the estimated constant so that population mortality is consistent with the year 2000 period life-table in Bell and Miller (2005).

We consider two specifications for the log productivity profiles in the model: one with linear frailty effects and one with quadratic frailty effects. For both specifications, we estimate a process for log wages (our proxy for labor productivity) of men using our dynamic panel estimator. Since we only observe wages of men who work, we employ a selection correction procedure in the estimation. Given that we did not find evidence that earnings impacts frailty, we treat frailty as exogenous. However, as we show below, the estimated effects of frailty on men's labor productivity are similar if we instead assume endogenous frailty.

Table 25: Distribution (%) of 24- to 26-year-old males across employment states by education and frailty percentile group

	Frailty percentile group				
	0–50th	50–70th	70–90th	90–95th	95–100th
High-school dropouts					
Employed	74.1	76.5	70.4	60.0	65.6
Non-employed	24.9	21.8	25.1	30.3	0.00
DI beneficiary	1.0	1.7	4.5	9.7	34.4
High-school graduates					
Employed	92.7	91.4	90.7	90.1	78.2
Non-employed	7.0	7.3	7.7	2.0	0.0
DI beneficiary	0.4	1.4	1.7	8.0	21.8
College graduates					
Employed	95.5	94.9	96.5	88.9	96.9
Non-employed	4.5	5.1	3.5	11.1	0.0
DI beneficiary	0.0	0.0	0.0	0.0	3.1

Note: Percent breakdown of 25- to 26-year-old males by employment state for each education and frailty percentile group. Authors' calculations using PSID, MEPS, and SSA data.

To conduct the selection correction in the first step, we follow Al-Sadoon et al. (2019) who show that in system GMM, selection bias is mainly due to correlation of the fixed effects in the selection and outcome processes. To correct for this selection bias they propose first estimating a selection equation that includes fixed effects and an exclusion restriction.[2] They then suggest including the estimated fixed effects as regressors in the outcome equation.

We use a fixed effect linear probability model of employment as our selection equation. We include the same set of regressors in our selection equation as in the outcome equation with the addition of exclusion restrictions. Following Low and Pistaferri (2015), we use "potential" government transfers interacted with frailty as our exclusion restrictions. "Potential" government transfers are defined as the sum of food stamps, AFDC/TANF payments, unemployment insurance benefits, and EITC payments that individuals would receive if they applied. These transfers vary across states. Moreover, they depend on marital status and number of kids (which also vary across individuals in our sample). Therefore, instead of using the amount of "potential" transfers as exclusion restrictions we use the interaction of state of residence, number of kids and martial status (a total of 436 combinations). These "potential" transfers do not directly impact individuals' wages or labor productivity but do create different work incentives for people with different frailty levels.

Table 27 shows the estimation results from this first stage for both the linear and quadratic frailty specifications. Table 28 shows the estimation results from both of the log productivity specifications both with and without the selection correction. The frailty effects are reported as the effect of one additional deficit. Notice that the coefficients on the high school dropout and high school graduate quadratic frailty terms have the wrong sign. The quadratic results for the high school dropout and high school graduate groups imply that productivity declines in frailty at lower values of frailty and increases in frailty at higher values. The increases occur at values of frailty above the 95th percentile of the frailty distribution and are, thus, estimated off a very small number of observations. The rate of decline in log productivity with frailty in both cases is very similar under the linear and quadratic specifications. In contrast, the coefficient on the quadratic term in the college regression is strongly significant, has the correct sign, and implies that frailty declines starting from the 76th percentile. Given these findings, we use the linear specification for the non-college groups and the quadratic specification for the college graduate group.

[2] An exclusion restriction is a variable that impacts the selection process but not the outcome process. In our case, a valid exclusion restriction should impact employment but not labor productivity or wages.

Table 26: Estimation of mortality probit

	Coefficient estimates
frailty	3.001***
	(0.098)
frailty^2	-0.500***
	(0.120)
age	-0.021***
	(0.005)
age^2	3.91e-4***
	(3.45e-5)
education (years)	0.003
	(0.002)
male	0.308***
	(0.010)
constant	-2.919***
	(0.183)
year fixed effects	included
Observations	213,114
Pseudo R^2	0.239

Note: Frailty is the fraction of deficit variables that the individual reports as a deficit and lies between 0 and 1. Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Estimated using 1998–2016 HRS data.

The results are robust to whether frailty is treated as exogenous or endogenous. Table 29 shows this for the linear specification.

Given these estimation results, we next remove the effects of frailty from our log wage observations. Then, to obtain the deterministic age effects, separately for each education group, we regress the adjusted log wages on a cubic polynomial in age and year dummies. We run these estimations both in our PSID sample of men and in a sample of men from the HRS. We splice together the estimated age profiles from the PSID and HRS to create the age profile we feed into the model. This is done to take advantage of the larger sample of older workers in the HRS relative to PSID. Given the sample sizes, we switch from the PSID to the HRS age profile at age 70 for non-college and age 65 for college.

Finally, we use the residuals from this regression to construct a set of variance-covariance moments. As in Guvenen (2009), we use these moments to estimate the stochastic component via GMM. For these last two steps, we group together high school dropouts and high school graduates in order to take advantage of larger sample sizes. The implication of this is that, while the effect of frailty on productivity is education group specific, the age effects and stochastic component are the same for both high school dropouts and high school graduates. Table 30 shows the estimation results from these steps.

Table 31 shows the results of the estimation of the zero frailty probits used in the model. Table 32 shows the results from the estimation of the nonzero frailty processes for each education group. Figures 4 and 5 are the counterparts for the high school dropout and college graduate groups to Figure 2 in the paper which shows the high school graduates. These estimations are done using the full PSID sample. Table 33 shows the results of the men's out-of-pocket medical expenditures estimation for each education group and type. These estimations are done using MEPS data on a sample of men only.

4.1 Comparing frailty effects on wages to Low and Pistaferri (2015)

Low and Pistaferri (2015) estimate the effect of *disability status* on wage for non-college men. They use self-reported responses in PSID to define three disability groups: no work limitations, moderate work limitations and severe work limitations. As the second column of Table 34 shows, they find that having

Table 27: Estimation of the effect of frailty on labor productivity for men. Stage 1: Selection equation regressions.

	Linear frailty effect	Quadratic frailty effect
$\text{frailty}_t \times$ HSD (one add. deficit)	-0.032	-0.050
	(0.216)	(0.216)
$\text{frailty}_t \times$ HSG (one add. deficit)	-0.031	-0.040
	(0.216)	(0.216)
$\text{frailty}_t \times$ CL (one add. deficit)	-0.019	-0.016
	(0.216)	(0.216)
${\text{frailty}_t}^2 \times$ HSD		0.000
		(0.000)
${\text{frailty}_t}^2 \times$ HSG		-0.000
		(0.000)
${\text{frailty}_t}^2 \times$ CL		-0.001***
		(0.000)
age	0.038***	0.039***
	(0.004)	(0.004)
age^2	-0.000***	-0.000***
	(0.000)	(0.000)
exclusion restrictions	total of 436 combinations	
joint p-value	0.000	0.000

Note: The left-hand-side variable is employment (1 if employed, 0 otherwise). The frailty effects are the effect of one additional deficit. Standard errors are in parenthesis. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

a moderate (severe) work limitation reduces wages by 5.7% (17.7%) relative to no work limitation.

To compare our estimates with theirs, we sort individuals in our PSID sample into the same three groups and calculate mean frailty by group. We then calculate the effects on wages of moving from the mean frailty of the no work limitation group to that of both the moderate and severe group. Mean frailty for each group is reported in column 1 of Table 34. Wage effects are reported in columns 3–6. We find that moderate (severe) work limitations reduce non-college wages by 12.9% (22.5%). The effects are larger than those found by Low and Pistaferri (2015). We also find that there is substantial variation in the effect within the non-college group. Workers with less than a high school degree experience significantly larger wage effects, on average, as compared to those with a high school degree. Effects, in general, decline with education. However, even college workers experience significant wage declines when they have severe declines in their health.

4.2 Identification and Sensitivity of Calibrated Parameters

Table 35 summarizes the parameters set directly using data, and Table 36 summarizes parameters calibrated using the model and their associated targeted moments. More specifically, the parameters in Table 36 target the following 107 moments: federal income tax receipts as a share of GDP (1 moment), the wealth-to-earnings ratio (1 moment), employment rates by frailty percentile group and age (5×12 moments), DI recipiency rates by frailty percentile group and age (5×8 moments), DI application success rates after $n_a = 0, 1, 2, 3$ previous tries (4 moments), and the share of DI applicants in the 25- to 64-year-old population (1 moment). The first two moments are associated with the discount factor β and the HSV tax parameter λ. The third set of 60 moments is used to determine the disutility parameters ϕ_0, ϕ_1, and ϕ_2. As discussed in Section 4 of the main text, these parameters are identified by the level and variation in employment by frailty after age 65. After this age, the DI application and success rates

Table 28: Estimating the effect of frailty on labor productivity for men. Stage 2: Bias correction.

	Linear frailty effect		Quadratic frailty effect	
	no correction	correction	no correction	correction
$\log(\text{wage}_{t-1})$	0.796***	0.841***	0.770***	0.814***
	(0.279)	(0.297)	(0.278)	(0.294)
$\log(\text{wage}_{t-2})$	-0.047	-0.119	-0.036	-0.105
	(0.243)	(0.264)	(0.240)	(0.261)
$\text{frailty}_t \times$ HSD (one add. deficit)	-0.044***	-0.048***	-0.089***	-0.097***
	(0.015)	(0.015)	(0.027)	(0.027)
$\text{frailty}_t \times$ HSG (one add. deficit)	-0.026***	-0.029***	-0.045***	-0.050***
	(0.008)	(0.008)	(0.015)	(0.015)
$\text{frailty}_t \times$ CL (one add. deficit)	0.006	0.006	0.023**	0.023**
	(0.004)	(0.005)	(0.009)	(0.010)
$\text{frailty}_t^2 \times$ HSD			0.007***	0.008***
			(0.002)	(0.002)
$\text{frailty}_t^2 \times$ HSG			0.003**	0.004**
			(0.001)	(0.001)
$\text{frailty}_t^2 \times$ CL			-0.004***	-0.004***
			(0.001)	(0.001)
selection term		0.053		0.081
		(0.077)		(0.072)
Observations	18,212	18,117	18,212	18,117
AR(1) test (p-value)	0.036	0.032	0.039	0.035
AR(2) test (p-value)	0.538	0.403	0.567	0.430
Hansen-Sargan test (p-value)	0.046	0.085	0.050	0.095
Diff. Hansen-Sargan test (p-value)	0.254	0.197	0.203	0.166
Diff. Hansen-Sargan test, Y-lag set	0.163	0.183	0.163	0.178

Note: The left-hand-side variable is log wage. The selection term is the predicted fixed effects from the regression in stage 1. The frailty effects are the effects of incurring one additional deficit. Standard errors are in parenthesis. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

(and their variation with frailty) do not have a direct impact on labor supply.

We are left with 45 moments: the DI recipiency rates by age and frailty, the average DI application success rates after each try, and the share of DI applicants in the population. We have 8 parameters to match these moments. Seven parameters associated with the probability of successful DI application,

$$\theta\left(j, f, n_a\right) = \begin{cases} \min\left\{1, \left(1 + \mathbf{1}_{\{j \geq 55\}}\left(\varrho - 1\right)\right)\vartheta\left(n_a\right) f^{\kappa}\right\}, & \text{if } f > \underline{f}, \\ 0, & \text{otherwise,} \end{cases}$$

where κ is the elasticity of the success rate with respect to frailty, $\underline{f}$ is the lowest value of frailty that has a chance of successful application, $\vartheta(n_a)$ (for $n_a = 0, 1, 2, 3$) are the loading factors after n_a previous tries, and ϱ is the loading factor after age 55. One parameter, ξ, which is the utility cost of application at first try.

The model is overidentified and variation in any of the parameters tends to affect all moments simultaneously (and vise versa). To help assess the sources of identification of these 8 parameters, following recommendations in Andrews et al. (2017), we report local elasticities (computed at the baseline calibration) of each of the 45 targeted moments with respect to each of the 8 parameters in Figure 6.

Consider first the loading factors on the probabilities of successfully obtaining DI, $\vartheta(n_a)$ (for $n_a = 0, 1, 2, 3$). Ignoring the cutoffs $\underline{f}$ and 1, counterparts to the overall DI success rates (the French and Song

Table 29: Estimating the effect of frailty on labor productivity for men: endogenous versus exogenous frailty.

	Exogenous frailty		Endogenous frailty	
	no correction	correction	no correction	correction
log(wage$_{t-1}$)	0.796***	0.841***	0.743***	0.724***
	(0.279)	(0.297)	(0.277)	(0.248)
log(wage$_{t-2}$)	-0.047	-0.119	0.024	0.028
	(0.243)	(0.264)	(0.266)	(0.249)
frailty$_t$ × HSD (one add. deficit)	-0.044***	-0.048***	-0.059**	-0.052*
	(0.015)	(0.015)	(0.028)	(0.028)
frailty$_t$ × HSG (one add. deficit)	-0.026***	-0.029***	-0.024*	-0.022
	(0.008)	(0.008)	(0.014)	(0.017)
frailty$_t$ × CL (one add. deficit)	0.006	0.006	0.002	0.005
	(0.004)	(0.005)	(0.018)	(0.019)
selection term		0.053		-0.010
		(0.077)		(0.304)
Observations	18,212	18,117	18,212	18,117
AR(1) test (p-value)	0.036	0.032	0.063	0.042
AR(2) test (p-value)	0.538	0.403	0.739	0.716
Hansen-Sargan test (p-value)	0.046	0.085	0.024	0.024
Diff. Hansen-Sargan test (p-value)	0.254	0.197	0.154	0.012
Diff. Hansen-Sargan test, Y-lag set	0.163	0.183	0.010	0.005

* $p < .1$, ** $p < .05$, *** $p < .01$

(2014) moments) can be calculated as

$$FS(n_a) = \frac{\sum_x \mu^N(\tilde{x}) I^D(\tilde{x}) \left(1 + \mathbf{1}_{\{j \geq 55\}} (\varrho - 1)\right) \vartheta(n_a) f_j^{\kappa}}{\sum_x \mu^N(\tilde{x}) I^D(\tilde{x})},$$

where $\tilde{x} \equiv (x, n_a) \equiv (j, a, f_j, s, \epsilon, \bar{e}, n_a)$ summarizes the state vector at age j including both $x \equiv (j, a, f_j, s, \epsilon, \bar{e})$ and n_a, $\mu^N(\tilde{x})$ is the measure of non-employed and $I_D(\tilde{x})$ is the indicator for decision to apply. Note that the summations are done over x, i.e., all the components of state space except the number of previous DI applications, n_a. The denominator is the measure of individuals with $n_a = 0, 1, 2$ or 3 previous tries who apply. The numerator is the subset of these individuals who become DI recipients.

The relative success rate after $n_a = 1, 2$ or 3 previous tries can be written as the ratio of the success rate after n_a previous tries to the success rate after $n_a = 0$ previous tries,

$$\frac{FS(n_a)}{FS(n_a = 0)} = \left(\frac{\frac{\sum_x \mu^N(\tilde{x}) I^D(\tilde{x}) \left(1 + \mathbf{1}_{\{j \geq 55\}} (\varrho - 1)\right) f_j^{\kappa}}{\sum_x \mu^N(\tilde{x}) I^D(\tilde{x})}}{\frac{\sum_x \mu^N(x,0) I^D(x,0) \left(1 + \mathbf{1}_{\{j \geq 55\}} (\varrho - 1)\right) f_j^{\kappa}}{\sum_x \mu^N(x,0) I^D(x,0)}} \right) \frac{\vartheta(n_a)}{\vartheta(0)}. \tag{5}$$

Equation 5 illustrates that the ratio $\frac{\vartheta(n_a)}{\vartheta(0)}$ has a direct effect on the relative DI application success rate after n_a tries. Note that the big term in parenthesis can also potentially vary with $\vartheta(n_a)$, but only through the effect that $\vartheta(n_a)$ has on the equilibrium distribution of individuals across the state space and the decision to apply for DI. Consistent with this expression, Figures 6b and 6d show that the relative success rates after $n_a = 1, 2, 3$ tries are especially sensitive to changes in the three loading factors, $\vartheta(n_a)$ for $n_a = 1, 2, 3$. In addition, the success rate after n_a previous tries increases the most in response to

Table 30: Estimation of labor productivity process after bias correction and removal of frailty effect

(a) Estimation results for deterministic component of wages (net of frailty effect).

	Ages 25-69 Non-college	Ages 25-64 College	Ages 70+ Non-college	Ages 65+ College
age	0.523***	0.540***	2.071	3.911
	(0.0938)	(0.130)	(1.465)	(2.647)
age^2	-0.0162***	-0.0154***	-0.0387	-0.0868
	(0.00309)	(0.00427)	(0.0332)	(0.0600)
age^3	0.000226***	0.000201***	0.000303	0.000845
	(4.38e-05)	(6.06e-05)	(0.000332)	(0.000601)
age^4	-1.18e-06***	-1.01e-06***	-8.38e-07	-3.06e-06
	(2.27e-07)	(3.14e-07)	(1.24e-06)	(2.24e-06)
constant	-3.489***	-3.839***	-36.88	-61.92
	(1.038)	(1.443)	(24.05)	(43.44)
year fixed effects	yes	yes	yes	yes
	17,078	11,465	13,286	6,144
R^2	0.034	0.059	0.042	0.019

(b) Estimation results for stochastic component of wages.

	Non-college	Col Graduates
var. of transitory shock	0.1164	0.1080
	(0.0073)	(0.0125)
var. of permanent shock	0.0055	0.0186
	(0.0034)	(0.0058)
var. of fixed effect	0.1164	0.1179
	(0.0157)	(0.0202)
persistence	0.9895	0.9684
	(0.0189)	(0.0118)

an increase in its corresponding n_ath loading factor. Thus, the loading factors $\vartheta(n_a)$ for $n_a = 1, 2, 3$ are pinned down by the success rates after 1, 2, and 3 tries relative to the success rate in the first try.

Next, consider κ, $\underline{f}$, and ϱ. These parameters are pinned down by the variation in DI recipiency rates by frailty percentile group and age. For each frailty percentile group and age, the measure of DI recipients is a combination of new entrants onto DI and ongoing DI recipients. The measure of age j and frailty f_j individuals who are new entrants onto DI is a weighted average of DI application success rates, $\theta(\cdot)$, one period before at age $j-1$,

$$m^{DI}(j, f_j) = \sum_{\tilde{x}} \text{weights}(\tilde{x}) \theta(j-1, f_{j-1}, n_a),$$

where

$$\text{weights}(\tilde{x}) = \mu^N(\tilde{x}) I_D(\tilde{x}) p(x) \pi^f(f_j | j-1, f_{j-1}, s),$$

and $\tilde{x} \equiv (j-1, a, f_{j-1}, s, \epsilon, \bar{e}, n_a)$ summarizes the state vector at age $j-1$ including both $x \equiv (j-1, a, f_{j-1}, s, \epsilon, \bar{e})$ and n_a, $\mu^N(\tilde{x})$ is the measure of non-employed at age $j-1$ and $I_D(\tilde{x})$ is the indicator for decision to apply. Plugging in the expression for $\theta(j-1, f_{j-1}, n_a)$ (and again ignoring the cutoffs $\underline{f}$

Table 31: Estimation of zero frailty probits

age	-0.0009
	(0.0028)
age^2	-0.0003***
	(0.00003)
high school grad dummy	-0.103***
	(0.017)
college grad dummy	0.293***
	(0.017)
constant	-0.497***
	(0.0622)
Observations	95,655
Pseudo R^2	0.083

Note: Standard errors are in parenthesis. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Zero frailty probits are estimated using PSID data.

and 1) we get the following expression for the measure of new DI applicants of age j and frailty f_j:

$$m^{DI}(j, f_j) = \sum_{\tilde{x}} \text{weights}(\tilde{x}) \left(1 + \mathbf{1}_{\{j-1 \geq 55\}} (\varrho - 1)\right) \vartheta(n_a) f_{j-1}^{\kappa}, \tag{6}$$

where the weights are as defined above.

Equation (6) shows that κ has two direct effects on the fraction of new DI entrants. On the one hand, increasing κ lowers the success rate after DI application at all ages and frailty levels. This is because frailty is a number smaller than one and the benchmark value of κ is larger than one. On the other hand, increasing κ increases the dispersion in success rates with frailty. In other words, when κ increases the success rate at lower frailty levels falls more than the success rate at higher frailty levels. So we should observe that increasing κ lowers the DI recipiency rates in all age and frailty groups, but less so in higher frailty groups. The top row in Figure 6a shows the elasticity of the DI recipiency rates at each age and frailty group with respect to κ. The pattern is as expected in higher frailty groups and in all but the youngest age groups. This indicates that the overall dispersion in DI recipiency rates with frailty pins down κ.

The success rates of lower frailty groups and young age groups are not affected as much by κ because of the cutoff for DI eligibility $\bar{f}$. Figure 6a shows that $\bar{f}$ has little impact on the DI recipiency rates of younger age groups. In particular, the frailty values of most individuals under age 35 are below the benchmark value of $\bar{f}$ which explains why increasing $\bar{f}$ appears to have no effect on these groups. As individuals age and their health declines, the frailty group most sensitive to small variations in the cutoff becomes relatively healthier. Dispersion in the sensitivity across the frailty groups also increases. Thus, while κ is pinned down by the overall level of dispersion in DI recipiency rates with frailty, $\bar{f}$ is pinned down by the increasing dispersion in DI recipiency rates with frailty as individuals age. $\bar{f}$ also helps us match the low DI recipiency rates in the youngest age groups.

The age loading factor ϱ also has a direct effect on the fraction of new DI entrants but, as Equation (6) shows, only after age 55. The third row in Figure 6a shows the elasticity of the DI recipiency rates by age and frailty group with respect to ϱ. Consistent with Equation (6), increases in ϱ increase the DI recipiency rates after age 55. Thus, the age pattern of DI recipiency rate identifies ϱ, in particular, the sharp increase in the take up rates after age 55.

Two parameters remain to be discussed: the loading factor in the first try, $\vartheta(0)$, and the utility cost of application, ξ. They are pinned down by the overall fraction of 25- to 64-year-olds on DI and the fraction of 25- to 64-year-olds who apply for DI, respectively. Comparing the elasticities of these two moments to

Table 32: Estimation of nonzero frailty process

(a) First Stage: deterministic component

	HS Dropout	HS Graduates	Col Graduates
age	1.2733	1.1170	0.9697
	(0.0584)	(0.0314)	(0.0633)
age^2	2.1698	1.4475	2.0202
	(0.2023)	(0.0977)	(0.3360)
age^3	-0.6161	-1.7021	-0.6399
	(0.4148)	(0.2355)	(0.7418)
age^4	4.2929	8.8316	3.1876
	(0.3631)	(0.2488)	(0.6051)
const.	-2.4895	-2.5784	-2.8193
	(0.0055)	(0.0029)	(0.0040)

Note: age is rescaled so that age = (age-25)/100.

	HS Dropout	HS Graduates	Col Graduates
ρ	0.9855	0.9973	0.9680
	(0.0019)	(0.0009)	(0.0019)
σ_α^2	0.2238	0.1217	0.1231
	(0.0119)	(0.0050)	(0.0055)
σ_u^2	0.0427	0.0497	0.0313
	(0.0038)	(0.0020)	(0.0023)
σ_ε^2	0.0263	0.0197	0.0273
	(0.0018)	(0.0007)	(0.0012)

(b) Second stage: Stochastic component

these two parameters reported in Figures 6b and 6d, it is clear that the fraction on DI is more sensitive to $\vartheta(0)$ while the application rate is more sensitive to ξ. This is as expected since $\vartheta(0)$ has a direct effect on the overall fraction of individuals on DI and ξ is the only parameter that has a direct effect on the number of DI applicants.

Figure 7 reports elasticities of calibrated parameters with respect to moments. This inverse mapping illustrates how variation or misspecification in targeted moments would influence estimated parameter values, taking into account the joint dependencies across moments in the data. In this calculation the weighting matrix is equal to the identity matrix. All parameters values are fairly insensitive to misspecification in the targeted moments. For example, a one percentage change in any of the targets changes our calibrated parameters by at most 0.2 percentage points.

5 Additional Assessment

To assess the model's performance with regards to non-targeted moments, we report employment and DI recipiency rates by age and frailty percentile groups for each education group. The employment rates are presented in Figure 8 and the DI recipiency rates are presented in Figure 9.

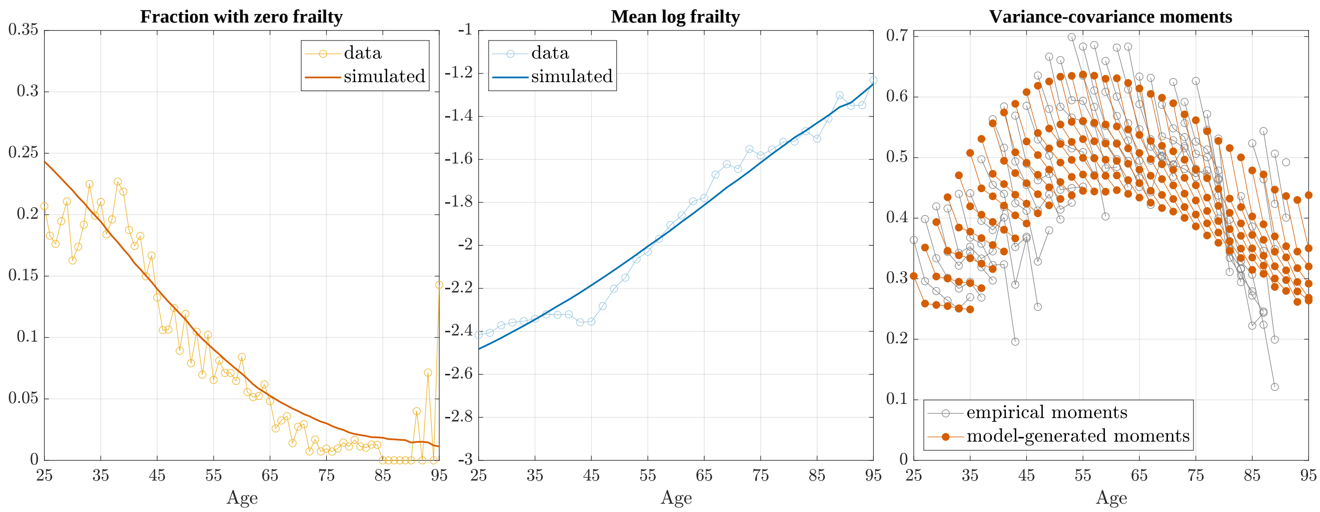


Figure 4: Estimation targets: auxiliary simulation model vs PSID data for high school dropouts. Left panel is the fraction with zero frailty by age, middle panel is mean log frailty by age for those with nonzero frailty, and right panel is the age-profile of the variance and covariances of log frailty residuals (the stochastic component of log frailty).

6 Additional Results

Figure 10 displays the ratios of lifetime earnings at the 5th, 10th, 90th, and 95th percentile relative to the median by age in the benchmark and the NFH economy. The figure shows that almost all of the difference between the variance of log lifetime earnings in the benchmark and the NFH economy is due to higher earnings at the bottom of the distribution in the NFH economy.

Health inequality in the model is due to both initial heterogeneity in frailty, captured by fixed effects and education, and idiosyncratic frailty shocks. To understand the relative importance of each of these components for lifetime earnings inequality, we conduct two related counterfactual experiments and report the results in Table 37. The fourth row of the table shows the variance of log lifetime earnings for a counterfactual economy that is identical to the benchmark, except that there are no frailty shocks. All inequality in frailty in this economy is due to initial fixed heterogeneity. The sixth row shows the same results for an economy that does not feature any initial heterogeneity in frailty due to fixed effects or education. All the inequality in frailty in this counterfactual is driven by frailty shocks.

Removing either only frailty shocks or only fixed heterogeneity reduces wage inequality and increases labor force participation leading to less inequality in lifetime earnings. However, removing frailty shocks leads to a larger reduction in lifetime earnings inequality than removing fixed heterogeneity. The main reason for this is that labor supply is higher absent shocks than absent fixed heterogeneity. This can be seen in Figure 11 which shows the fractions employed by age in the benchmark economy, the no-frailty-heterogeneity economy, the economy with no frailty shocks, and the economy with no frailty fixed effects. Absent shocks most individuals are too healthy to qualify for DI and the average employment rate is above that in the baseline at all ages. Absent fixed heterogeneity, all newborn individuals are healthy and the employment rate initially rises above the baseline level leading to a reduction in lifetime earnings inequality. However, due to the highly persistent frailty shocks, the fraction of individuals whose health is poor enough that they choose to exit the labor force and apply for DI increases with age. By age 35, employment rates begin to decline eventually reaching similar levels to those in the benchmark economy.

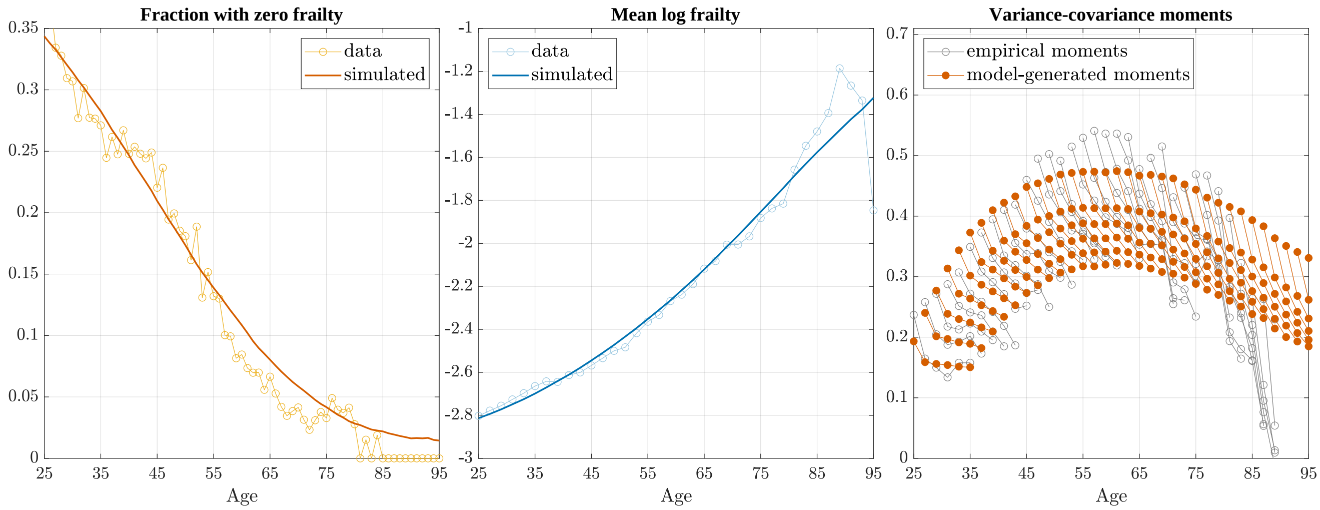


Figure 5: Estimation targets: auxiliary simulation model vs PSID data for college graduates. Left panel is the fraction with zero frailty by age, middle panel is mean log frailty by age for those with nonzero frailty, and right panel is the age-profile of the variance and covariances of log frailty residuals (the stochastic component of log frailty).

6.1 Aggregate effects of health inequality

We report the aggregate implications of removing health inequality and removing its effect via each of the five channels through which health operates in the model in Table 38. Each column shows the change in GDP per capita, aggregate consumption, aggregate capital, aggregate labor services, aggregate hours, and labor productivity (GDP per hour) for each of our counterfactual economies relative to the benchmark. The first column shows that removing all inequality in frailty raises GDP per capita by 3.91 percent and aggregate consumption by 2.39 percent. It also increases hours worked (employment) by 6.01 percent. As we explain in Section 6 of the paper, removing inequality in frailty mainly increases participation of workers at the bottom of the income/wage distribution. Since these are on average the less productive workers, the resulting GDP per hour (per employed worker) falls by 1.98 percent.

Columns 2 through 6 show that the main drivers of the aggregate effects are the DI and disutility channels. Notice, also, that the effect of DI on consumption is significantly less than the effect on GDP. This is due to the fact that removing only the DI channel has opposite effects on labor supply at young and old ages. Moreover, while it reduces the aggregate disability benefit, it increases the fraction of individuals who are eligible for means-tested welfare transfers. These opposing effects aggregate to a smaller relative impact on consumption of the DI channel (as compared to the relative impact of the disutility channel), even-though it is a significantly more important channel in terms of affecting individual's labor supply and income inequality. Finally, removing frailty inequality in mortality increases survival and tilts the age distribution of the model towards older (mostly retired) individuals. For this reason it has a negative impact on all aggregate measures.

Table 39 shows the DI recipiency rates (top panel), employment rates (middle panel), and the share of people who receive means-tested transfers (bottom panel) in the benchmark economy, NFH economy and the five additional counterfactual economies. Looking at the first two columns, the effect of removing frailty inequality on DI recipiency is large for all three educations groups. Although, college graduates have very low DI usage in the benchmark so the increase in employment is concentrated among high school dropouts and to a lesser extent high school graduates. Notice that the effect on the fraction receiving means-tested transfers is small. The effect of removing frailty inequality on means-tested program usage is due to a balance of two opposing forces. On the one hand, shutting down the DI channel pushes young

Table 33: Estimation of out-of-pocket medical expenditures

	Estimation of log of out of pocket medical expenditures								
	High School Dropouts			High School Graduates			College Graduates		
	on medicare	working	not working	on medicare	working	not working	on medicare	working	not working
age	0.19	-0.23	0.42	-0.08	-0.03	-0.05	0.47	-0.11	-0.75
	(0.10)	(0.09)	(0.22)	(0.07)	(0.06)	(0.16)	(0.16)	(0.08)	(0.30)
age^2	-0.0024	0.00577	-0.00948	0.00238	0.00166	0.00165	-0.00717	0.00339	0.0163
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
age^3	0.0000114	-0.0000391	0.0000749	-0.0000149	-0.0000115	-0.00000646	0.0000359	-0.0000235	-0.000104
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
frailty	21.1	29.15	26.66	15.12	25.01	23.07	12.32	19.89	21.68
	(0.87)	(0.84)	(1.40)	(0.71)	(0.52)	(1.24)	(0.90)	(0.72)	(2.28)
$frailty^2$	-49.27	-71.78	-62.71	-35.9	-66.46	-55.72	-32.42	-49.45	-62.04
	(2.77)	(4.76)	(4.87)	(2.41)	(3.05)	(4.83)	(3.34)	(4.63)	(9.30)
$frailty^3$	35.86	55.31	47.43	27.32	54.53	42.95	26.16	37.55	51.05
	(2.45)	(6.36)	(4.56)	(2.27)	(4.28)	(5.03)	(3.33)	(6.77)	(9.67)
constant	-2.138	3.876	-5.45	3.535	1.525	1.232	-5.263	3.977	12.97
	(1.90)	(1.27)	(3.01)	(1.30)	(0.78)	(2.28)	(3.15)	(1.04)	(4.20)
Obs.	7160	17232	3068	10941	46641	5138	5182	24998	1456
R^2	0.162	0.205	0.312	0.113	0.181	0.278	0.0694	0.142	0.259

Table 34: Effect of work limitation on wages (% decline in wages relative to no limitation)

	mean	Low and Pistaferri (2015)	Our estimation				
	frailty	non-college	non-college	HSD	HSG	college	all
No limitation	0.07						
Mod. limitation	0.22	-5.7	-12.9	-18.9	-11.4	-0.1	-8.5
Severe limitation	0.32	-17.7	-22.5	-33.1	-20.0	-11.6	-18.6

frail workers out of the labor force and onto these programs. On the other hand, removing the disutility and mortality channels creates additional incentives to work at older ages reducing the usage of these programs.

6.2 Alternative measures of inequality

Our findings show that health inequality increases lifetime earnings inequality. The increase is driven by the negative impacts of poor health on the labor supply and earnings of individuals in the bottom of the lifetime earnings distribution. This increase is offset by SSDI/SSI and means-tested transfers. However, the offset is only partial. As a result, health inequality also has an impact on inequality in disposable income and consumption.

Table 40 shows the variance of log lifetime disposable income in the benchmark economy, the economy with no frailty heterogeneity, the economy with only individual fixed effect heterogeneity (and no shock) and the economy with only frailty shocks (and no fixed effect heterogeneity), and Figure 15a presents the variance of log lifetime disposable income by age in the benchmark and NFH economy.[3] Removing health inequality causes the variance of log lifetime disposable income to fall by 7.7, 7.8, and 6.8 percent at ages 45, 55, and 65, respectively. This finding implies that about two-thirds of the inequality in lifetime earnings due to poor health is undone through the tax and transfer system.

Interestingly, the relatively smaller impacts of health inequality on lifetime disposable income inequality do not translate into smaller impacts of health inequality on consumption inequality. Table 41 show the variance of log current consumption in the benchmark economy, the economy with no frailty

[3]Disposable income as the sum of labor earnings and transfers net of all taxes. Lifetime disposable income at each age is the sum of disposable income to date. We have done all the calculations with an alternative definition that includes capital income. The results are very similar.

Table 35: Parameters chosen outside the model

Parameter	Description	Value
Demographics		
J	maximum age	70 (94 yo)
R	retirement/SS eligibility age	41 (65 yo)
ν	population growth rate	0.02
Preferences		
γ	curvature of utility function	2
μ	weight on consumption	0.5
Employment dynamics		
$\sigma(s)$	job separation rates (HSD, HS, CL)	0.27, 0.15, 0.06
φ	job finding rate after non-employment	0.52
χ	% earnings lost due to unemployment	34.6
Technology		
α	capital share	0.33
δ	depreciation rate	0.07
r	return on assets	0.04
Government policies		
τ_{SS}, τ_{med}	SS and Medicare tax rates	0.124, 0.029
τ_K	capital tax rate	0.3
τ	tax progressivity parameter	0.036
$\underline{b}$	SSI payment (% of ave. earnings)	13
$\underline{c}$	minimum consumption (2000 $'s)	$4,375
G	government purchases (% of GDP)	12.8

heterogeneity, the economy with only individual fixed effect heterogeneity (and no shock) and the economy with only frailty shocks (and no fixed effect heterogeneity), and Figure 15a presents the variance of log current consumption by age in the benchmark and NFH economy. Removing health inequality significantly reduces the variance of log current consumption. The variance of log consumption at age 65 is 18.2 percent lower in the NFH economy as compared to the benchmark.

Why are the impacts on consumption inequality so much larger than the impacts on lifetime disposable income inequality? The primary reason is that removing health inequality reduces wealth inequality. Less educated individuals in poor health in the benchmark have lower savings than less educated individuals in the NFH economy for three reasons. First, they have less income. Second, they have less incentives to accumulate wealth due to a lower life expectancy. Third, poor health (or the risk of poor health) means they are more impacted by the negative effects of the means-tested transfer program on incentives to save.[4]

Relative to the case of the variance of log lifetime earnings, initial fixed effect heterogeneity plays a larger role in accounting for the decline in the variance of log lifetime disposable income. This is for two reasons. First, individuals with a high initial fixed frailty component may never work. These individuals have positive disposable lifetime income but because they have zero lifetime earnings are not accounted for in the variance of log lifetime earnings calculations. Second, DI provides more insurance against frailty shocks than high initial and permanent frailty because benefits are based on past earnings.

[4]It is well documented that means-tested transfer programs distort savings incentives and that their distortionary effects are larger for lower income individuals. See, for instance, Hubbard et al. (1995).

Table 36: Parameters calibrated using the model and associated targeted moments

Parameter	Description	Value
Preferences and taxes		
β	discount factor	0.986
ξ	utility cost of first DI application	0.07
λ	tax scale effect	0.92
Disutility of work $= \phi_0 + \phi_1 f^{\phi_2}$		
ϕ_0	baseline	0.62
ϕ_1	scale effect	6.2
ϕ_2	curvature effect	6.0
Prob. successful DI app. $= \min\left\{1, \left(1 + \mathbf{1}_{\{j \geq 55\}} (\varrho - 1)\right) \vartheta (n_a) f^{\kappa}\right\}$ if $f > \underline{f}$ and 0 otherwise		
$\vartheta(n_a = 0)$	scale effect at initial try	6.84
$\vartheta(n_a = 1)$	scale effect after 1 try	3.98
$\vartheta(n_a = 2)$	scale effect after 2 tries	3.86
$\vartheta(n_a \geq 3)$	scale effect after 3+ tries	3.93
κ	curvature effect	3
ϱ	loading factor after age 55	1.8
$\underline{f}$	threshold for nonzero probability	0.196

Targeted moment $\rightarrow$ associated parameter(s)[a]	Data	Model
Wealth-earnings ratio $\rightarrow \beta$	3.2	3.2
Federal income tax (% of GDP) $\rightarrow \lambda$	8	8
DI applicants (% of 25–64 yo population) $\rightarrow \underline{f}$	2	2
Employment rates by frailty percentile group and age $\rightarrow (\phi_0, \phi_1, \phi_2)$	Figure 3	
DI recipiency rates by frailty percentile group and age $\rightarrow (\kappa, \varrho, \xi)$	Figure 3	
DI application success rates after n_a tries $\rightarrow \vartheta(n_a)$	Figure 3	

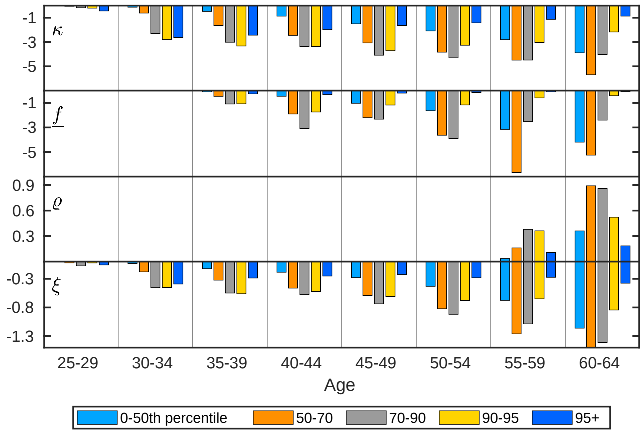
[a] All of these parameters are jointly calibrated. Here, we only point to the parameter(s) that are more closely associated with the respective moments.

6.3 Alternative measures of health

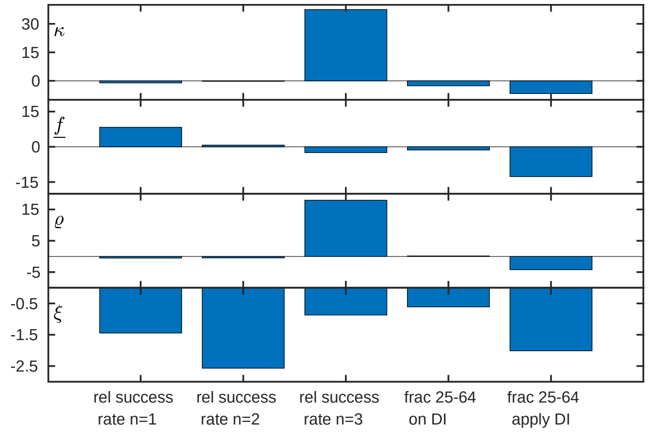
Tables 42–47 report education and age-specific transition rates across good and bad health states for men ages 25 to 50 in the PSID, and across good health, bad health, and death for men over age 50 in the HRS. Transition rates are reported using two ways of converting SRHS to a binary measure: classifying 'fair' and 'poor' SRHS as bad health and the rest as good health, and classifying only 'poor' SRHS as bad health. Table 48 reports the initial distributions of men ages 25 and 26 across these health states in the PSID.

6.4 Robustness of welfare results to treatment of men on DI at age 25

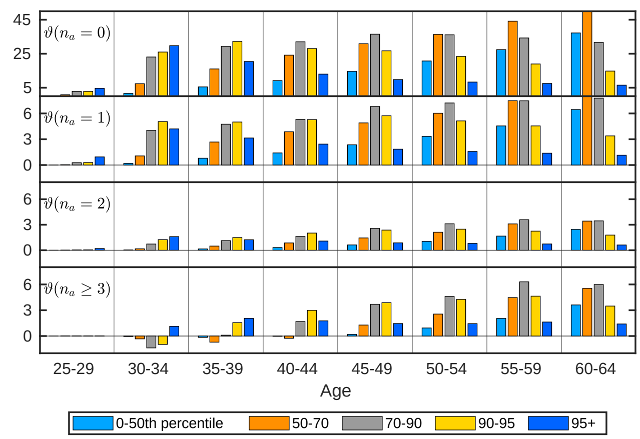
In our benchmark economy, 2 percent of men start off already on DI at age 25 to be consistent with the fact that 2 percent of men aged 24–26 years are on DI in the data. In the 'No DI' economy in Section 6.3 these individuals are assumed to have zero productivity and start out non-employed. Consequently, they qualify immediately for the means-tested transfer program which they stay on their entire lives. While men who are already on DI at age 25 likely have very low productivity, the assumption that their productivity is zero may be too extreme. An alternative assumption is that the fixed effect in their productivity process is similar to those of men who transition to DI at ages 26–30 in the benchmark economy and the initial value of their persistent shock at age 25 is similar to persistent shock value of



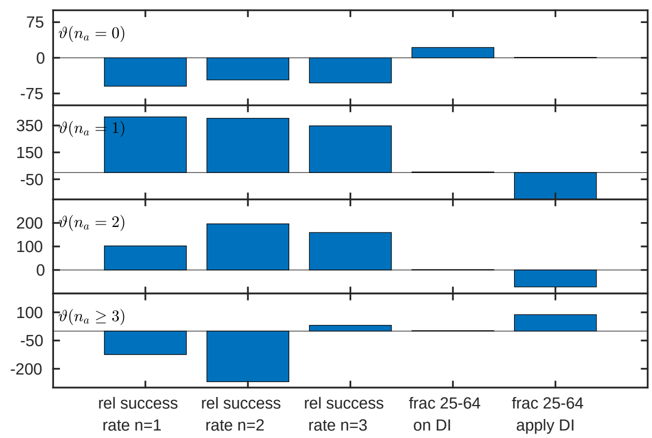
(a) DI recipiency rates with respect to κ, $\underline{f}$, ϱ, and ξ



(b) Relative DI application success rates, fraction on DI and fraction apply for DI with respect to κ, $\underline{f}$, ϱ, and ξ



(c) DI recipiency rates with respect to $\vartheta(n_a = 0)$, $\vartheta(n_a = 1)$, $\vartheta(n_a = 2)$, $\vartheta(n_a \geq 3)$

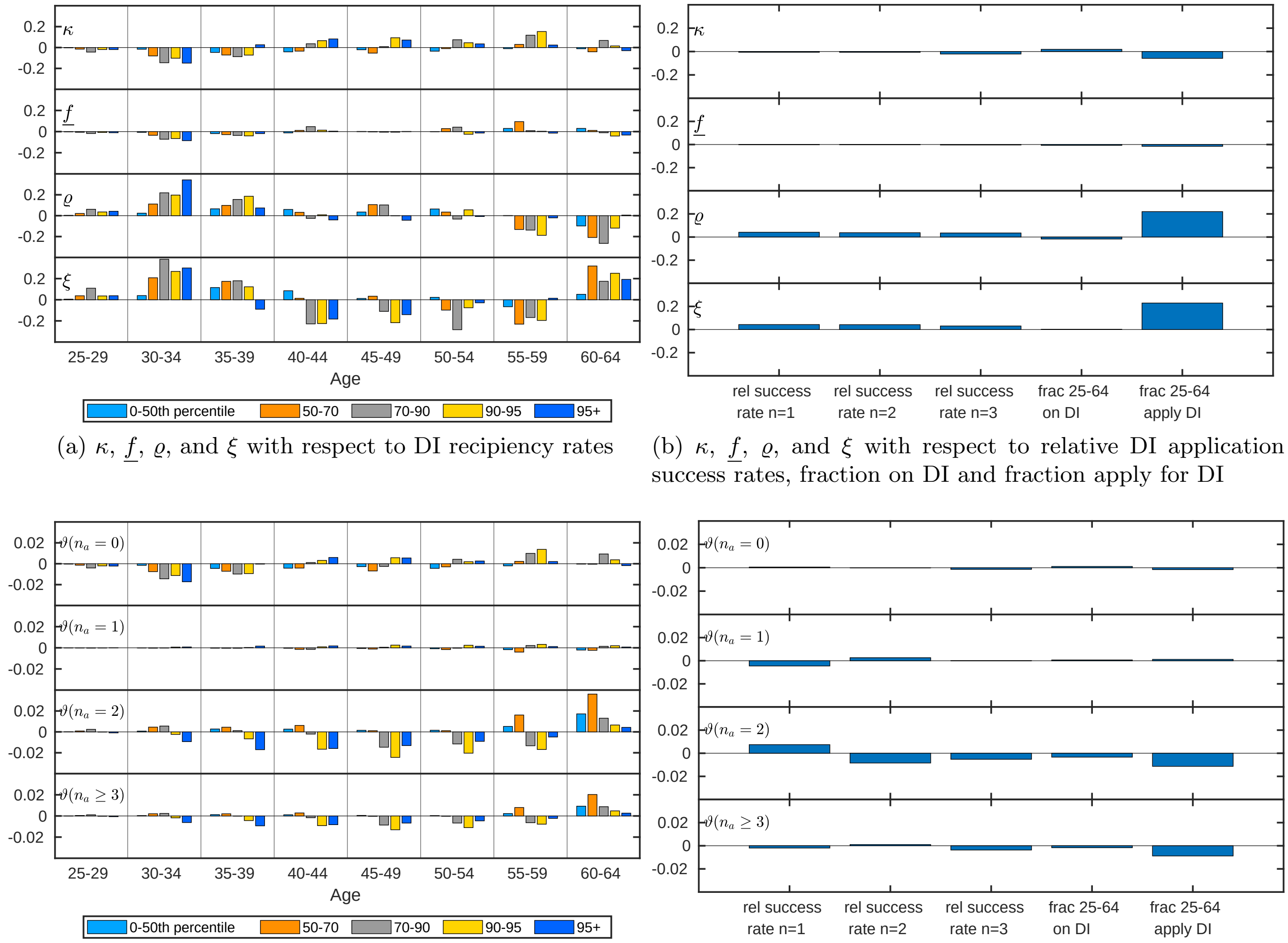


(d) Relative DI application success rates, fraction on DI and fraction apply for DI with respect to $\vartheta(n_a = 0)$, $\vartheta(n_a = 1)$, $\vartheta(n_a = 2)$, $\vartheta(n_a \geq 3)$

Figure 6: Jacobian: derivative of moments w.r.t. parameters.

these men in the period that they apply for DI. We now investigate the implications for the main welfare results from removing DI under this alternative assumption.

Under the alternative assumption, absent DI, we give individuals who were on DI at age 25 in the benchmark the same labor productivity process as everyone else. However, we set their productivity fixed effect component to the average fixed effect of men who transition to DI between ages 26 and 30 in the benchmark economy, and we set their initial persistent shock value to the average persistent shock value of these men in the period that they apply for DI. Table 49 reports the welfare implications of removing DI under this alternative assumption. The welfare results in Table 49 can be compared to those in Table 10 in the paper. Under the assumption that those initially on DI have positive productivity, the welfare losses of the high-school dropouts and graduates from removing DI are smaller and the welfare gains for college graduates are larger. The welfare results change the most of the high-school dropouts because they account for most of the individuals initially on DI in the benchmark. However, even for the high-school dropouts, the changes are small and, before they know their education type, individuals in the model still prefer the economy with the DI program once they take into account the tax implications of removing it.



(a) κ, $\underline{f}$, ϱ, and ξ with respect to DI recipiency rates

(b) κ, $\underline{f}$, ϱ, and ξ with respect to relative DI application success rates, fraction on DI and fraction apply for DI

(c) $\vartheta(n_a = 0)$, $\vartheta(n_a = 1)$, $\vartheta(n_a = 2)$, $\vartheta(n_a \geq 3)$ with respect to DI recipiency rates

(d) $\vartheta(n_a = 0)$, $\vartheta(n_a = 1)$, $\vartheta(n_a = 2)$, $\vartheta(n_a \geq 3)$ with respect to relative DI application success rates, fraction on DI and fraction apply for DI

Figure 7: Lambda: derivative of parameters w.r.t. moments.

6.5 Additional Welfare Results

More detailed results from the experiments where we replace the DI program with one that has no redistribution across education groups are reported in Table 50.

7 Measurement of lifetime earnings in NLSY79

In order to assess the model's performance in matching the inequality of lifetime earnings we use NLSY79 data. This comparison cannot be done using PSID because of the relatively short panel. On the other hand, NLSY79 is a sample of representative young individuals who were between the ages of 14 and 22 in 1979 (when they were first interviewed). We use data from 1982 to 2018 which allows us to calculate lifetime earnings over the ages 25 to 60 for 2,901 male respondents.

Our sample consists of men between ages 25 and 60 in the NLSY79 "Cross Section Sample" with complete information on hours and earnings over their lifecycle. We deflate the earnings data using PCE. In our structural model we ignore the intensive margin of labor supply. This is motivated by our empirical analysis that finds no impact from health on hours worked conditional on employment. In order to construct a comparable moment in the data, we calculate hourly wages conditional on employment.

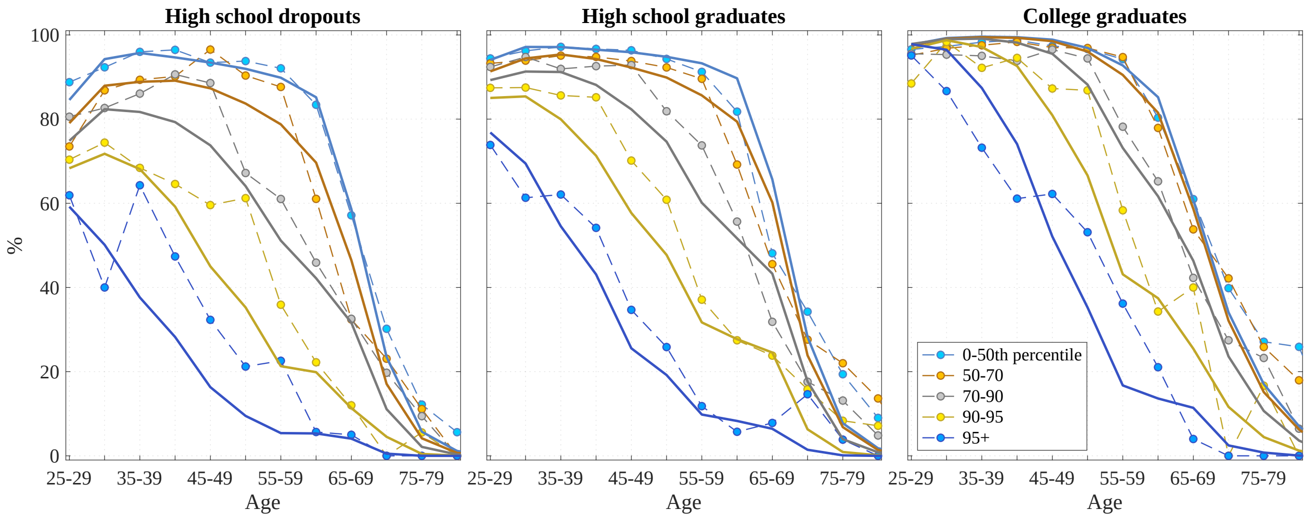


Figure 8: Calibration assessment: model versus data. Solid (dashed) lines are the model (data). The panels show the male employment rates by frailty percentile groups and age for each education group: high school dropouts, high school graduates, and college graduates. Source for the data PSID.

Table 37: Variance of log lifetime earnings: frailty shocks versus initial frailty heterogeneity.

	age 35	age 45	age 55	age 65	age 75
Benchmark	0.435	0.502	0.582	0.586	0.520
No frailty heterogeneity (NFH)	0.388	0.395	0.421	0.434	0.445
% Δ relative to benchmark	-11.0	-21.5	-27.7	-26.0	-14.5
No frailty shocks	0.377	0.406	0.446	0.462	0.441
% Δ relative to benchmark	-13.3	-19.3	-23.4	-21.2	-15.1
No frailty fixed effect	0.391	0.430	0.490	0.505	0.496
% Δ relative to benchmark	-10.1	-14.5	-15.8	-13.8	-4.7

Note: In the "No frailty heterogeneity" counterfactual all individuals have the average frailty age profile. "No frailty shocks" removes ex post shocks but retains all fixed heterogeneity. "No frailty fixed effect" removes fixed heterogeneity but retains all shocks.

We use this variable to calculate lifetime earnings.

We use the same definition of employment as we use in the PSID sample. In any given year, individuals are considered employed if they worked at least 520 hours during the year and non-employed if they worked less than 520 hours. Observations where labor force status is deemed employed based on hours but wages are reported as less than $4 an hour are dropped. When calculating lifetime earnings, we assign earnings of zero to those individuals who are classified as non-employed.

To check whether our NLSY79 sample is comparable to our PSID sample (which is the main sample we use to estimate the wage process we feed into the model), we compare the age profiles of the variance of log current earnings across the two samples. Figure 16 shows that the age profiles are very similar. It also shows that the variance of log current earnings in the model matches the data moments closely. Finally, Figure 5 in the paper shows the variance of log lifetime earnings in the NLSY79 sample and in the model, as well as, the fraction of men with zero lifetime earnings. The model matches this non-targeted moment very closely.

The left panel of Figure 17 shows the ratios of the 5th and 95th percentiles of lifetime earnings to the median in our NLSY sample and the benchmark economy. The model shows slightly less inequality

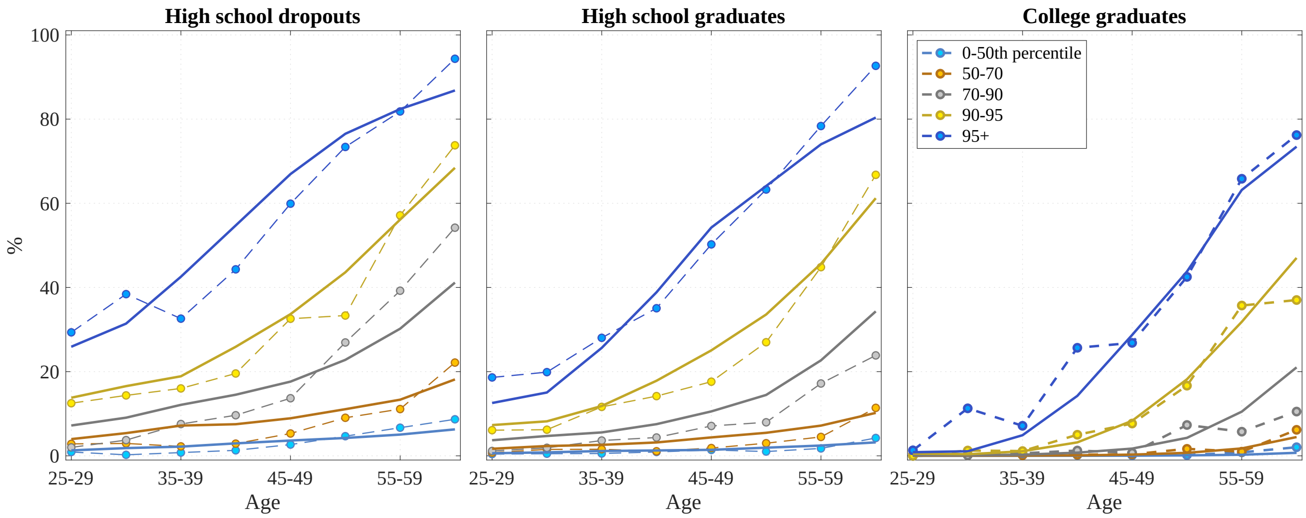
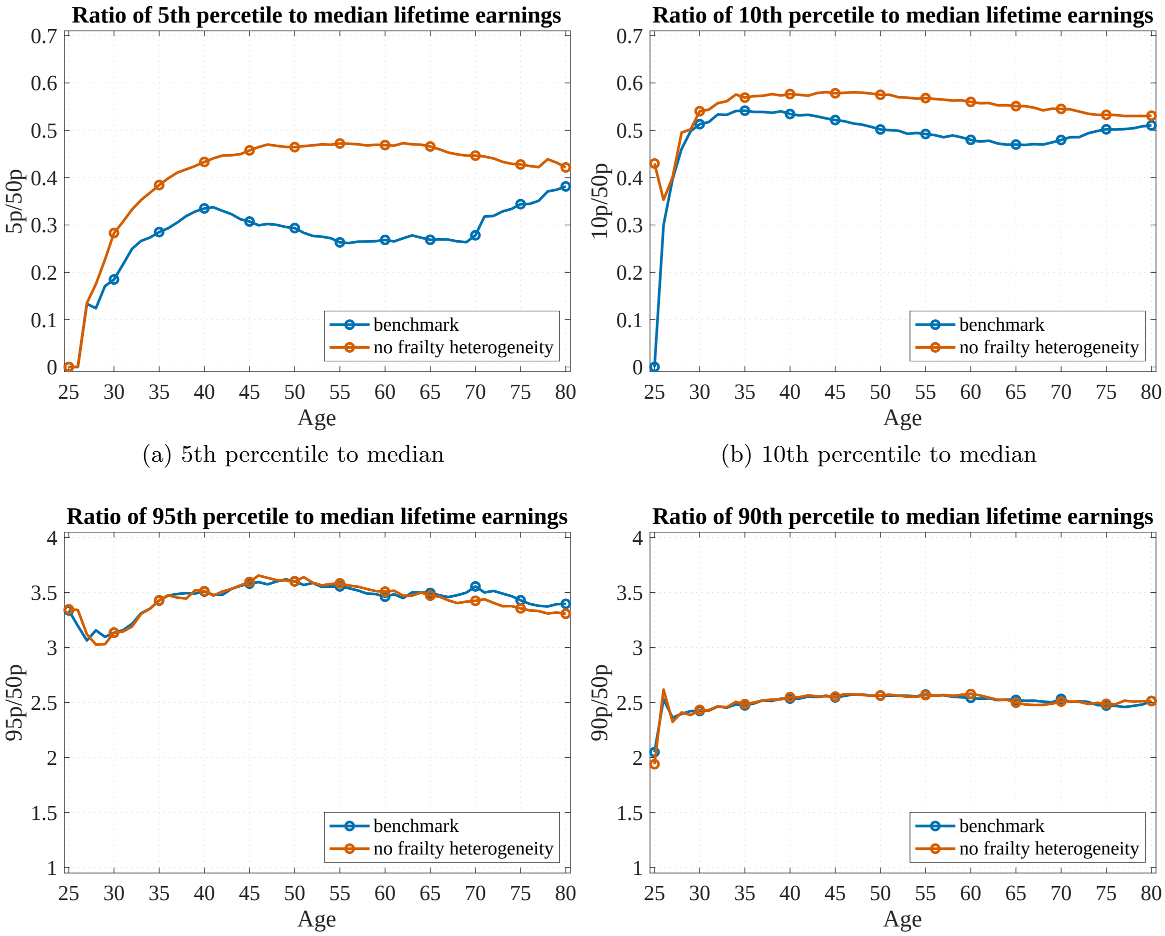


Figure 9: Calibration assessment: model versus data. Solid (dashed) lines are the model (data). The panels show the male DI recipiency rates by frailty percentile groups and age for each education group: high school dropouts, high school graduates, and college graduates. Source for the data is MEPS and Social Security Administration.

at the bottom and slightly more inequality at the top. In other words, the poorest individuals in the model are not as poor as they are in the NLSY sample, relative to the median. On the other hand richest individuals are richer relative to the median when compared to our NLSY sample. However, the gap between the model and data is small, especially at the bottom of the distribution. The right panel shows the same pattern for ratio of the 10th percentile and the 90th percentiles of lifetime earnings relative to the median.



(a) 5th percentile to median

(b) 10th percentile to median

(c) 95th percentile to median

(d) 90th percentile to median

Figure 10: Inequality in lifetime earnings in the benchmark economy (blue) and the no-frailty-heterogeneity economy (red).

8 Persistence of health: SRHS vs Frailty

Here we compare the persistence of health when it is measured by self-reported health status (SRHS) and when it is measured by frailty. The following is taken from the supplemental appendix to Hosseini et al. (2022). We include this here for convenience. We invite the reader to see Hosseini et al. (2022) and it's supplemental appendix for more details.

SRHS partitions individuals into different categories according to their own subjective health assessment. One way to track the evolution of SRHS is to document how the size of each partition changes with age. Using PSID data we plot the fraction of individuals in each self-reported health category by age in Figure 18. The blue lines in Figures 18a, 18b, 18c, 18d, and 18e show the fraction individuals who report SRHS of 'excellent', 'very good', 'good', 'fair', and 'poor', respectively. The first observation here is that the fraction of those who report 'excellent' and 'very good' health decline by age, while the fraction of those who report 'fair' and 'poor' health increase by age.

To compare the evolution of frailty (a continuous variable) with categorical SRHS we likewise partition individuals into categories using frailty. Specifically, we partition individuals within each five-year age group into five frailty categories labeled 'excellent', 'very good', 'good', 'fair', and 'poor' (same as the SRHS categories). The cutoff values of frailty that determine which category is assigned are age-independent and determined such that the distribution of individuals across frailty and SRHS categories

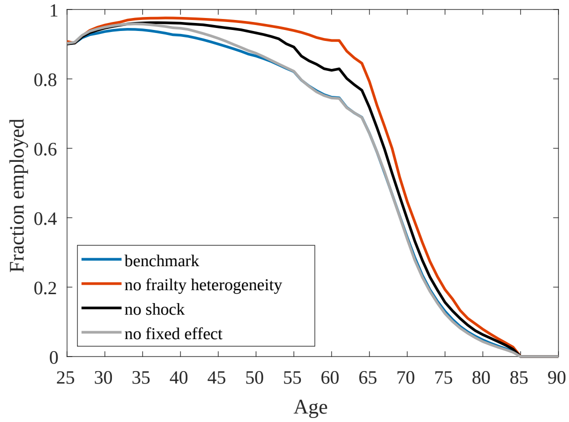


Figure 11: Fraction employed by age in the benchmark economy, the no-frailty-heterogeneity economy, the economy with no frailty shocks, and the economy with no frailty fixed effects.

Table 38: Aggregate Effect of Healthy Inequality

	NFH in model	NFH in DI	NFH in Disutility	NFH in Labor prod.	NFH in Med. Exp.	NFH in Mortality
	% change relative to benchmark					
GDP	3.91	2.81	1.05	0.74	0.38	−1.29
Consumption	2.39	1.25	0.84	0.33	0.35	−1.43
Capital	3.91	2.81	1.05	0.74	0.38	−1.29
Labor input	3.91	2.81	1.05	0.74	0.38	−1.29
Hours	6.01	3.38	1.15	1.16	0.46	−0.94
GDP per Hour	−1.98	−0.54	−0.10	−0.42	−0.08	−0.35

Note: Each column shows the difference in aggregate measure between the respective counterfactual and benchmark. NFH: no frailty heterogeneity. NFH in DI: probability of DI acceptance is the same for all individuals at same age, NFH in Labor Prod.: there is no heterogeneous effect of frailty on wage, NFH in Disutility: there is no heterogeneous effect of frailty on disutility from work, NFH in Med. Exp.: no heterogeneous effect of frailty on out of pocket medical expenditures, NFH in Mortality: no heterogeneous effect of frailty on mortality.

is the same for the age group 25 to 29. For example, the fraction of 25- to 29-year-olds with SRHS of 'excellent' is 26.5 percent. We set the cutoff value for 'excellent' frailty such that 26.5 percent of 25- to 29-year-olds are also in the 'excellent' frailty category. The resulting cutoff value for frailty is 0.036. At each age, individuals with a frailty value less than 0.036 are assigned to the frailty category 'excellent'. Next, we find the cut off value for the 65th percentile (65 percent of the sample have a SRHS of 'excellent' or 'very good' at age 25 to 29): 0.071. In each age group, anyone whose frailty measure is larger than 0.036 but smaller than 0.071 is assigned to the frailty category of 'very good', and anyone whose frailty value is exactly 0.071 is randomly assigned to either the 'very good' or 'good' category. The other two cutoffs, are chosen accordingly at the 91st and 99th percentiles and determine the assignment of the remaining individuals to the 'good', 'fair', and 'poor' frailty categories.[5] Using this procedure, the frailty and SRHS categories of the age group 25 to 29 are perfectly aligned (by construction).[6]

The red lines in Figures 18a, 18b, 18c, 18d, and 18e show the fraction individuals who fall in the frailty categories of 'excellent', 'very good', 'good', 'fair', and 'poor', respectively. Notice that fractions of 'excellent' and 'very good' frailty categories decline at a faster rate (relative to SRHS categories).

[5]See Hosseini et al. (2022) for all cutoff values.

[6]There is nothing special about age 25-to-29. We show, in the supplemental appendix of Hosseini et al. (2022), that our findings do not change if we choose another age group to define cutoffs.

Table 39: Effects of removing health inequality (overall and via different channels) on DI recipiency, employment, and receipt of means-tested transfers

	Benchmark	NFH in model	NFH in DI	NFH in Disutility	NFH in Labor prod.	NFH in Med. Exp.	NFH in Mortality
			DI Recipiency Rate (% of 25- to 64-year-olds)				
ALL	6.04	0.92	2.01	5.59	5.50	6.11	6.77
HSD	13.73	1.69	5.95	12.87	12.60	13.94	15.28
HSG	7.37	1.36	2.47	6.85	6.71	7.46	8.18
CG	1.60	0.00	0.05	1.38	1.41	1.59	1.83
			Employment Rate (% of 25- to 74-year-olds)				
ALL	83.70	90.04	86.53	84.58	84.67	84.00	82.81
HSD	72.96	85.77	76.87	74.25	75.04	73.18	71.30
HSG	82.22	89.46	85.72	83.25	83.27	82.43	81.20
CG	89.27	92.35	90.77	89.79	89.75	89.71	88.97
			Means-tested Transfers Recipiency Rate (%)				
ALL	3.41	3.09	4.45	3.31	3.10	3.16	3.47
HSD	7.41	6.05	10.34	7.27	6.46	6.78	7.45
HSG	4.26	3.85	5.54	4.12	3.89	3.91	4.29
CG	0.97	0.94	1.08	0.94	0.93	0.98	1.01

Note: The top (middle) [bottom] panel shows DI recipiency (employment) [means-tested transfers recipiency] rates in the benchmark and each counterfactual economy. HSD: high school dropout, HSG: high school graduate, CG: college graduate. NFH: no frailty heterogeneity. NFH in DI: probability of DI acceptance is the same for all individuals at same age and determined by the average frailty profile, NFH in Labor Prod.: effect of frailty on labor productivity is determined by the average frailty profile, NFH in Disutility: disutility from work is determined by average frailty profile, NFH in Med. Exp.: out-of-pocket medical expenditures are determined by the average frailty profile, NFH in Mortality: mortality is determined by the average frailty profile.

Moreover, frailty categories of 'fair' and 'poor' grow faster (relative to SRHS categories). The takeaway here is that health deteriorates faster when measured using frailty in comparison to SRHS.

Next, we look at the transition probabilities across different health categories. This allows us to compare the persistence of SRHS with that of frailty. To this end, we report the transition probabilities across five categories of SRHS (Table 51) and frailty (Table 52) for three different 25-year age groups. The frailty categories are defined by cutoffs that equate the share in each SRHS and frailty category at ages 25 to 29, as described above.

The table shows that the frailty index is more persistent than SRHS. Notice that, for all age groups, the diagonal values are all higher for the frailty index relative to SRHS. For example, individuals between the ages of 50 and 75, with frailty category 'excellent' have a 78.6% chance of maintaining this status while individuals in the same age group, and with 'excellent' SRHS have only a 57.1% chance. The difference in persistence is largest at the poor health end of the spectrum. Once an individual's frailty index is high enough that s/he is assigned to the 'poor' frailty category the probability s/he is there two years later is 73.2%, 83.4%, and 86.86% respectively for age groups 25–49, 50–74, and 75+. In contrast, individuals who report a SRHS status of 'poor' have only 41%, 59.3%, and 59.7% chance, respectively, of reporting poor health two years later.

The take away from these tables is that the frailty categories are significantly more persistent than the SRHS categories. In other words, it is much more likely for individuals to change their assessment of their own health than it is for their frailty index to cross the fixed thresholds that determine comparable frailty categories.

Table 40: Variance of log lifetime disposable income: frailty shocks versus initial frailty heterogeneity

	age 35	age 45	age 55	age 65	age 75
Benchmark	0.328	0.364	0.390	0.390	0.365
No frailty heterogeneity (NFH)	0.307	0.336	0.360	0.363	0.359
% Δ relative to benchmark	-6.3	-7.7	-7.8	-6.8	-1.8
No frailty shocks	0.318	0.356	0.381	0.382	0.365
% Δ relative to benchmark	-3.0	-2.3	-2.2	-2.1	-0.1
No frailty fixed effect	0.315	0.351	0.374	0.373	0.369
% Δ relative to benchmark	-3.8	-3.6	-4.1	-4.2	1.1

Note: In the "No frailty heterogeneity" counterfactual all individuals have the average frailty age profile. "No frailty shocks" removes ex post shocks but retains all fixed heterogeneity. "No frailty fixed effect" removes fixed heterogeneity but retains all shocks.

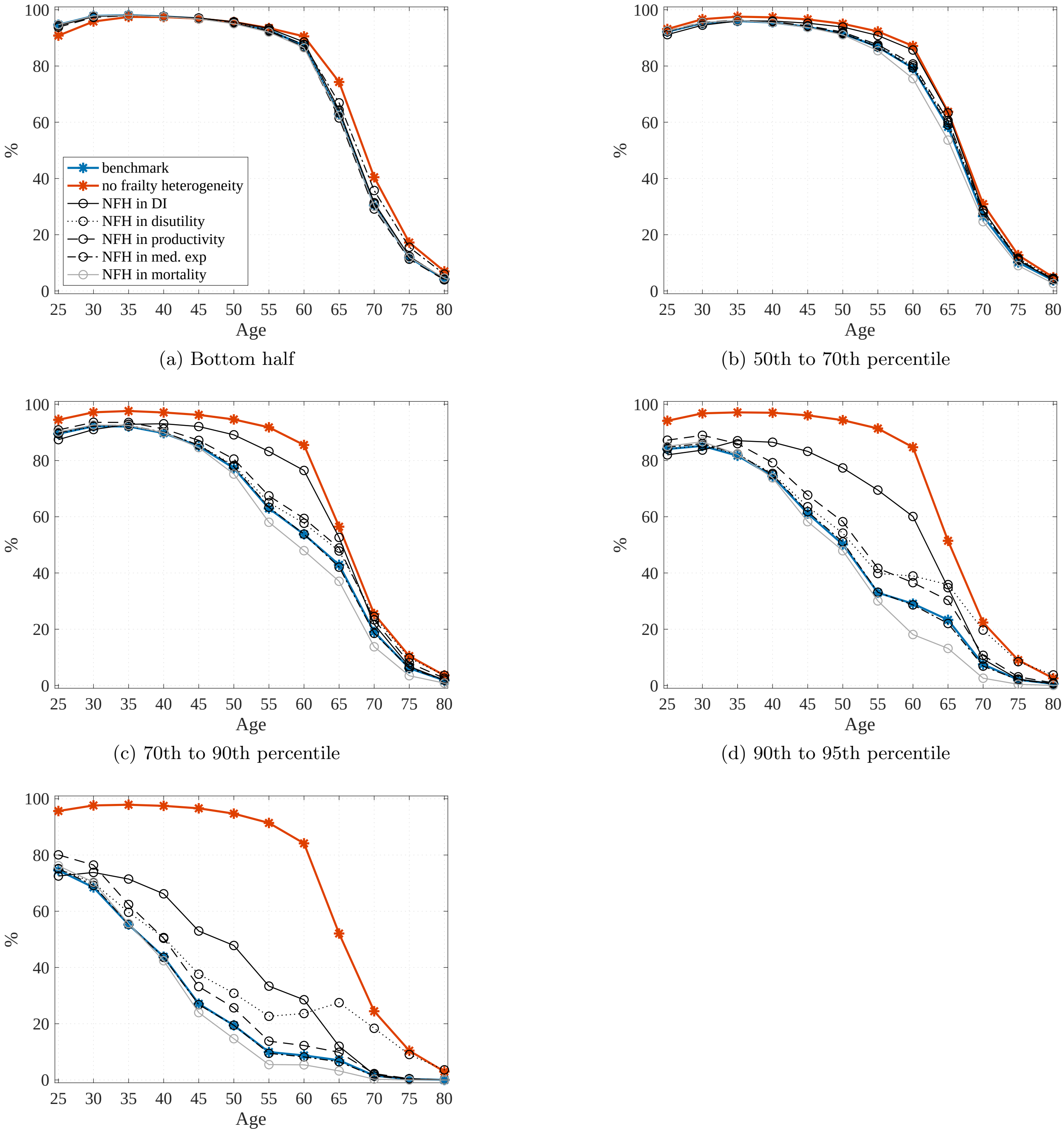
Table 41: Variance of log consumption.

	age 35	age 45	age 55	age 65	age 75
Benchmark	0.372	0.450	0.566	0.600	0.580
No frailty heterogeneity (NFH)	0.329	0.383	0.478	0.491	0.523
% Δ relative to benchmark	-11.7	-15.0	-15.5	-18.2	-9.8
No frailty shocks	0.347	0.417	0.522	0.549	0.562
% Δ relative to benchmark	-6.7	-7.5	-7.7	-8.5	-3.0
No frailty fixed effect	0.347	0.431	0.539	0.576	0.586
% Δ relative to benchmark	-6.7	-4.3	-4.7	-4.1	1.1

Note: In the "No frailty heterogeneity" counterfactual all individuals have the average frailty age profile. "No frailty shocks" removes ex post shocks but retains all fixed heterogeneity. "No frailty fixed effect" removes fixed heterogeneity but retains all shocks.

Age		**Bad=Fair/Poor**				**Bad=Poor**		
		Good	Bad	Dead		Good	Bad	Dead
25–29	Good	88.55	7.19	0.00	Good	94.41	1.48	0.00
	Bad	38.70	57.52	0.00	Bad	35.16	57.28	0.00
30–34	Good	89.66	6.76	0.00	Good	95.93	0.67	0.00
	Bad	33.32	64.96	0.00	Bad	45.06	55.92	0.00
35–39	Good	89.81	6.88	0.00	Good	93.84	2.04	0.00
	Bad	17.35	75.78	0.00	Bad	38.35	62.62	0.00
40–44	Good	89.89	7.01	0.00	Good	95.27	1.84	0.00
	Bad	19.46	78.46	0.00	Bad	49.91	47.84	0.00
45–49	Good	88.17	9.06	0.00	Good	94.89	2.21	0.00
	Bad	20.45	75.78	0.00	Bad	24.09	71.12	0.00

Table 42: Annual transition probabilities across good SRHS, bad SRHS, and death for men without a high-school degree ages 25 to 49. Transition rates when bad SRHS is defined as SRHS equals fair or poor (poor only) are on the left (right). Data source: PSID.



(a) Bottom half

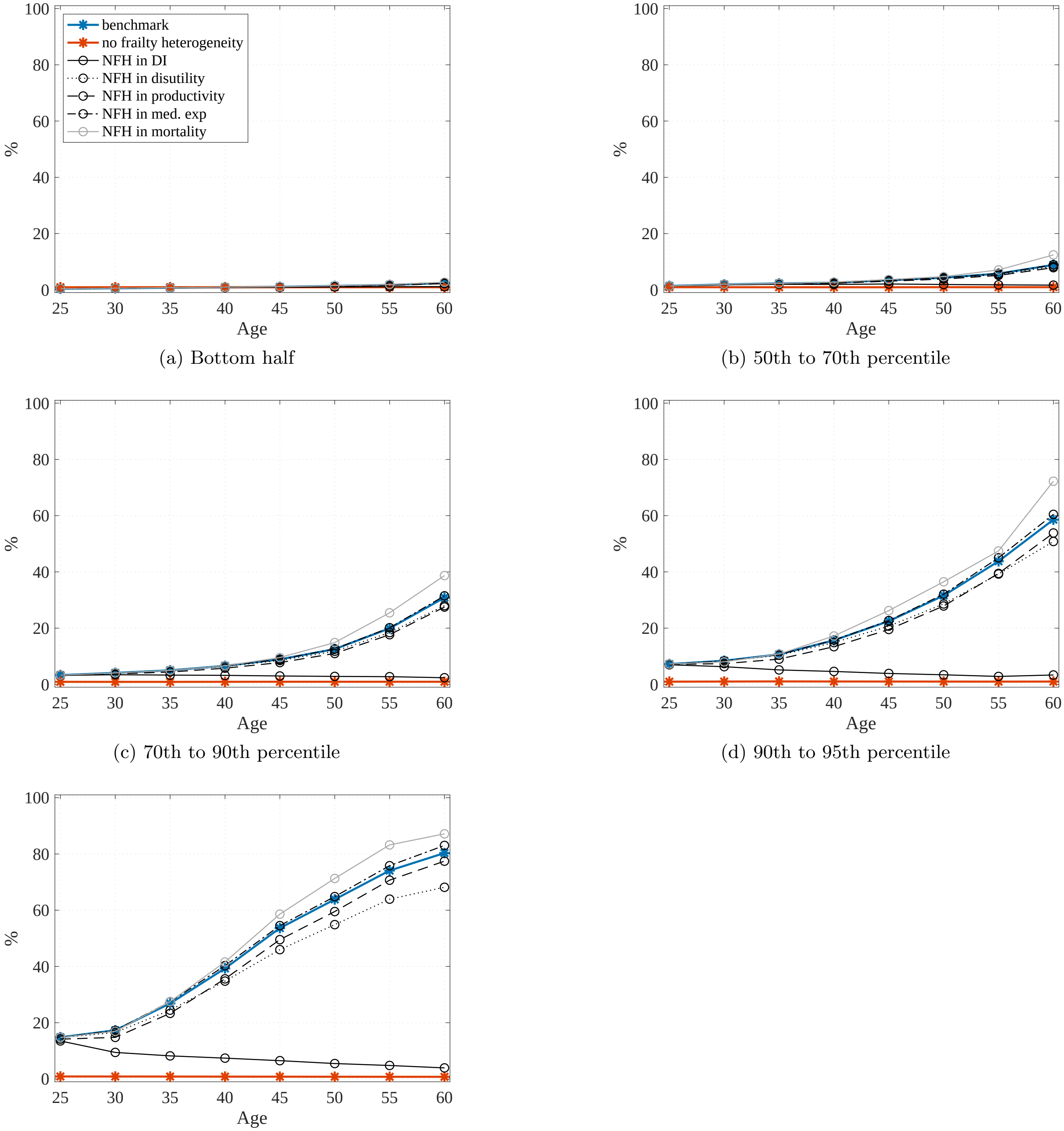
(b) 50th to 70th percentile

(c) 70th to 90th percentile

(d) 90th to 95th percentile

(e) 95th and above

Figure 12: Employment rates by age and frailty percentile groups: comparison between the benchmark (blue), no-frailty-heterogeneity (red), and additional five counterfactual (black and grey) economies.



(a) Bottom half

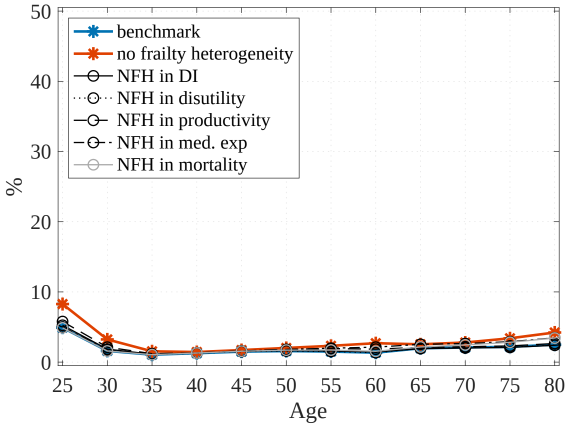
(b) 50th to 70th percentile

(c) 70th to 90th percentile

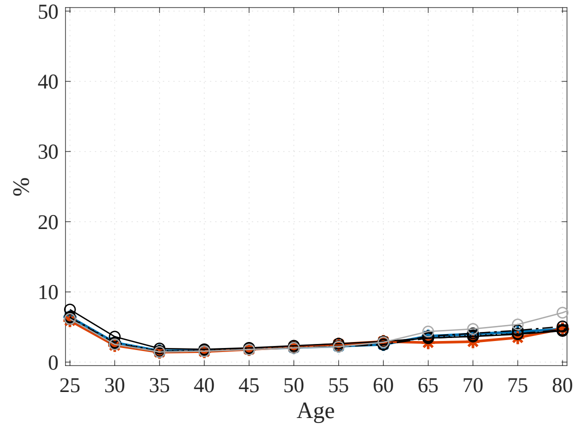
(d) 90th to 95th percentile

(e) 95th and above

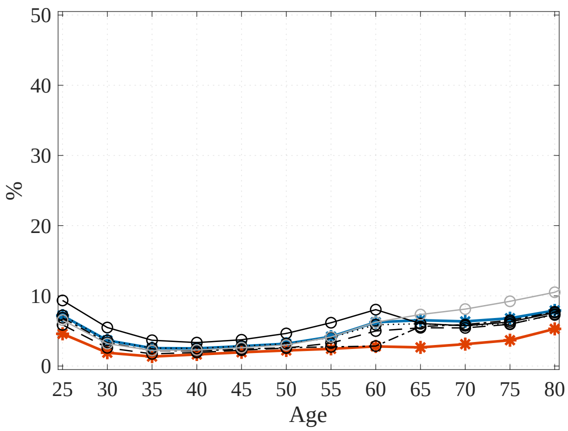
Figure 13: DI recipiency rates by age and frailty percentile groups: comparison between the benchmark (blue), no-frailty-heterogeneity (red), and additional five counterfactual (black and grey) economies.



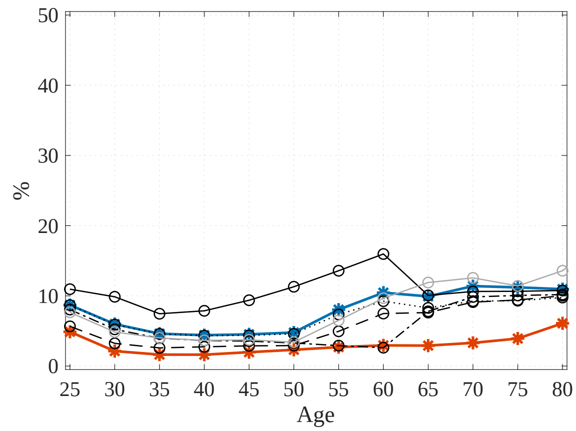
(a) Bottom half



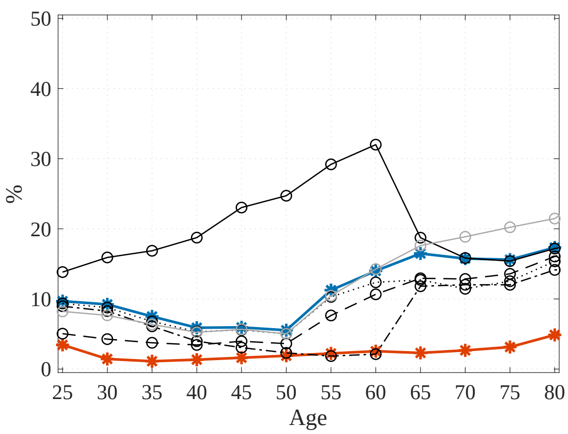
(b) 50th to 70th percentile



(c) 70th to 90th percentile

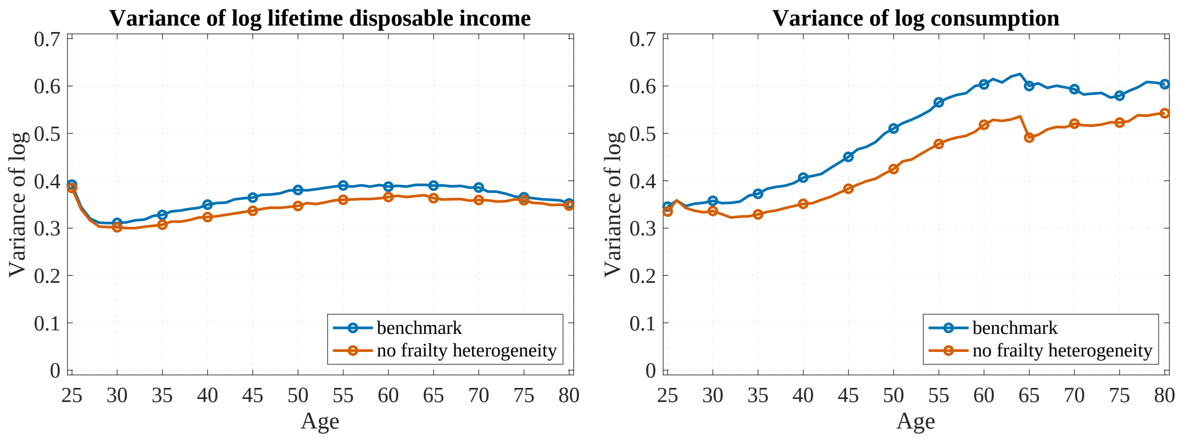


(d) 90th to 95th percentile



(e) 95th and above

Figure 14: Means-tested transfer recipiency rates by age and frailty percentile groups: comparison between the benchmark (blue), no-frailty-heterogeneity (red), and additional five counterfactual (black and grey) economies.



(a) Variance of log lifetime disposable income (b) Variance of log consumption

Figure 15: Panel (a) is the variance of log lifetime disposable (defined as the sum of labor earnings and transfers net of taxes). Panel (b) is the variance of log consumption. The blue line is the benchmark and the red line is the no-frailty-heterogeneity economy.

Age		**Bad=Fair/Poor**				**Bad=Poor**		
		Good	Bad	Dead		Good	Bad	Dead
	Good	93.21	3.96	0.00	Good	96.51	0.71	0.00
25–29	Bad	32.58	64.92	0.00	Bad	55.88	39.50	0.00
	Good	94.54	3.64	0.00	Good	97.42	0.73	0.00
30–34	Bad	30.15	67.32	0.00	Bad	43.71	52.73	0.00
	Good	94.73	3.18	0.00	Good	97.27	0.61	0.00
35–39	Bad	21.12	76.48	0.00	Bad	23.75	73.42	0.00
	Good	94.11	4.03	0.00	Good	97.42	0.81	0.00
40–44	Bad	26.02	72.25	0.00	Bad	32.35	61.30	0.00
	Good	94.25	4.41	0.00	Good	97.35	1.17	0.00
45–49	Bad	19.40	78.14	0.00	Bad	22.98	75.25	0.00

Table 43: Annual transition probabilities across good SRHS, bad SRHS, and death for men with a high-school degree only ages 25 to 49. Transition rates when bad SRHS is defined as SRHS equals fair or poor (poor only) are on the left (right). Data source: PSID.

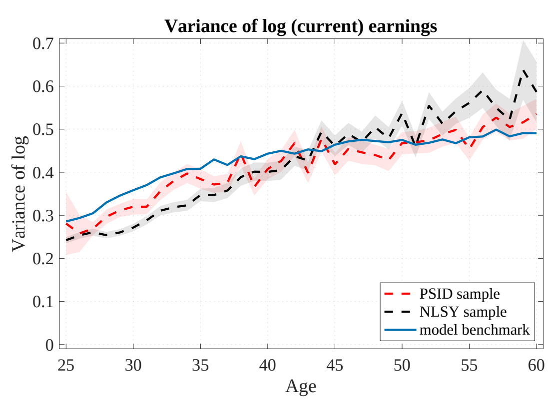


Figure 16: The variance of log current earnings in PSID sample (red), the NLSY79 sample (black), and the benchmark economy (blue).

Age		Bad=Fair/Poor				Bad=Poor		
		Good	Bad	Dead		Good	Bad	Dead
	Good	98.41	0.74	0.00	Good	99.01	0.10	0.00
25–29	Bad	21.91	75.49	0.00	Bad	29.46	70.69	0.00
	Good	98.30	0.96	0.00	Good	99.02	0.26	0.00
30–34	Bad	23.45	76.65	0.00	Bad	29.47	70.66	0.00
	Good	97.26	1.85	0.00	Good	98.92	0.11	0.00
35–39	Bad	42.94	53.79	0.00	Bad	34.77	65.44	0.00
	Good	97.57	1.88	0.00	Good	99.31	0.18	0.00
40–44	Bad	32.13	67.97	0.00	Bad	24.51	75.56	0.00
	Good	96.93	2.40	0.00	Good	99.23	0.06	0.00
45–49	Bad	34.91	63.74	0.00	Bad	34.70	65.45	0.00

Table 44: Annual transition probabilities across good SRHS, bad SRHS, and death for men with a college degree ages 25 to 49. Transition rates when bad SRHS is defined as SRHS equals fair or poor (poor only) are on the left (right). Data source: PSID.

Age		Bad=Fair/Poor				Bad=Poor		
		Good	Bad	Dead		Good	Bad	Dead
	Good	85.64	14.13	0.23	Good	94.54	4.68	0.78
50–54	Bad	16.50	81.03	2.47	Bad	27.03	69.22	3.75
	Good	86.10	13.47	0.43	Good	94.60	4.56	0.84
55–59	Bad	12.88	83.75	3.36	Bad	22.45	71.02	6.53
	Good	86.57	12.66	0.76	Good	94.21	4.47	1.32
60–64	Bad	12.79	83.26	3.95	Bad	21.89	71.02	7.09
	Good	84.41	13.70	1.89	Good	92.54	4.83	2.63
65–69	Bad	14.21	80.15	5.64	Bad	22.40	68.38	9.21
	Good	83.20	14.27	2.53	Good	91.58	5.40	3.02
70–74	Bad	12.82	80.04	7.15	Bad	21.07	65.52	13.41
	Good	79.25	17.42	3.33	Good	87.85	7.23	4.92
75–79	Bad	12.90	75.41	11.70	Bad	18.25	62.53	19.22
	Good	77.43	17.38	5.19	Good	84.38	8.65	6.97
80–84	Bad	10.63	73.43	15.94	Bad	18.85	56.04	25.11
	Good	69.90	20.97	9.13	Good	80.22	8.02	11.75
85–89	Bad	13.99	62.48	23.53	Bad	16.02	49.42	34.56
	Good	63.75	19.81	16.44	Good	73.38	7.06	19.55
90–110	Bad	12.86	56.98	30.16	Bad	16.24	46.62	37.14

Table 45: Annual transition probabilities across good SRHS, bad SRHS, and death for men without a high school degree ages 50 to 94. Transition rates when bad SRHS is defined as SRHS equals fair or poor (poor only) are on the left (right). Data source: HRS.

Age		Bad=Fair/Poor				Bad=Poor		
		Good	Bad	Dead		Good	Bad	Dead
50–54	Good	93.59	6.13	0.29	Good	97.37	2.08	0.54
	Bad	20.56	76.69	2.75	Bad	30.97	64.09	4.94
55–59	Good	92.59	6.86	0.54	Good	97.02	2.15	0.82
	Bad	17.42	79.19	3.40	Bad	23.56	69.92	6.51
60–64	Good	92.05	7.16	0.80	Good	96.57	2.24	1.19
	Bad	18.01	76.98	5.01	Bad	23.68	65.61	10.72
65–69	Good	91.50	7.60	0.90	Good	96.01	2.48	1.51
	Bad	16.52	76.84	6.64	Bad	20.13	65.65	14.22
70–74	Good	89.11	9.27	1.62	Good	94.44	2.87	2.69
	Bad	15.09	75.31	9.60	Bad	19.82	63.08	17.11
75–79	Good	86.18	10.74	3.07	Good	91.84	3.97	4.19
	Bad	14.93	73.40	11.67	Bad	18.47	59.93	21.60
80–84	Good	82.79	12.62	4.59	Good	88.70	4.44	6.85
	Bad	13.47	70.73	15.80	Bad	20.29	56.15	23.56
85–89	Good	73.27	18.80	7.93	Good	81.75	7.06	11.19
	Bad	15.11	60.16	24.73	Bad	20.50	39.56	39.94
90–110	Good	67.44	13.89	18.66	Good	75.39	5.83	18.77
	Bad	13.29	57.33	29.38	Bad	12.57	36.51	50.92

Table 46: Annual transition probabilities across good SRHS, bad SRHS, and death for men with a high school degree only ages 50 to 94. Transition rates when bad SRHS is defined as SRHS equals fair or poor (poor only) are on the left (right). Data source: HRS.

Age		Bad=Fair/Poor				Bad=Poor		
		Good	Bad	Dead		Good	Bad	Dead
50–54	Good	96.93	2.86	0.21	Good	98.92	0.77	0.31
	Bad	23.69	72.33	3.98	Bad	27.18	61.62	11.20
55–59	Good	96.74	2.96	0.30	Good	98.79	0.75	0.47
	Bad	21.02	75.96	3.03	Bad	26.20	68.90	4.91
60–64	Good	96.35	3.30	0.34	Good	98.42	0.90	0.68
	Bad	21.09	73.07	5.84	Bad	23.27	66.51	10.22
65–69	Good	95.23	4.09	0.68	Good	97.56	1.29	1.15
	Bad	19.32	72.40	8.27	Bad	21.63	60.07	18.30
70–74	Good	93.03	5.53	1.44	Good	96.54	1.52	1.94
	Bad	19.15	73.25	7.60	Bad	21.55	64.23	14.21
75–79	Good	90.87	7.06	2.07	Good	94.72	2.42	2.86
	Bad	15.26	74.52	10.21	Bad	23.43	58.75	17.81
80–84	Good	87.09	9.35	3.56	Good	91.03	3.44	5.53
	Bad	12.76	67.83	19.42	Bad	14.05	52.73	33.22
85–89	Good	82.66	10.98	6.35	Good	84.86	5.82	9.32
	Bad	10.53	63.74	25.72	Bad	19.79	41.45	38.76
90–94	Good	79.90	10.36	9.74	Good	84.53	4.33	11.14
	Bad	13.72	62.31	23.97	Bad	12.95	48.56	38.49

Table 47: Annual transition probabilities across good SRHS, bad SRHS, and death for men with a college degree ages 50 to 94. Transition rates when bad SRHS is defined as SRHS equals fair or poor (poor only) are on the left (right). Data source: HRS.

	Bad=Fair/Poor			Bad=Poor	
	Good	Bad		Good	Bad
HSD	0.87	0.13	HSD	0.99	0.01
HSG	0.93	0.07	HSG	0.99	0.01
COL	0.98	0.02	COL	1.00	0.00

Table 48: Distribution across SRHS by education for men ages 25 to 26. The left-hand-side (right-hand-side) columns are for bad SRHS defined as SRHS equals fair or poor (poor only). Data source: PSID.

Table 49: Effects of removing DI (alternative treatment of men on DI at age 25)

	Benchmark	No DI benefits & tax		
		P.E.	G.E.1	G.E.2
Welfare (% relative to benchmark)				
All	n.a.	0.13	-0.11	-0.40
HSD	n.a.	-2.05	-2.27	-3.23
HS	n.a.	-0.20	-0.44	-0.80
CL	n.a.	1.67	1.40	1.54
Variance				
log lifetime earnings (at age 65)	0.586	0.470	0.471	0.455
log lifetime disp. income (at age 65)	0.390	0.378	0.379	0.373
log consumption (overall)	0.513	0.516	0.516	0.506
Change relative to benchmark (%)				
GDP	n.a.	2.84	2.82	3.15
Consumption	n.a.	2.99	2.71	3.13
Capital	n.a.	2.84	2.82	3.15
Labor input	n.a.	2.84	2.82	3.15
Hours	n.a.	3.87	3.81	4.92
GDP per hour	n.a.	-0.99	-0.95	-1.68
Fraction (%)				
Working (25- to 74-year-olds)	83.70	86.97	86.92	87.85
On DI (25- to 64-year-olds)	6.04	0.00	0.00	0.00
On means tested transfers (all)	3.41	5.39	5.44	4.46
Policy Variables				
Payroll tax rate (%)	12.40	10.50	10.50	10.50
Min. consumption (2000 $)	$4375	$4375	$4375	$4089
Tax function parameter (λ)	0.9200	0.9200	0.9178	0.9200

Note: P.E. is the economy without DI in *partial equilibrium*: SSDI/SSI benefits and corresponding fraction of payroll tax are removed. G.E.1 and G.E.2 are *general equilibrium* economies with the overall government budget balanced by adjusting the income tax and minimum consumption, respectively.

Table 50: Additional statistics on the effects of DI program without redistribution

	Benchmark	No DI		DI w/o redistribution	
		P.E.	G.E.1	P.E.	G.E.1
% working (25- to 74-year-olds)					
ALL	83.70	86.70	86.65	83.49	83.49
HSD	72.96	77.41	77.31	72.05	72.04
HS	82.22	86.03	85.95	82.07	82.06
CL	89.27	90.64	90.63	89.19	89.19
% on DI (25- to 64-year-olds)					
ALL	6.04	0.00	0.00	6.09	6.09
HSD	13.73	0.00	0.00	13.96	13.96
HS	7.37	0.00	0.00	7.42	7.42
CL	1.60	0.00	0.00	1.58	1.58
% on means-tested transfers					
ALL	3.41	5.68	5.74	3.63	3.63
HSD	7.41	14.60	14.71	8.50	8.52
HS	4.26	7.02	7.11	4.47	4.47
CL	0.97	1.02	1.03	0.93	0.93
DI payroll tax rate (%)					
HSD	12.4	10.5	10.5	15.6	15.6
HS	12.4	10.5	10.5	13.2	13.2
CL	12.4	10.5	10.5	11.1	11.1
Tax function parameter (λ)	0.9200	0.9200	0.9173	0.9200	0.9196

Note: 'Remove DI' is an economy in which SSDI/SSI benefits and corresponding fraction of payroll tax are removed. 'DI w/o Redistribution' is an economy without redistribution in DI, in which DI benefits of each education group are paid for by education-specific payroll taxes. P.E. is *partial equilibrium* and G.E.1 is *general equilibrium* where the overall government budget is balanced by adjusting the income tax.

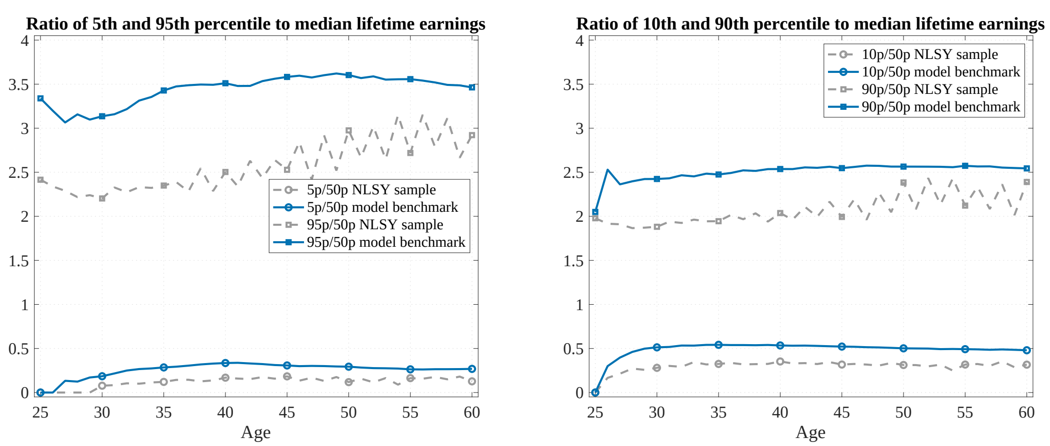
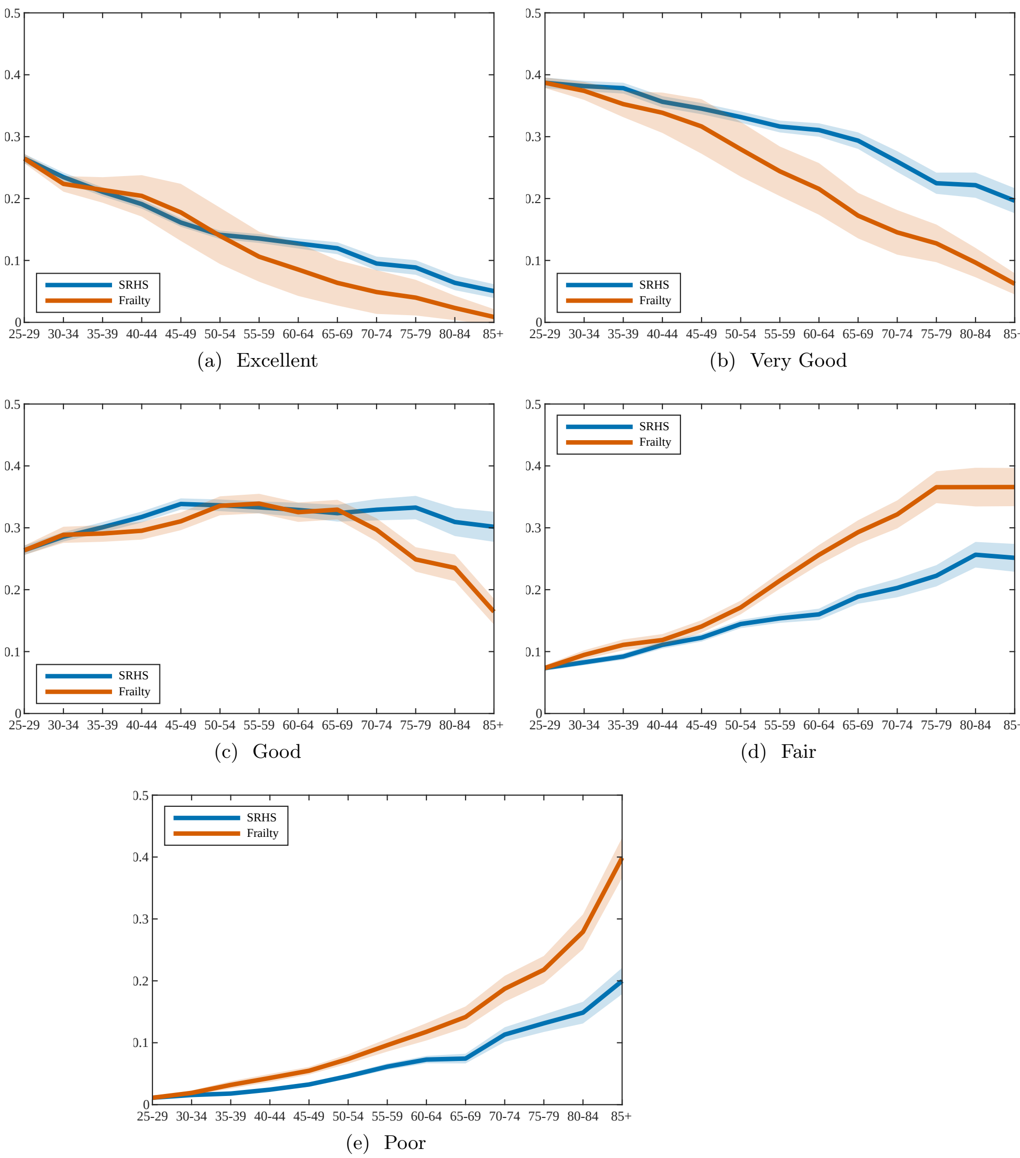


Figure 17: Left panel is the ratio of 5th percentile and 95th percentile of lifetime earnings in NLSY sample (gray) and the benchmark economy (blue). Right panel is the ratio of 10th percentile and 90th percentile of lifetime earnings in NLSY sample (gray) and the benchmark economy (blue).



(a) Excellent (b) Very Good (c) Good (d) Fair (e) Poor

Figure 18: Distribution of health status by age. The blue line shows the share of each SRHS at each 5 year age group. The red line shows the share of each frailty category at each 5 year age group. Cutoffs for frailty categories are chosen so that the distributions are aligned at ages 25 to 29. Shaded areas are 95 percent confidence intervals based on 1000 independent draws of the joint distribution of age, frailty, and SRHS. Source: authors' calculations using PSID.

	Excellent	Very Good	Good	Fair	Poor
	Ages 25 to 49				
	Excellent	Very Good	Good	Fair	Poor
Excellent	53.4	33.1	11.1	2.0	0.3
Very Good	15.7	55.2	25.3	3.3	0.4
Good	6.2	27.7	52.0	12.6	1.4
Fair	3.0	10.9	34.7	43.5	7.9
Poor	2.4	3.8	14.6	38.3	41.0
	Ages 50 to 74				
	Excellent	Very Good	Good	Fair	Poor
Excellent	57.1	32.2	7.9	2.2	0.6
Very Good	11.4	58.7	25.9	3.4	0.7
Good	2.6	21.5	57.1	16.1	2.7
Fair	1.4	5.3	28.6	51.1	13.6
Poor	0.9	2.8	8.2	28.9	59.3
	Ages 75 and older				
	Excellent	Very Good	Good	Fair	Poor
Excellent	42.2	31.1	18.0	5.2	3.5
Very Good	8.4	48.3	28.7	9.8	4.9
Good	2.7	17.1	53.2	20.2	6.9
Fair	1.0	5.9	20.3	50.2	22.6
Poor	0.9	4.7	10.1	24.6	59.7

Table 51: Two-year transition probabilities (%) across SRHS categories within three age groups: 25–49 year-olds, 50–74 year-olds, and ages 75 and older constructed using the PSID sample. Rows sum to one.

	Excellent	Very Good	Good	Fair	Poor
	Ages 25 to 49				
	Excellent	Very Good	Good	Fair	Poor
Excellent	80.5	17.0	2.3	0.2	0.1
Very Good	3.1	80.2	15.4	1.1	0.2
Good	0.1	8.0	79.0	11.7	1.3
Fair	0.1	0.6	11.7	75.5	12.1
Poor	0.1	0.8	2.6	23.3	73.2
	Ages 50 to 74				
	Excellent	Very Good	Good	Fair	Poor
Excellent	78.6	17.8	3.0	0.5	0.1
Very Good	1.8	75.6	20.2	2.0	0.4
Good	0.0	4.3	78.6	15.9	1.1
Fair’	0.0	0.2	7.7	76.4	15.7
Poor	0.0	0.1	0.6	16.0	83.4
	Ages 75 and older				
	Excellent	Very Good	Good	Fair	Poor
Excellent	64.5	25.5	5.5	2.7	1.8
Very Good	0.7	60.9	28.1	7.3	3.0
Good	0.0	3.4	61.0	30.6	5.1
Fair	0.0	0.2	6.2	67.2	26.4
Poor	0.0	0.0	0.5	12.8	86.8

Table 52: Two-year transition probabilities (%) across frailty health categories within three age groups: 25–49 year-olds, 50–74 year-olds, and ages 75 and older constructed using the PSID sample. Rows sum to one.

References

Al-Sadoon, M. M., S. Jiménez-Martín, J. M. Labeaga et al., "Simple methods for consistent estimation of dynamic panel data sample selection models," Technical Report, 2019. 29

Andrews, I., M. Gentzkow and J. M. Shapiro, "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics* 132 (2017), 1553–1592. 32

Arellano, M. and S. Bond, "Some Tests of Specification for Panel Data: Monte Carol Evidence and and Application to Employment Equations," *The Review of Economic Studies* 58 (1991), 277–297. 5

Arellano, M. and O. Bover, "Another look at the instrumental variable estimation of error-components models," *Journal of econometrics* 68 (1995), 29–51. 4

Arellano, M. A. and B. E. Honoré, "Panel Data Models: Some Recent Developments," *Handbook of Econometrics* 5 (2001), 3229–3296. 7

Beck, T., R. Levine and N. Loayza, "Finance and the Sources of Growth," *Journal of financial economics* 58 (2000), 261–300. 4

Bell, F. C. and M. L. Miller, "Life tables for the United States social security area," *Social Security Administration Publications* (2005). 28

Blundell, R. and S. Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics* 87 (1998), 115–143. 4, 11

———, "GMM estimation with persistent panel data: an application to production functions," *Econometric reviews* 19 (2000), 321–340. 4, 11

Bond, S. R., "Dynamic panel data models: a guide to micro data methods and practice," *Portuguese economic journal* 1 (2002a), 141–162. 5, 16

———, "Dynamic panel data models: a guide to micro data methods and practice," *Portuguese economic journal* 1 (2002b), 141–162. 5

French, E. and J. Song, "The effect of disability insurance receipt on labor supply," *American economic Journal: economic policy* 6 (2014), 291–337. 32

Guvenen, F., "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics* 12 (2009), 58–79. 30

Hansen, L. P., "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50 (1982), 1029–1054. 4

Heathcote, J., F. Perri and G. Violante, "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006," *Review of Economic Dynamics* 13 (2010), 15–51. 1

Hosseini, R., K. Kopecky and K. Zhao, "The Evolution of Health over the Life Cycle," *Review of Economic Dynamics* 45 (2022), 237–263. 2, 3, 17, 23, 46, 47

Hubbard, R. G., J. Skinner and S. P. Zeldes, "Precautionary Savings and Social Insurance," *Journal of Political Economy* 103 (1995), 360–399. 40

Low, H. and L. Pistaferri, "Disability insurance and the dynamics of the incentive insurance trade-off," *American Economic Review* 105 (2015), 2986–3029. 29, 30, 31, 39

Nickell, S., "Biases in dynamic models with fixed effects," *Econometrica: Journal of the Econometric Society* (1981), 1417–1426. 5

Roodman, D., "A note on the theme of too many instruments," *Oxford Bulletin of Economics and statistics* 71 (2009), 135–158. 5, 8

Sargan, J. D., "The estimation of economic relationships using instrumental variables," *Econometrica: Journal of the econometric society* (1958), 393–415. 4

Searle, S., A. Mitnitski, E. Gahbauer, T. Gill and K. Rockwood, "A standard procedure for creating a frailty index," *BMC Geriatrics* 8 (2008), 1. 1

Staiger, D. and J. H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica* 65 (1997), 557–586. 9, 10

Stock, J. H. and M. Yogo, "Testing for weak instruments in linear IV regression," in D. W. Andrews and J. H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg* (Cambridge, UK: Princeton University Press, 2005), 80–108. 9

Wintoki, M. B., J. S. Linck and J. M. Netter, "Endogeneity and the dynamics of internal corporate governance," *Journal of Financial Economics* 105 (2012), 581–606. 9, 10

Yang, Y. and L. C. Lee, "Dynamics and heterogeneity in the process of human frailty and aging: evidence from the US older adult population," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* (2009), gbp102. 1