Lecture 6: Topics in Intertemporal Labor Supply

a. the extensive margin

b. wage and earnings generating functions, partial insurance models

Some References

extensive margin:

Marco Bianchi, Björn R. Gudmundsson, Gylfi Zoega. "Iceland's Natural Experiment in Supply-Side Economics." AER 91(5) 2001, pp. 1564-1579.

Raj Chetty, Adam Guren, Day Manoli, and Andrea Weber. "Does Indivisible Labor Explain the Di¤erence between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." NBER Macroeconomics Annual 2012.

Day Manoli and Andrea Weber. "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions." July 2013. Available at http://weber.vwl.uni-mannheim.de/2476.0.html

wage and earnings generating functions, partial insurance models:

John Abowd and David Card (1989), "On the Covariance Structure of Earnings and Hours Changes." Econometrica, 57(2): 411–45.

Altonji, Joseph G & Segal, Lewis (1996). "Small-Sample Bias in GMM Estimation of Covariance Structures," Journal of Business & Economic Statistics, 14(3): 353-66.

Michael Baker and Gary Solon (2003). "Earnings Dynamics and Inequality among Canadian Men, 1976–1992: Evidence from Longitudinal Income Tax Records." JOLE 21(2): 289-321

Steven Haider and Gary Solon (2006). "Life-Cycle Variation in the Association between Current and Lifetime Earnings." AER 96(4): 1308-1320.

Fatih Guvenen (2007). "Learning Your Earning: Are Labor Income Shocks Really Very Persistent? AER 97 (3): 687-712

Øivind Anti Nilsen et al. (2012). "Intergenerational Earnings Mobility Revisited: Estimates Based on Lifetime Earnings." Scandanavian J of Economics 114(1): 1–23.

Uta Schoenberg (2007). "Testing for Asymmetric Employer Learning." JOLE 25(4): 651-691.

Richard Blundell, Luigi Pistaferri and Ian Preston (2008), "Consumption Inequality and Partial Insurance" AER 98(5): 1997-1921

Jonathan Heathcote, Kjetil Storeslettern and Giovanni Violante (2014), "Consumption and Labor Supply with Partial Insurance: An Analytical Framework" AER 107(7): 2075-2126

The Extensive margin

A lot of the labor supply literature ignores the *extensive margin* – workers who don't work for a year are dropped. However, variation in the number of workers is potentially important for understanding aggregate movements in hours:

(a) some people do miss an entire year of work in downturns

(b) the elasticity of participation w.r.t. wages can be relatively high, even if η is small.

There is a literature in macro arguing that the extensive margin is highly elastic, and that the extensive margin needs to be taken into consideration in both tax policy analysis and in macro modeling (see Chetty et al for a discussion of this literature).

Chetty et al present a meta analysis of various quasi-experimental studies that measure the effects of either permanent changes in (after tax) wages, or temporary changes, on employment rates. They use the former to obtain estimates of compensated elasticities of participation; the latter provide estimates of the Frisch elasticities of participation. An interesting paper is the one by Bianchi et al (2001), on the effects of a tax holiday created in Iceland when the country switched tax systems and everyone was untaxed for a single year (1987). You may find it instructive to read the paper because it is almost impossible to understand what the original authors did (or why), despite the very clear research design.

Looking at Chetty et al's Table 1, notice that the typical compensated elasticity is around 0.25, while the typical Frish elasticity is around 0.3. These are not much different than the elasticities people have obtained for the intensive margin.

Manoli and Weber (2013) is a very recent attempt to look at one of the important extensive margins : variation in the length of time people work. This paper uses an RD design to study the effects of a benefit that is paid to workers who retire after certain tenure milestones: see their Figure 1 at the end of the notes. For example, if you retire with 11-14 years of tenure you get 1/3 of a year of salary, whereas if you retire with 15-19 years you get 1/2 year of salary.Since workers start jobs at different ages, there is a smooth distribution of people across the tenure distribution at different ages, and Manoli and Weber find strong evidence that some workers appear to delay retirement to get the benefits – see their Figure 4, which shows spikes in tenure just after the milestones.

They use a variant of the bunching style estimator we discussed in Lecture 3 to relate the fraction of people who retire at the threshold point to the relative size of the extra severance payment available for those who reach the threshold. Specifically, they smooth the density of retirement tenures (see Figure 13), then for each milestone they calculate

$$\frac{\Delta p}{p_t} = \frac{\sum_{k=0}^{11}(r_{t+k} - r_{t+k}^S)}{\sum_{k=0}^{11} r_{t+k}^S}$$

where r_{t+k} is the fraction of retirements at month k after the milestone tenure level t, and r_{t+k}^S is the corresponding smoothed fraction. In words, this is the excess mass of retirements in the year just after the milestone. They then calculate a simple "extensive margin elasticity":

$$\epsilon = \frac{\frac{\Delta p}{p_t}}{\frac{(1-\tau(sev))SP_t}{(1-\tau(earn))y_t}},$$

where SP_t is the extra severance pay after the milestone t (e.g., 1/6 of a year of salary for reaching 15 years), $\tau(sev)$ is the tax rate on severance pay, which is low, and $\tau(earn)$ is the effective tax rate on an additional year of work, which is about 80% – arising from a combination of a 30 payroll/income tax and a 50% replacement rate from the pension system. See Table 4 at the end of the notes for the calculations. In the top panel they calculate the denominator of the elasticity using the "rules" - these are around 0.1 to 0.3 . A problem is that tax records show that some people get the severance even if they retire a bit early, and others don't seem to get it even if they pass the milesone – see Figure 8 at the end of the notes for the distributions of SP payment fractions. So in Panel B of Table 4 they calculate the denominator by estimating the relative gain in severance pay from exceeding the milestone – similar to the first stage in a fuzzy RD. These elasticities are larger because the gain in realized SP is smaller than the gain implied by the formula.

Wage and Earnings Generating Functions, Labor Supply with Incomplete Markets

As discussed in lecture 5, an important question for interpreting the reaction of hours to wage changes is to what extent wage innovations are expected to persist? Persist shocks have income effects that offset the inter-temporal substitution effect; transitory shocks have small (igorable) income effects. This question is closely related to a longstanding question in macro: how much of any observed rise in labor earnings today is expected to persist into the future? If earnings changes are purely tranistory, then a fall/rise in earnings today has little impact of workers' lifetime earnings and they would not be expected to lower or raise their consumption by much, assuming they have access to "perfect" credit markets.

We begin in this lecture by discussing methods for estimating micro level "wage generating functions" (WGF) and "earnings generating functions" (EFG). These are models of the stochastic processes that generate wages and earnings over time - typically for male earners who work every period for sure, or for families with at least one earner who works in every period.

WGF's

Recall that Pistaferri (2003) assumes that an appropriate model for individual wages is:

$$\begin{aligned} \log w_{it} &= \omega_i + u_{it} \quad , \\ u_{it} &= u_{it-1} + \zeta_{it} \end{aligned}$$

where the ζ_{it}'s are uncorrelated over time. This is a *pure random walk* model, in which $E[\log w_{it+j}|\log w_{it}] = \log w_{it}$. A more general model is

$$\begin{aligned} \log w_{it} &= \omega_i + x_{it}\beta_t + u_{it} + e_{it} \qquad (1) \\ u_{it} &= \alpha u_{it-1} + \zeta_{it} \quad , \end{aligned}$$

where e_{it} and ζ_{it} are serially uncorrelated and uncorrelated with each other. This model includes a fixed component ω_i, a component attributable to observables x_{it}, an AR(1) component u_{it}, and a purely transitory component e_{it}. We will discuss how to estimate the parameters of this model using simple method of moments.

A standard method is to first regress $\log w_{it}$ on x_{it}, and treat the residuals r_{it} as estimates of the combined error component $\omega_i + u_{it} + e_{it}$. Then we form the covariance matrix C of the residuals and fit a model to the vector of elements of C. Let

$$\begin{aligned} \sigma_\omega^2 &= var[\omega_i] \\ \sigma_{u0}^2 &= var[u_{i0}], \\ v_t &= var[\zeta_{it}] \end{aligned}$$

Notice that:we can write

$$r_{it} = \omega_i + \alpha^t u_{i0} + \alpha^{t-1}\zeta_{i1} + ... + \alpha^t \zeta_{it-1} + \zeta_{it} + e_{it}$$

which implies that

$$\begin{aligned} var[r_{i1}] &= \sigma_\omega^2 + \alpha^2\sigma_{u0}^2 + v_1 + var[e_{i1}], \\ var[r_{it}] &= \sigma_\omega^2 + \alpha^{2t}\sigma_{u0}^2 + v_t + \alpha^2 v_{t-1} + ... + \alpha^{2(t-1)}v_1 + var[e_{it}], \\ cov[r_{it}, r_{is}] &= \sigma_\omega^2 + \alpha^{s+t}\sigma_{u0}^2 + \alpha^{t-s}v_s + \alpha^{t-s+2}v_{s-1} + ... + \alpha^{s+t-2}v_1, \quad (s < t) \end{aligned}$$

The term σ_{u0}^2 represents an initial conditions effect: it is the effect of the dispersion in the pre-sample value of u_{it}, which gradually fades out if $\alpha < 1$. It is a matter of algebra to show that if $var[e_{it}]$ is constant, and all the $v_t' s$ are constant (i.e., $v_t = v$), and if $\sigma_{u0}^2 = v/(1-\alpha^2)$, (its steady state value) then the variances of r_{it} are all constant. If $var[e_{it}]$ and all the $v_t' s$ are constant but $\sigma_{u0}^2 < v/(1-\alpha^2)$, the variances of r_{it} rise over time.

As written, the model in equation (1) assumes that the permanent component of wage heterogeneity (ω_i) contributes a fixed amount (σ_ω^2) to the variance of wages in all periods, and to the covariances at all leads/lags. If there is skill biased technical change, we might expect that differences in wages between people with different levels of skill will rise over time. One way to build that idea into (1) is to assume that there are a set of loading factors ψ_t that vary over time, with $\psi_1 = 1$ for some base period:

$$\begin{aligned} \log w_{it} &= \psi_t(\omega_i + x_{it}\beta_t + u_{it} + e_{it}) \qquad (2) \\ &= x_{it}\beta_t' + \psi_t(\omega_i + u_{it} + e_{it}) \end{aligned}$$

where $\beta_t' = \psi_t \beta_t$. Notice that I am assuming here that all 4 components are scaled by the same loading factor in each period. In general that need not be true. For example, if you think that e_{it} includes both productivity components and measurement error, then this component may not get scaled up/down over time the same as the pure productivity components. Equation (2) leads to expressions for the variances and covariances of the wage residuals that are relatively simple but incorporate an alternative source of non-stationarity. Card and Lemieux (AER, March 1994) used a model like (2) to evaluate the role of rising return to skill in leading to widening wage differences between black and white workers. Baker and Solon (2003) use a model like (2) to look at earnings dynamics in Canada.

Several recent studies (eg Haider and Solon, 2006; Schoenberg, 2007) have argued that the loading factor on the permanent component ω_i rises with age (rather than, or in addition to, changing over time). There are several explanations for this: one explored by Schoenberg is that it takes time for the market to figure out who is high ability. Another is that high ability people invest more in on-the-job training in their youth, depressing their wages relative to their long term average. The recent paper by Nilsen et al. (2012) shows data from several different countries suggesting that there is a lifecycle pattern in the loading factor on the permanent component of earnings.

A third class of earnings models assumes that there are person-specific growth rates in wages or earnings (for an early version, see Ashenfelter and Card, 1985; Guvenen 2007 is the most prominent recent paper taking this line). For example, ignoring the $x's$ and the loading factors, suppose:

$$\log w_{it} = \omega_i + \rho_i t + u_{it} + e_{it} \tag{3}$$

where

$$\begin{aligned} \sigma_\rho^2 &= var[\rho_i] \\ \sigma_{\rho\omega} &= cov[\rho_i, \omega_i] \\ 0 &= cov[\rho_i, u_{it}] \\ 0 &= cov[\rho_i, e_{it}] \end{aligned}$$

In this setup, the random trend is allowed to be correlated with the permanent component, but not the transitory components. This implies that:

$$\begin{aligned} var[r_{i1}] &= \sigma_\omega^2 + \sigma_\rho^2 + 2\sigma_{\rho\omega} + \alpha^2\sigma_{u0}^2 + v_1 + var[e_{i1}], \\ var[r_{it}] &= \sigma_\omega^2 + t^2\sigma_\rho^2 + 2t\sigma_{\rho\omega} + \alpha^{2t}\sigma_{u0}^2 + v_t + \alpha^2 v_{t-1} + ... + \alpha^{2(t-1)}v_1 + var[e_{it}], \\ cov[r_{it}, r_{is}] &= \sigma_\omega^2 + st\sigma_\rho^2 + (s+t)\sigma_{\rho\omega} + \alpha^{s+t}\sigma_{u0}^2 + \alpha^{t-s}v_s + \alpha^{t-s+2}v_{s-1} + ... + \alpha^{s+t-2}v_1, \quad (s < t) \end{aligned}$$

Notice that a random trend generates a very specific form of non-stationarity, with quadratic growth rates in the variances and covariances. An interesting feature of a random trend model is that it implies a positive correlation between growth rates of wages for the same individual in different periods. Taking first differences of equation (3):

$$\Delta \log w_{it} = \rho_i + \Delta u_{it} + \Delta e_{it}$$

Notice that if e_{it} is an i.i.d. process, then Δe_{it} is an MA(1) with 1st order autocorrelation of $-1/2$. If u_{it} is a random walk, then Δu_{it} is serially uncorrelated. If u_{it} is an AR(1) then Δu_{it} and Δu_{is} are correlated, but for t and s far apart, $cov(\Delta u_{it}, \Delta u_{is}) \rightarrow 0$. Thus, one way to look for the presence of a random trend is to see whether wage changes for the same individual at long lags are correlated. Does someone who had faster wage growth from age 25 to 30 have faster wage growth between 40 and 45?

Estimation Methods

In general, for any specific model of the wage generating process, we can write

$$vecltr[C] = m = f(\theta)$$

where θ represents the parameters in the wage process. The method of moments idea is to find a value for θ that gives the best fit to the empirical estimates of m. Call $\widehat{m}$ the estimate of m. In general an element of $\widehat{m}$ is some term in the empirical covariance matrix $\widehat{C}$, say

$$\widehat{m}_k = cov[r_{it}, r_{is}] = \frac{1}{N}\sum_i r_{it} r_{is} = \frac{1}{N}\sum_i m_{ki}$$

(since the residuals have zero mean by construction we don't have to deviate from means). We can construct the sampling variance of the element $\widehat{m}_k$ by

$$\frac{1}{N}\sum_i (m_{ki} - \widehat{m}_k)^2$$

which is just the variance of the second moment in the sample, divided by N, and the sampling covariance between estimates of any two elements $\widehat{m}_k$ and $\widehat{m}_h$ by

$$\frac{1}{N}\sum_i (m_{ki} - \widehat{m}_k)(m_{hi} - \widehat{m}_h).$$

Under regularity conditions (basically, iid sampling and finite *fourth* moments), the vector of estimates of the second moments will have a standard normal distribution with

$$\sqrt{N}(\widehat{m} - m) \rightarrow N(0, V)$$

Moreover, the matrix

$$\widehat{V} = \frac{1}{N}\sum_i (m_i - \widehat{m})(m_i - \widehat{m})'$$

is a consistent estimate of V.

For estimation, one simple choice is least squares:

$$\min_\theta [\widehat{m} - f(\theta)]'[\widehat{m} - f(\theta)]$$

Various GLS variants are also possible. Consider a positive definite matrix A (of the right dimension): then we can use the objective:

$$\min_{\theta}[\widehat{m} - f(\theta)]' A [\widehat{m} - f(\theta)]. \tag{4}$$

Chamberlain (1982) presented the following theorem. Assume:
1. $\widehat{m} \to f(\theta^0)$ almost surely
2. f is continuous in θ in some neighborhood Θ that contains θ^0
3. $f(\theta) = f(\theta^0)$ for θ in $\Theta \Rightarrow \theta = \theta^0$ (i.e, we have identification)
4. $A \to \Psi$ a positive definite matrix

Then the gls estimator $\widehat{\theta}$ based on equation (1) converges almost surely to θ^0.

If in addition:
5. $\sqrt{N}(\widehat{m} - f(\theta^0)) \to N(0, V)$
6. f is 2x continuously differentiable for θ in some neighborhood of θ^0, and

$$F = F(\theta^0) \equiv \frac{\partial f(\theta^0)}{\partial \theta}$$

has full rank, then

$$\sqrt{N}(\widehat{\theta} - \theta^0) \to N(0, \Delta)$$

where

$$\Delta = (F'\Psi F)^{-1} F'\Psi V \Psi F (F'\Psi F)^{-1}.$$

It can also be shown that the optimal choice for A is one such that $A \to V^{-1}$, in which case $\Delta = (F'V^{-1}F)^{-1}$. Notice that the least squares choice $A = I$ leads to the var-cov:

$$\Delta_{ols} = (F'F)^{-1} F'VF(F'F)^{-1}$$

which looks just like the variance matrix you get in a regression model with non-spherical errors when you use OLS. In applications we need to estimate F and V: we will use $\widehat{F} = F(\widehat{\theta})$ and some estimate of $\widehat{V}$.

A nice feature of the optimal weight matrix is that under the null, the minimand

$$N[\widehat{m} - f(\theta)]' V^{-1} [\widehat{m} - f(\theta)]$$

has an asymptotic χ^2 distribution, with degrees of freedom equal to the difference between the number of moments and the number of elements of θ. This provides a general specification test of the validity of the model $m = f(\theta)$. For other weighting matrices there is a similar overall goodness of fit statistic:

$$N[\widehat{m} - f(\theta)]' R^{-} [\widehat{m} - f(\theta)]$$

where R^{-} is a generalized inverse of the matrix $R = (I - F(F'AF)^{-1}F'A)V(I - F(F'AF)^{-1}F'A)$. (This matrix has rank at most equal to the difference between the number of moments and the number of columns of F, which is the number of elements in θ).

As a practical matter the optimal choice for the weighting matrix can lead to substantial problems in small samples. This was not well understood at the time of Abowd-Card, but was pointed out in the paper by Altonji and Segel (1996). It is generally agreed that when the moments of interest are all (roughly) scaled the same (as is true when we consider covariances of log wage residuals) the least squares objective is sensible.

EGF's

Abowd and Card present EGF's for three samples. They show that the first difference of log earnings has a covariance structure that has mainly 0's after more than 2 lags. For example, here is a fragment of the var-cov of changes in (residualized) log annual earnings for continuously employed men in the PSID:

Dy_{69}	Dy_{70}	Dy_{71}	Dy_{72}	Dy_{73}	Dy_{74}
Dy_{70}	0.161				
Dy_{71}	-0.036	0.158			
Dy_{72}	-0.007	-0.064	0.170		
Dy_{73}	0.000	-0.002	-0.062	0.134	
Dy_{74}	0.003	0.000	-0.007	-0.036	0.129

Notice that:

(a) the variances of the changes is nonstationary

(b) $cov[Dy_t, Dy_{t-1}]$ is negative and about 30% as large as the variance (i.e the first order correlation is about -0.3);

(c) $cov[Dy_t, Dy_{t-2}] = cov[Dy_t, Dy_{t-3}] = cov[Dy_t, Dy_{t-4}] \approx 0$

On this basis, a simple EGF that one could posit is:

$$\begin{aligned} log\, y_{it} &= \alpha_i + z_{it} + u_{it} \\ z_{it} &= z_{it-1} + \xi_{it} \end{aligned}$$

where $\{\xi_{it}\}$ are serially uncorrelated, but with different variances over time, and $\{u_{it}\}$ are also serially uncorrelated with potentially time-varying variances. One interpretation of the $u's$ is as measurement errors. Under this "random walk plus noise" assumption:

$$Dy_{it} = log\, y_{it} - log\, y_{it} = \xi_{it} + u_{it} - u_{it-1}$$

If the u's are stationary this model implies that

$$\begin{aligned} var[Dy_{it}] &= var[\xi_{it}] + 2\sigma_u^2 \\ cov[Dy_{it}, Dy_{it-1}] &= -\sigma_u^2 \\ cov[Dy_{it}, Dy_{it-j}] &= 0,\ j > 1 \end{aligned}$$

Blundell, Pistaferri and Preston partial insurance model

The idea in this paper is to combine data on earnings and consumption and use it to estimate "semi-structural" parameters showing the responsiveness of consumption to the permanent and transitory components of earnings. The model is fit to data from the PSID on continuously married families over the period from 1980 to 1992. A problem to begin is that the PSID only has consumption data on food spending. So BPP use data from the CEX to estimate a set of models of the form:

$$log(food - spending_{it}) = x_{it}\delta + \beta log(c_{it}) + e_{it}$$

where c_{it} is total non-durable spending. The estimate a model that allows β to vary by year, by education of the household head, and the number of kids (though none of the interactions are large or significant). The main estimate is $\widehat{\beta} = 0.85$ with std. error. of 0.15 (which is super precise). They then "invert" the model and apply it to the PSID to get *imputed* total non-durable consumption of each family in each year. Thereafter c refers to imputed log of total consumption, which is (essentially) $1.17 \times log(food - spending)$.

The next step of the paper is to specify an EGF. They assume:

$$\begin{aligned} log\, y_{it} &= P_{it} + v_{it} \\ P_{it} &= P_{it-1} + \zeta_{it} \\ v_{it} &= \varepsilon_{it} + \theta\varepsilon_{it-1} \end{aligned}$$

which implies that:

$$\Delta log\, y_{it} = \zeta_{it} + \varepsilon_{it} + (\theta - 1)\varepsilon_{it-1} - \theta\varepsilon_{it-2}$$

Evidence on the goodness of fit of this class of models is shown in their Table 3 (see end of notes). This looks a lot like the A-C EGF for males.

A third step is to specify how consumption is related to earnings. They assume:

$$\Delta c_{it} = \phi\zeta_{it} + \psi\varepsilon_{it} + \xi_{it}$$

They actually present the model as if ϕ and ψ vary across consumers (and over time for consumers), but for the main analysis these are treated as constant. This is a "semi-structural" equation that nests some interesting cases.

a) If people simply adjust consumption to income then $\phi = \psi = 1$

b) If people can borrow and lend freely, but don't knowζ_{it} until it happens, then ϕwill be large but less than 1 (they argue that simulations suggest $\phi = 0.8$) whereas ψ should be pretty small.

Notice that with the EGF and the consumption model we can derive the covariance structure of earnings and consumption changes:

$$\begin{aligned} cov(\Delta log\, y_{it}, \Delta log\, y_{is}) &= var[\zeta_{it}] + var[\Delta v_{it}]\; s = t \\ &= (\theta - 1)var[\varepsilon_{it-1}] + \theta(1 - \theta)var[\varepsilon_{it-2}]\; s = t - 1 \\ &= -\theta var[\varepsilon_{it-2}]\; s = t - 2 \\ &= 0\; s < t - 2 \\ cov(\Delta log\, y_{it}, \Delta c_{it}) &= \phi var[\zeta_{it}] + \psi var[\varepsilon_{it}] \\ cov(\Delta log\, y_{it+1}, \Delta c_{it}) &= \psi(\theta - 1)var[\varepsilon_{it}] \\ cov(\Delta log\, y_{it-1}, \Delta c_{it}) &= 0 \end{aligned}$$

Note that if $\theta \neq 1$ than ψ is identified from $cov(\Delta log\, y_{it+1}, \Delta c_{it})$.

In their estimation, BPP allow a serially uncorrelated but non-stationary measurement error in consumption. Their table 6 reports estimates of the main parameters for their model the whole sample and 4 subgroups (with and without college; born in the 1930s vs. 1940s). Their key estimates for the whole sample are:

$$\begin{aligned} \theta &= 0.113\ (0.025) \\ \phi &= 0.642\ (0.095) \\ \psi &= 0.053\ (0.044) \end{aligned}$$

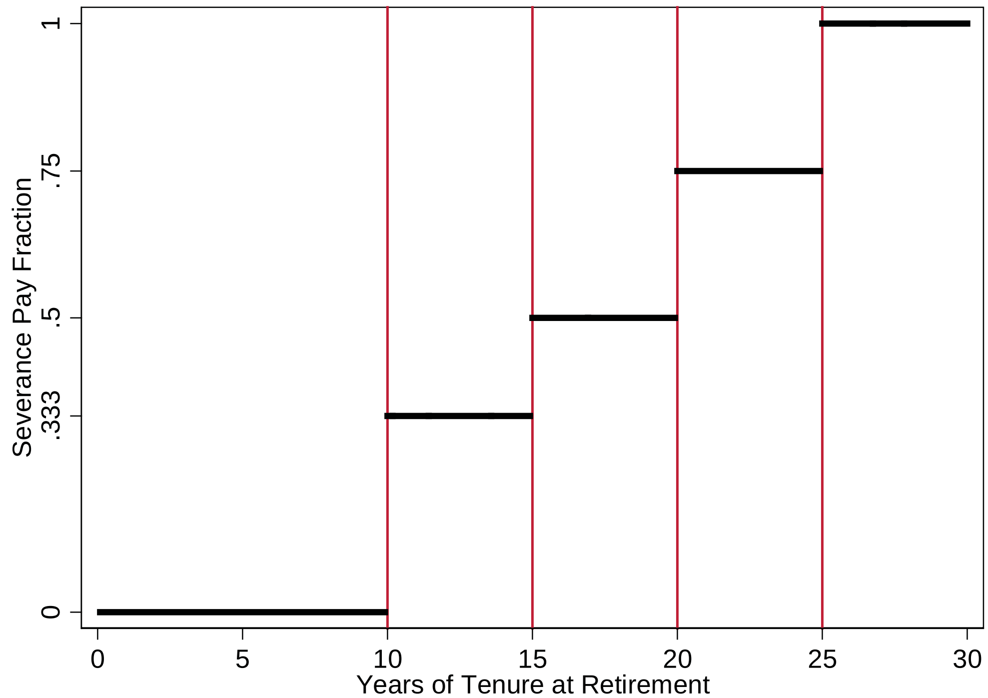
The main heterogeneity is in ϕ.They estimate $\phi = 0.94$ (0.18) for families headed by a male with less than college education, and $\phi = 0.42$ (0.09) for families headed by a male with some college.

TABLE 1
Extensive Margin Elasticity Estimates from Quasi-Experimental Studies

Study	Elasticity	Standard Error	Population and Variation
A. Steady State (Hicksian) Elasticities			
1. Juhn, Murphy, and Topel (1991)	0.13	0.02	Men, skill-specific trends, 1971-1990
2. Eissa and Liebman (1996)	0.30	0.10	Single Mothers, U.S. 1984-1990
3. Graversen (1998)	0.24	0.04	Women, Denmark 1986 tax reform
4. Meyer and Rosenbaum (2001)	0.43	0.05	Single Women, U.S. Welfare Reforms 1985-1997
5. Devereux (2004)	0.17	0.17	Married Women, U.S. wage trends 1980-1990
6. Eissa and Hoynes (2004)	0.15	0.07	Low-Income Married Men & Women, U.S. EITC expansions 1984-1996
7. Liebman and Saez (2006)	0.15	0.30	Women Married to High Income Men, U.S. tax reforms 1991-97
8. Meghir and Phillips (2010)	0.40	0.08	Low-Education Men, U.K. wage trends, 1994-2004
9. Blundell, Bozio, and Laroque (2011)	0.30	n/a	Prime-age Men and Women, U.K., tax reforms 1978-2007
Unweighted Mean	**0.25**		
B. Intertemporal Substitution (Frisch) Elasticities			
10. Carrington (1996)	0.43	0.08	Full Population of Alaska, Trans-Alaska Pipline, 1968-83
11. Gruber and Wise (1999)	0.23	0.07	Men, Age 59, variation in social security replacement rates
12. Bianchi, Gudmunndsson, and Zoega (2001)	0.42	0.07	Iceland 1987 zero tax year
13. Card and Hyslop (2005)	0.38	0.03	Single Mothers, Canadian Self Sufficiency Project
14. Brown (2009)	0.18	0.01	Teachers Near Retirement, California Pension System Cutoffs
15. Manoli and Weber (2011)	0.25	0.01	Workers Aged 55-70, Austria severance pay discontinuities
Unweighted Mean	**0.32**		

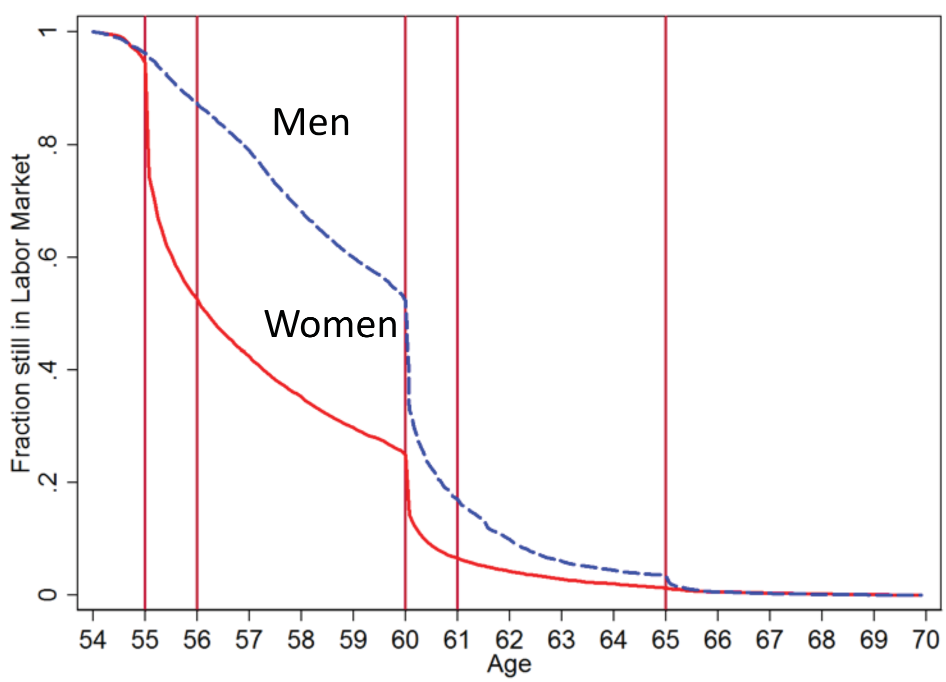
Notes: This table reports elasticities of employment rates with respect to wages, defined as the log change in employment rates divided by the log change in net-of-tax wages. Where possible, we report elasticities from the authors' preferred specification. When estimates are available for multiple populations or for multiple specifications without a stated preference among them, we report an unweighted mean of the relevant elasticities. See Appendix B for details on sources of estimates.

Fig. 1. Payment Amounts based on Tenure at Retirement



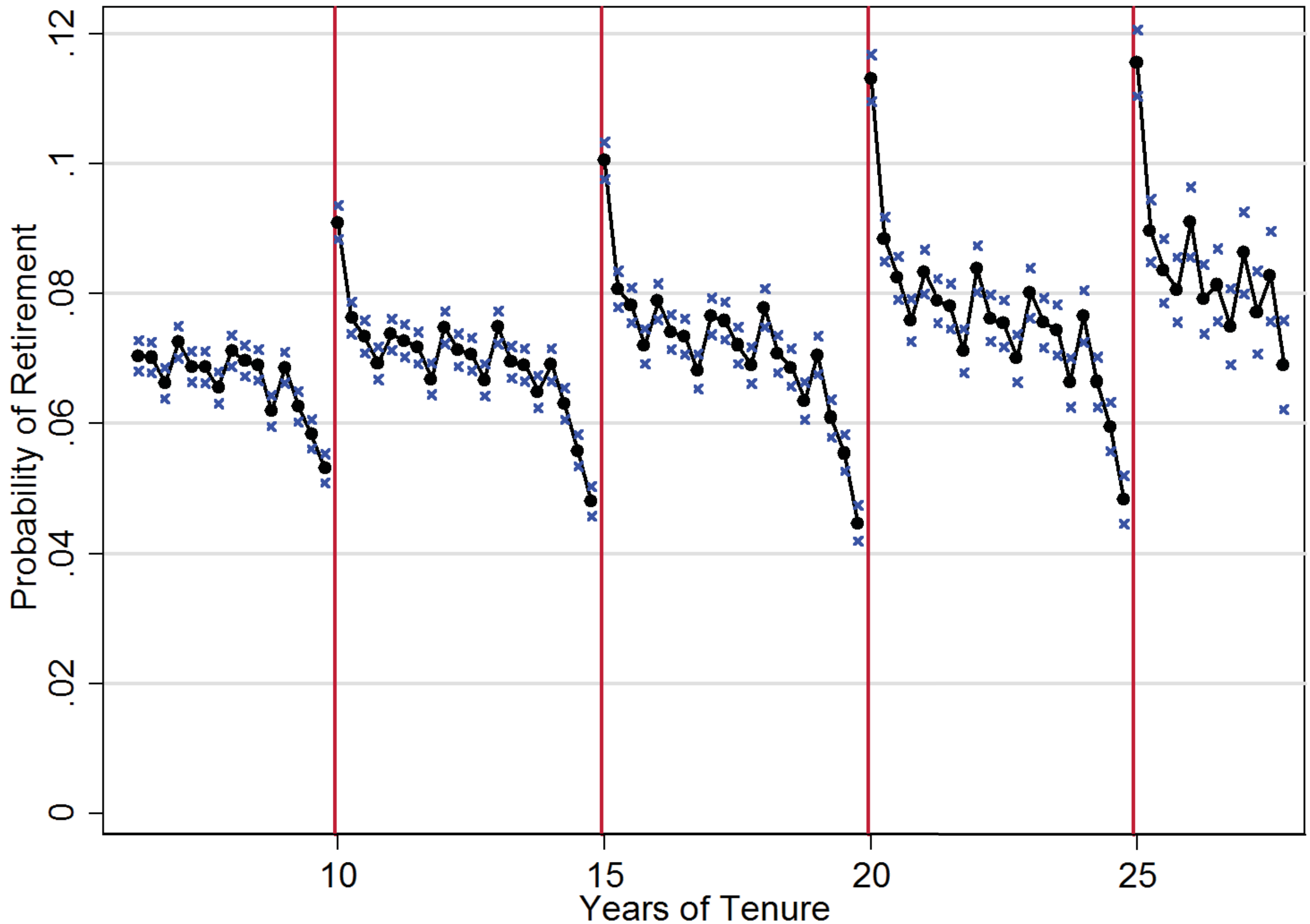
Notes: The employer-provided severance payments are made to private sector employees who have accumulated sufficient years of tenure by the time of their retirement. Tenure is defined as uninterrupted employment time with a given employer and retirement is based on claiming a government-provided pension. The payments must be made within 4 weeks of claiming a pension according to the following schedule.

Fig. 2. Exits from Labor Force into Retirement



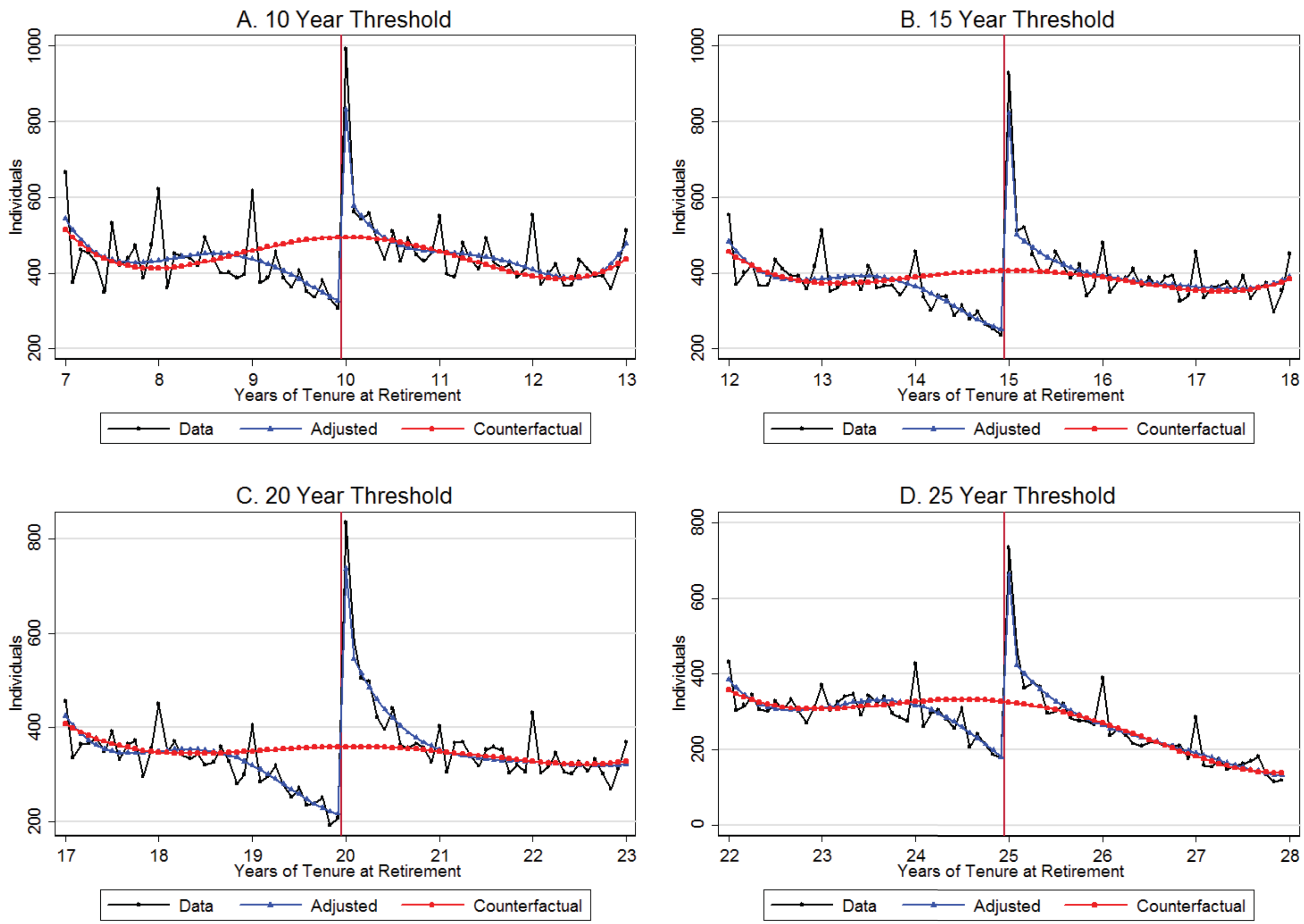
Notes: The survival functions are computed at a monthly frequency using birthdates and last observed job ending dates. The solid red line is the survival function for women; the Early Retirement Age and Normal Retirement Age for women are respectively 55 and 60. The dashed blue line is the survival curve for men; the Early Retirement Age and Normal Retirement Age for men are respectively 60 and 65. Prior to age 60, men can retire through disability pensions.

Fig. 4. Controlling for Covariates



Notes: We regress a quarterly retirement indicator on quarterly tenure dummies and controls for age, gender, calendar years, citizenship, blue collar job status, industry, region, current calendar quarter, job starting month, earnings histories, firm size, health and years of experience. The black circles are the estimated coefficients on the tenure dummies. The blue x's above and below each circle represent +/- 2 standard errors around each point estimate.

Fig. 13. Estimating the Changes in Retirements



Notes: This figure combines plots for the observed retirement frequencies (black squares), the seasonally adjusted retirement frequencies (blue triangles) and the counterfactual retirement frequencies (red circles).

Table 4: Estimation Results

Panel A: Legislated Δ Sev Pay Fraction

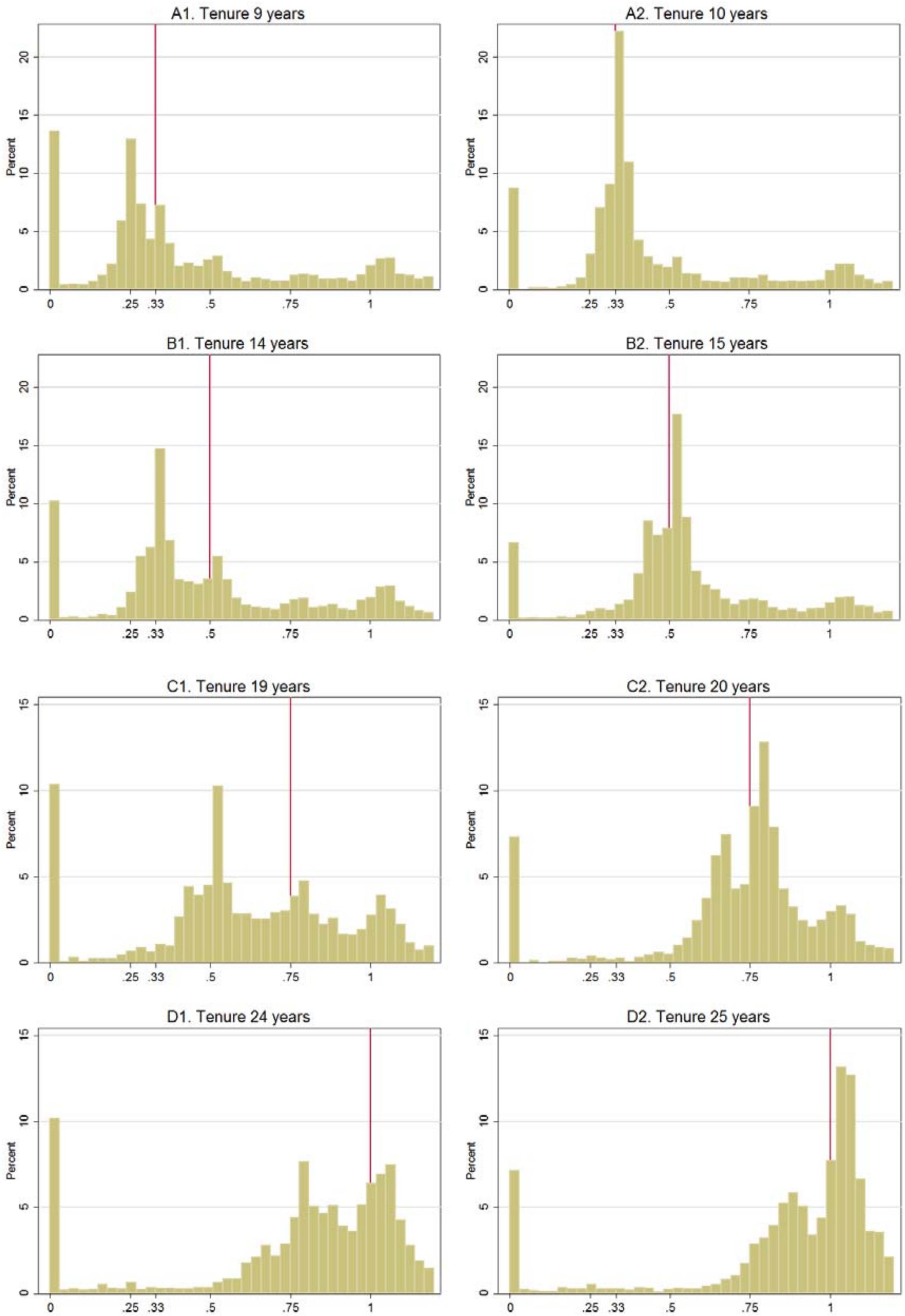
	10 Year Threshold N=21,729	15 Year Threshold N=19,724	20 Year Threshold N=15,588	25 Year Threshold N=18,461	Average
Change in Retirement Probabilities	0.1414	0.2424	0.3777	0.2123	0.2434
	(0.0224)	(0.0273)	(0.0330)	(0.0251)	(0.0146)
Δ Sev Pay Fraction	0.3333	0.1667	0.2500	0.2500	0.2500
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Change in Net-of-Tax Rate	1.5667	0.7833	1.1750	1.1750	1.1750
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Elasticity	0.0902	0.3094	0.3214	0.1807	0.2254
	(0.0143)	(0.0349)	(0.0281)	(0.0214)	(0.0138)

Panel B: Estimated Δ Sev Pay Fraction

	10 Year Threshold N=21,729	15 Year Threshold N=19,724	20 Year Threshold N=15,588	25 Year Threshold N=18,461	Average
Change in Retirement Probabilities	0.1414	0.2424	0.3777	0.2123	0.2434
	(0.0233)	(0.0277)	(0.0350)	(0.0251)	(0.0157)
Δ Sev Pay Fraction	0.0620	0.1056	0.1202	0.0514	0.0848
	(0.0046)	(0.0058)	(0.0049)	(0.0070)	(0.0028)
Change in Net-of-Tax Rate	0.2916	0.4963	0.5651	0.2415	0.3986
	(0.0215)	(0.0275)	(0.0229)	(0.0331)	(0.0131)
Elasticity	0.4848	0.4883	0.6684	0.8790	0.6301
	(0.0892)	(0.0622)	(0.0683)	(0.1668)	(0.0559)

Notes: Numbers in parentheses are bootstrapped standard errors based on 1000 replications. For each tenure threshold, estimation results are based on the sample of observations that have a binding sev pay schedule. Table 2 provides the exact sample definitions. The Change in the Net-of-Tax Rate is mechanically computed using Δ Sev Pay Fraction: Change in Net-of-Tax Rate = (1-0.06)*(Δ Sev Pay Fraction)/(1-0.80).

Fig. 8. Severance Pay Fractions at Different Tenure Levels



Notes: This figure presents the distribution of the severance pay fraction at a given level of tenure at retirement. The severance pay fraction is computed using data from income tax records. Specifically, the fraction is computed as the severance pay in the year of retirement divided by average income in the 3 years prior to retirement. Years of tenure at retirement are computed using job start and exit dates from social security records. The vertical red lines in each plot indicate the legislated severance pay fraction at retirement based on the given level of tenure at retirement.

the degree of insurance to these shocks for the entire sample and for different subgroups of the population.

A. *The Autocovariance of Consumption and Income*

The impact of the deterministic effects $\mathbf{Z}_{it}$ on log income and (imputed) log consumption is removed by separate regressions of these variables on year and year-of-birth dummies, and on a set of observable family characteristics (dummies for education, race, family size, number of children, region, employment status, residence in a large city, outside dependent, and presence of income recipients other than husband and wife). We allow for the effect of most of these characteristics to vary with calendar time. We then work with the residuals of these regressions, labelled $c_{i,t}$ and $y_{i,t}$.[19]

To pave the way to the formal analysis of partial insurance, Table 3 reports unrestricted minimum distance estimates of several moments of the income process for the whole sample: the variance of unexplained income growth, $\operatorname{var}(\Delta y_t)$, the first-order autocovariances, $(\operatorname{cov}(\Delta y_{t+1}, \Delta y_t))$, and the second-order autocovariances, $(\operatorname{cov}(\Delta y_{t+2}, \Delta y_t))$. Estimates are reported for each year. Table 4 repeats the exercise for our new panel data measure of consumption. Finally, Table 5 reports minimum distance estimates of contemporaneous and lagged consumption-income covariances. As noted above, some of the moments are missing because consumption data were not collected in the PSID in the 1987–1988 period.

Looking at Table 3, one can notice the strong increase in the variance of income growth, rising by more than 30 percent by 1985. Also notice the blip in the final year (in 1992 the PSID converted the questionnaire to electronic form and imputations of income were done by machine). The absolute value of the first-order autocovariance also increases until the mid-1980s and then is stable or even declines. Second- and higher-order autocovariances (which, from equation (7), are informative about the presence of serial correlation in the transitory income component) are small and only in few cases statistically significant. At least at face value, this evidence seems to tally quite well with a canonical MA(1) process in growth, as implied by an income process given by the sum of a martingale permanent component

Table 3—The Autocovariance Matrix of Income Growth

Year	$\operatorname{var}(\Delta y_t)$	$\operatorname{cov}(\Delta y_{t+1}, \Delta y_t)$	$\operatorname{cov}(\Delta y_{t+2}, \Delta y_t)$
1980	0.0832	−0.0196	−0.0018
	(0.0089)	(0.0035)	(0.0032)
1981	0.0717	−0.0220	−0.0074
	(0.0075)	(0.0034)	(0.0037)
1982	0.0718	−0.0226	−0.0081
	(0.0051)	(0.0035)	(0.0026)
1983	0.0783	−0.0209	−0.0094
	(0.0066)	(0.0034)	(0.0042)
1984	0.0805	−0.0288	−0.0034
	(0.0055)	(0.0036)	(0.0032)
1985	0.1090	−0.0379	−0.0019
	(0.0180)	(0.0074)	(0.0038)
1986	0.1023	−0.0354	−0.0115
	(0.0077)	(0.0054)	(0.0038)
1987	0.1116	−0.0375	0.0016
	(0.0097)	(0.0051)	(0.0046)
1988	0.0925	−0.0313	−0.0021
	(0.0080)	(0.0042)	(0.0032)
1989	0.0883	−0.0280	−0.0035
	(0.0067)	(0.0059)	(0.0034)
1990	0.0924	−0.0296	−0.0067
	(0.0095)	(0.0049)	(0.0050)
1991	0.0818	−0.0299	NA
	(0.0059)	(0.0040)	
1992	0.1177	NA	NA
	(0.0079)		

[19] To the extent that these regressions remove changes that are unexpected by the individuals, we might expect this to change the relative degree of persistence in the remaining shocks, but not the insurance parameters. For example, by removing the effect of education-time on income and consumption, we are also removing the increase in inequality due to, say, changing education premiums (Attanasio and Davis 1996). If we omit the education variables from our first stage, we find that it makes only a small difference to the estimated insurance parameters (for example, the estimate of ϕ in Table 6 below is 0.71 instead of 0.64). The same qualitative comment applies to the other variables whose effect is removed in the first stage.

TABLE 4—THE AUTOCOVARIANCE MATRIX OF CONSUMPTION GROWTH

Year	$\text{var}(\Delta c_t)$	$\text{cov}(\Delta c_{t+1}, \Delta c_t)$	$\text{cov}(\Delta c_{t+2}, \Delta c_t)$
1980	0.1275	−0.0526	0.0022
	(0.0097)	(0.0076)	(0.0056)
1981	0.1197	−0.0573	0.0025
	(0.0116)	(0.0084)	(0.0043)
1982	0.1322	−0.0641	0.0006
	(0.0110)	(0.0087)	(0.0060)
1983	0.1532	−0.0691	−0.0056
	(0.0159)	(0.0100)	(0.0067)
1984	0.1869	−0.1003	−0.0131
	(0.0173)	(0.0163)	(0.0089)
1985	0.2019	−0.0872	NA
	(0.0244)	(0.0194)	
1986	0.1628	NA	NA
	(0.0184)		
1987	NA	NA	NA
1988	NA	NA	NA
1989	NA	NA	NA
1990	0.1751	−0.0602	−0.0057
	(0.0221)	(0.0062)	(0.0067)
1991	0.1646	−0.0696	NA
	(0.0142)	(0.0100)	
1992	0.1467	NA	NA
	(0.0130)		

and a serially uncorrelated transitory component. Since evidence on second-order autocovariances is mixed, however, in estimation we allow for MA(1) serial correlation in the transitory component $(v_{i,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1})$.[20]

While income moments are informative about shifts in the income distribution (and on the temporary or persistent nature of such shifts), they cannot be used to make conclusive inference about shifts in the consumption distribution. For this purpose, one needs to complement the analysis of income moments with that of consumption moments and of the joint income-consumption moments. This is done in Tables 4 and 5. Table 4 shows that the variance of imputed consumption growth also increases quite strongly in the early 1980s, peaks in 1985, and then it is essentially flat afterward. Note the high value of the level of the variance, which is clearly the result of our imputation procedure. The variance of consumption growth captures in fact the genuine association with shocks to income, but also the contribution of slope heterogeneity and measurement error.[21] The absolute value of the first-order autocovariance of consumption growth should be a good estimate of the variance of the imputation error. This is in fact quite high. Second-order and higher consumption growth autocovariances are mostly statistically insignificant and economically small.

Table 5 examines the association, at various lags, of unexplained income and consumption growth. The contemporaneous covariance should be informative about the effect of income shocks on consumption growth if measurement errors in consumption are orthogonal to measurement errors in income. This covariance increases in the early 1980s and then is flat or even declining afterward.

From (9), the covariance between current consumption growth and one-period-ahead income growth $\text{cov}(\Delta c_t, \Delta y_{t+1})$ should reflect the extent of insurance with respect to transitory shocks (i.e., $\text{cov}(\Delta c_t, \Delta y_{t+1}) = 0$ if there is full insurance of transitory shocks). We note that in the pure self-insurance case with infinite horizon and MA(1) transitory component, the impact of transitory shocks on consumption growth is given by the annuity value $r(1 + r - \theta)/(1 + r)^2$. With a small interest rate, this will be indistinguishable from zero, at least statistically. In fact, this covariance is hardly statistically significant and economically close to zero. At the foot of Table 5 we present the p-values for the joint significance tests of the autocovariances $E(\Delta c_t, \Delta y_{t+j})(j \geq 1)$. These p-values also detect advance information. If future income shocks were known to the consumer in earlier periods, then consumption should adjust before the observed shock occurs. This should show up in significant autocovariances between changes in consumption and future

[20] We also estimated the autocovariances of income growth at lags greater than two and find that none of them is statistically significant. These results are available from the authors upon request.

[21] To a first approximation, the variance of consumption growth that is not contaminated by error can be obtained by subtracting twice the (absolute value of) first-order autocovariance $\text{cov}(\Delta c_{t+1}, \Delta c_t)$ from the variance $\text{var}(\Delta c_t)$.

Table 5—The Consumption-Income Growth Covariance Matrix

Year	$\text{cov}(\Delta y_t, \Delta c_t)$	$\text{cov}(\Delta y_{t+1}, \Delta c_t)$	$\text{cov}(\Delta y_t, \Delta c_{t+1})$
1980	0.0040	0.0013	0.0053
	(0.0041)	(0.0039)	(0.0037)
1981	0.0116	−0.0056	−0.0043
	(0.0036)	(0.0032)	(0.0036)
1982	0.0165	−0.0064	−0.0006
	(0.0036)	(0.0031)	(0.0039)
1983	0.0215	−0.0085	−0.0075
	(0.0045)	(0.0049)	(0.0043)
1984	0.0230	−0.0030	−0.0119
	(0.0052)	(0.0043)	(0.0050)
1985	0.0197	−0.0035	−0.0035
	(0.0068)	(0.0047)	(0.0065)
1986	0.0179	−0.0015	NA
	(0.0048)	(0.0052)	
1987	NA	NA	NA
1988	NA	NA	NA
1989	NA	NA	0.0030
			(0.0040)
1990	0.0077	0.0045	−0.0016
	(0.0045)	(0.0065)	(0.0042)
1991	0.0112	0.0011	−0.0071
	(0.0044)	(0.0049)	(0.0042)
1992	0.0082	NA	NA
	(0.0048)		
Test $\text{cov}(\Delta y_{t+1}, \Delta c_t) = 0$ for all t			p-value 25%
Test $\text{cov}(\Delta y_{t+2}, \Delta c_t) = 0$ for all t			p-value 27%
Test $\text{cov}(\Delta y_{t+3}, \Delta c_t) = 0$ for all t			p-value 74%
Test $\text{cov}(\Delta y_{t+4}, \Delta c_t) = 0$ for all t			p-value 68%

incomes. We find no statistical evidence, however, that this is the case.

The covariance between current consumption growth and past income growth $\text{cov}(\Delta c_{t+1}, \Delta y_t)$ plays no role in the PIH model with perfect capital markets, but may be important in alternative models where liquidity constraints are present (a standard excess sensitivity argument; see Marjorie Flavin 1981). The estimates of this covariance in Table 5 are also close to zero.

To sum up, the evidence suggests that a simple permanent-transitory framework for income shocks with time-varying second-order moments in these shocks provides a good representation of the income process for families in the PSID over this period. Overall we find only weak evidence that transitory shocks affect consumption growth. In the sensitivity results reported below, however, we find that there is evidence of significant responsiveness to transitory shocks for low-wealth families and for the low-income poverty sample of the PSID.

B. *Insurance*

Our focus here will be on the variances of the permanent and the transitory shock, σ_ζ^2 and σ_ε^2, on the partial insurance coefficients for the permanent shock (ϕ) and for the transitory shock (ψ), and the way these parameters vary over time, as well as among different groups in the population. Our estimates are based on a generalization of moments (7)–(9). In particular, to account for our imputation procedure, we allow consumption to be measured with error, and we allow the variance of the measurement error in consumption to vary with time. This is to capture the fact that the imputation error is scaled by a time-varying budget elasticity which induces nonstationarity. We also consider an MA(1) process for the transitory error component of income $(v_{i,a,t} = \varepsilon_{i,t} + \theta\varepsilon_{i,t-1})$, and estimate the MA(1) parameter θ. Finally, we allow for i.i.d. unobserved heterogeneity in the individual consumption gradient, and estimate its variance (σ_ξ^2).

We present the results of three specifications: one for the whole sample (the "baseline" specification), one where the parameters are estimated separately by education (college versus no college), and one where parameters are estimated separately by cohort (born 1930s versus born 1940s).[22] We also allow for some time nonstationarity. In particular, in all specifications we let the variances of the permanent and the transitory shock, σ_ζ^2 and σ_ε^2, respectively, vary with calendar time. As for the partial insurance coefficients for the permanent shock (ϕ) and for the transitory shock (ψ), we assume that they take on two different values, before and after 1985. This is consistent with the evidence in Figure 1, which divides the sample period into a period of rapid

[22] Results for the younger cohort (born in the 1950s) and the older cohort (born in the 1920s) are less reliable because these cohorts are not observed for the whole sample period. We thus omit them.

Table 6—Minimum-Distance Partial Insurance and Variance Estimates

		Whole sample	No college	College	Born 1940s	Born 1930s
σ_ζ^2	1979–81	0.0102	0.0067	0.0099	0.0074	0.0057
(Variance perm. shock)		(0.0035)	(0.0037)	(0.0053)	(0.0035)	(0.0072)
	1982	0.0207	0.0154	0.0252	0.0210	0.0166
		(0.0041)	(0.0053)	(0.0060)	(0.0061)	(0.0075)
	1983	0.0301	0.0317	0.0233	0.0184	0.0246
		(0.0057)	(0.0075)	(0.0089)	(0.0058)	(0.0086)
	1984	0.0274	0.0333	0.0176	0.0219	0.0224
		(0.0049)	(0.0074)	(0.0060)	(0.0077)	(0.0102)
	1985	0.0293	0.0287	0.0204	0.0187	0.0333
		(0.0096)	(0.0073)	(0.0151)	(0.0066)	(0.0225)
	1986	0.0222	0.0173	0.0312	0.0222	0.0111
		(0.0060)	(0.0068)	(0.0101)	(0.0077)	(0.0114)
	1987	0.0289	0.0202	0.0354	0.0307	0.0079
		(0.0063)	(0.0073)	(0.0098)	(0.0080)	(0.0111)
	1988	0.0157	0.0117	0.0183	0.0155	0.0007
		(0.0069)	(0.0079)	(0.0110)	(0.0076)	(0.0099)
	1989	0.0185	0.0107	0.0274	0.0176	0.0217
		(0.0059)	(0.0101)	(0.0061)	(0.0082)	(0.0182)
	1990–92	0.0134	0.0092	0.0216	0.0081	0.0063
		(0.0042)	(0.0045)	(0.0065)	(0.0059)	(0.0091)
σ_ε^2	1979	0.0415	0.0465	0.0302	0.0314	0.0342
(Variance trans. shock)		(0.0059)	(0.0096)	(0.0056)	(0.0054)	(0.0070)
	1980	0.0318	0.0330	0.0284	0.0269	0.0306
		(0.0039)	(0.0053)	(0.0059)	(0.0056)	(0.0072)
	1981	0.0372	0.0364	0.0253	0.0319	0.0267
		(0.0035)	(0.0053)	(0.0046)	(0.0058)	(0.0064)
	1982	0.0286	0.0376	0.0214	0.0264	0.0342
		(0.0039)	(0.0063)	(0.0042)	(0.0049)	(0.0078)
	1983	0.0286	0.0372	0.0186	0.0190	0.0284
		(0.0037)	(0.0063)	(0.0037)	(0.0045)	(0.0077)
	1984	0.0351	0.0405	0.0305	0.0223	0.0453
		(0.0039)	(0.0059)	(0.0051)	(0.0047)	(0.0100)
	1985	0.0380	0.0356	0.0496	0.0280	0.0504
		(0.0075)	(0.0056)	(0.0130)	(0.0062)	(0.0115)
	1986	0.0544	0.0474	0.0452	0.0261	0.0672
		(0.0058)	(0.0076)	(0.0085)	(0.0060)	(0.0153)
	1987	0.0480	0.0520	0.0421	0.0440	0.0499
		(0.0054)	(0.0082)	(0.0071)	(0.0093)	(0.0095)
	1988	0.0383	0.0472	0.0343	0.0386	0.0543
		(0.0047)	(0.0074)	(0.0060)	(0.0068)	(0.0148)
	1989	0.0369	0.0539	0.0219	0.0360	0.0493
		(0.0068)	(0.0126)	(0.0051)	(0.0070)	(0.0132)
	1990–92	0.0506	0.0536	0.0345	0.0429	0.0753
		(0.0040)	(0.0062)	(0.0049)	(0.0060)	(0.0127)
θ		0.1132	0.1268	0.1086	0.1324	0.1706
(Serial correl. trans. shock)		(0.0247)	(0.0318)	(0.0341)	(0.0442)	(0.0470)
σ_ξ^2		0.0105	0.0074	0.0141	0.0122	0.0001
(Variance unobs. slope heterog.)		(0.0041)	(0.0079)	(0.0040)	(0.0064)	(0.0090)
ϕ		0.6423	0.9439	0.4194	0.7928	0.6889
(Partial insurance perm. shock)		(0.0945)	(0.1783)	(0.0924)	(0.1848)	(0.2393)
ψ		0.0533	0.0768	0.0273	0.0675	−0.0381
(Partial insurance trans. shock)		(0.0435)	(0.0602)	(0.0550)	(0.0705)	(0.0737)
p-value test of equal ϕ		23%	99%	8%	81%	18%
p-value test of equal ψ		75%	33%	29%	76%	4%

Notes: This table reports DWMD results of the parameters of interest. We also estimate time-varying variances of measurement error in consumption (results not reported for brevity). See the main text for details. Standard errors in parentheses.