Lecture 6: Topics in Intertemporal Labor Supply
a. the extensive margin
b. structural 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

structural labor supply:

Pierre Olivier Gourinchas and Jonathan Parker. (2002) "Consumption Over the Life Cycle" *Econometrica* 70 (No. 1):47-89.

Michael Keane and Kenneth Wolpin. "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Calro Evidence." *Review of Economics and Statistics 76 (November 1994): 648-672.*

Michael Keane, Petra Todd and Kenneth Wolpin. "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications." Chapter 4 in Handbook of Labor Economics Volume 4a.

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.

Structural Methods

The idea of fully structural modelling is to estimate the parameters of the utility function that drives choices within and between period. Some advantages of this approach:

1) the model can be solved for the value of the marginal utility of wealth for an agent in a given period, conditional on the state variables he or she sees at that point. This makes it possible to assess the wealth effects of wage changes, and the net effect (via intertemporal substitution and wealth effects) on labor supply

2) the model can be used to assess out of sample policy changes, like a revision in social security, on outcomes at all stages of the lifecycle

There are also some costs:

3) because of computational complexity many simplifications have to be made.

4) it is often very hard to understand where identification is coming from - in most cases parameters are identified by a combination of functional form assumptions and general features of the data. There is rarely local identification based on specific design features, as occurs in IV or RD approaches to estimation of simpler 'reduced form' models

A basic example.

We will discuss a simple dynamic labor supply model that illustrates the idea of interpolation of the value function (or, actually the derivative of the value function) using a regression approximation. To keep things very simple, we will assume that wages take on only a limited set of values (say $w_1, w_2 ... w_J$) and $\pi_{ij} = P(w_t = w_i | w_{t-1} = w_j)$ are known. There will be two state variables: the wage, and assets. The value function at time t will be denoted $V_t(A_t, w_t)$. When the wage takes on only discrete values this is just a set of J functions $V_t(A_t, w_j)$. What is relevant for dynamic consumption and hours choices are the derivatives $\partial V_t(A, w_j)/\partial A = \lambda_t(A, w_j)$. The solution method will involve working backward from the retirement period, and at each period solving for the optimal choices of consumption and hours in that period, as a function of the wage in that period, assets, and the approximations to $\partial V_{t+1}(A, w_j)/\partial A$. With these in hand we can then compute $\partial V_t(A, w_j)/\partial A$ at each of a finite set of values for A. We will then fit a regression model to these points to get an approximating model for $\partial V_t(A, w_j)/\partial A$ at every level of A. We then continue working backward to obtain the optimal consumption and hours functions in each period for each wage and level of assets,

$$c_t^*(A_t, w_t)$$
$$h_t^*(A_t, w_t).$$

In applications these functions can be used to compute a likelihood for the observed data for a sample of people who are observed at various points in time, or to compute hours and consumption profiles that are matched to observed profiles. We defer a discussion of how to use the estimated optimal response functions till the end of the lecture.

Let's assume the within period utility function is separable:

$$U(c,h)=u(c)-d(h).$$

with $d(0)=0$. Let's also assume that agents work until an exogenous age R, then retire. At that point the agent becomes eligible for a pension p. In addition to the pension amount, an agent with (beginning-of-period) wealth A_R buys an annuity and receives a per-period payment of rA_R for the rest of his/her life. For purposes of modeling labor supply at earlier ages we can therefore consider the value function for period R :

$$V_R(A_R)=\sum_{j=0}^{\infty}\frac{U(p+rA_R,0)}{(1+r)^j}=\frac{1}{r}u(p+rA_R)$$

where $U(c,h)$ is the within-period utility function, and I have simplified things by assuming that the agents' discount rate and the annuity price are equivalent (with separable preferences this means that the agent wants to set consumption constant for all remaining periods). A similar setup is used by Gourinchas and Parker (2002). Note that the function $V_R(A_R)$ inherits properties from $u(.)$, so if u depends on some parameter τ then the same parameter shifts V_R.

Now let's go back to period $R-1$. In this period the agent faces a wage w_{R-1}, and has assets A_{R-1}. The value function for this period is

$$V_{R-1}(A_{R-1},w_{R-1})=\max_{c_{R-1},h_{R-1}}u(c_{R-1})-d(h_{R-1})+\frac{1}{1+r}[\frac{1}{r}u(p+r(1+r)(A_{R-1}+w_{R-1}h_{R-1}-c_{R-1})].$$

Note that there is no uncertainty left once we get to $R-1$. So we can solve for the optimal choice in this period very easily, to get a starting value function for our backward recursion.

The f.o.c.'s for period $R-1$ are:

$$\begin{aligned}u'(c_{R-1}) &= \lambda_{R-1}=u'(p+r(1+r)(A_{R-1}+w_{R-1}h_{R-1}-c_{R-1}))\\ d'(h_{R-1}) &= \lambda_{R-1}w_{R-1}.\end{aligned}$$

Now lets assume

$$\begin{aligned}d(h) &= \frac{1}{1+1/\eta}h^{1+1/\eta}\\ u(c) &= \log c\end{aligned}$$

so the f.o.c. for hours implies:

$$h_{R-1}=w_{R-1}^{\eta}\ c_{R-1}^{-\eta},$$

which means optimal earnings in period $R-1$ are

$$w_{R-1}h_{R-1} = w_{R-1}^{1+\eta}\ c_{R-1}^{-\eta}$$

Now all we have to do is find an optimal choice for c_{R-1}. Equating marginal utility of consumption in period $R-1$ and R means that the levels of consumption are equal, so we are looking for a leve of c that satisfies:

$$\begin{aligned} c &= p + r(1+r)(A_{R-1} + w_{R-1}^{1+\eta}\ c^{-\eta} - c) \\ \Rightarrow & \ c = \frac{r(1+r)}{1+r(1+r)}A_{R-1} + \frac{1}{1+r(1+r)}p + \frac{r(1+r)}{1+r(1+r)}w_{R-1}^{1+\eta}c^{-\eta} \end{aligned}$$

This has to be solved numerically. It has the form

$$c = f(c) = k + \gamma c^{-\eta}$$

and notice that k is pretty big and γ is small. Its not hard to solve this by iterative methods.[1] With this we have now obtained numerically

$$c_{R-1}^{*}(A_{R-1}, w_{R-1})$$

(this also depends on η, p, r). We can then obtain $h_{R-1}^{*}(A_{R-1}, w_{R-1})$.

Now notice that

$$\partial V_{R-1}(A, w_{R-1})/\partial A = \lambda_{R-1}^{*}(A_{R-1}, w_{R-1}) = \frac{1}{c_{R-1}^{*}(A_{R-1}, w_{R-1})}.$$

This is the function we are going to need to take expectations over in solving for optimal choices at period $R-2$. In particular, if in period $R-2$ the wage is $w_{R-2} = w_i$ then we are going to need to calculate

$$E_{R-2}[\partial V_{R-1}(A, w_i)/\partial A] = \sum_j \frac{1}{c_{R-1}^{*}(A, w_j)}\pi_{ji},$$

treating A as an endogenous variable that depends on $c_{R-2}, w_{R-2}, h_{R-2}$, and A_{R-2}.

Our method is as follows. First, using the procedure above, we calculate $c_{R-1}^{*}(A, w_j)$ for a grid of values of A and each possible value of w_j. In a test program, I measured all monetary units in $1000's$ and assumed that the possible values for A are $1, 2...1,000$ (i.e., up to a million). I assumed that w takes on

[1] I used this method: start with the initial guess $c_1 = k$. Now $f(c) = f(c_1) + (c - c_1)f'(c_1)$, so setting $c = f(c)$ gives a new guess

$$c_2 = \frac{f(c_1) - c_1 f'(c_1)}{1 - f'(c_1)}.$$

This converges in 3-4 iterations.

values of $10, 20....100$ (i.e., 10,000, 20,000... 100,000), and that $p = 20$ (i.e., 20,000). Then I formed a simple $n^{th} - order$ polynomial approximation:

$$\frac{1}{c^*_{R-1}(A, w_j)} = b_{0j} + b_{1j}A + b_{2j}A^2 + ...b_{nj}A^n$$

For my test program I found that $n = 4$ gets an extremely good fit. Now notice that once we have these coefficients, the expected derivative of the $R-1$ value function is:

$$\begin{aligned} E_{R-2}[\partial V_{R-1}(A, w_i)/\partial A] &= \sum_j (b_{0j} + b_{1j}A + b_{2j}A^2 + ...b_{nj}A^n)\pi_{ji} \\ &= \sum_j b_{0j}\pi_{ji} + \sum_j b_{1j}\pi_{ji}A + ... + \sum_j b_{nj}\pi_{ji}A^n \\ &= b^i_0 + b^i_1 A + b^i_2 A^2 + ... + b^i_n A^n \end{aligned}$$

where the coefficients $b^i_0, b^i_1...b^i_n$ depend on the wage in $R-2$ via the weights π_{ji}. Notice the benefit of having a discrete first-order process for wages: given the J approximating polynomials, all we have to do to form the expectation for a given wage in $R-2$ is weight the approximating polynomials by the appropriate transition probabilities.

Now we are ready to solve the optimal choices for c and h in $R-2$. Specifically, the Bellman equation is:

$$V_{R-2}(A_{R-2}, w_{R-2}) = \max_{c_{R-2}, h_{R-2}} u(c_{R-2}) - d(h_{R-2}) + \frac{1}{1+r} E_{R-2}[V_{R-1}(A_{R-1}, w_{R-1}|w_{R-2})].$$

And the f.o.c. are:

$$\begin{aligned} u'(c_{R-2}) &= \lambda_{R-2} = E_{R-2}[\partial V_{R-1}(A_{R-1}, w_{R-1}|w_{R-2})/\partial A_{R-1}] \\ d'(h_{R-2}) &= \lambda_{R-2} w_{R-2} \\ &\Rightarrow h_{R-2} = w^{\eta}_{R-2}\ c^{-\eta}_{R-2} \\ &\Rightarrow w_{R-2}h_{R-2} = w^{1+\eta}_{R-2}\ c^{-\eta}_{R-2} \end{aligned}$$

So we need to solve

$$\frac{1}{c_{R-2}} = b^i_0 + b^i_1 A + b^i_2 A^2 + ... + b^i_n A^n$$

where

$$A = (1+r)(A_{R-2} + w^{1+\eta}_{R-2}\ c^{-\eta}_{R-2} - c_{R-2}).$$

Thus for each value of A_{R-2} and each possible value of the wage w_i we need to solve the root of the function $g(c; A_{R-2}, w_i)$, where:

$$g(c; A_{R-2}, w_i) = \frac{1}{c} - \sum_k b^i_k((1+r)(A_{R-2} + w^{1+\eta}_i c^{-\eta} - c)^k = 0.$$

Again, a numerical solution is needed.[2] The solution is

$$c^*_{R-2}(A_{R-2}, w_{R-2})$$

(which also depends on η, p, r). We can then get $h^*_{R-2}(A_{R-2}, w_{R-2})$.

Finally, going backward one step we will need to evaluate

$$E_{R-3}[\partial V_{R-2}(A_{R-2}, w_{R-2})/\partial A_{R-2}|w_{R-3} = w_i] = \sum_j \frac{1}{c^*_{R-2}(A_{R-2}, w_j)} \pi_{ji}.$$

Thus we can proceed backwards, by estimating the approximating polynomial functions and repeating the previous steps.

Some comments:

1) Notice in this algorithm, everything is summarized by the approximating polynomial coefficients for $\lambda^*_t(A_t, w_j)$. For example, if we use a fourth order polynomial, and have 10 possible wage values, the relevant information for period t, given the transition matrix elements π_{ij}, and the parameters (η, p, r), is contained in 50 numbers. The algorithm proceeds by getting the numbers sequentially from $R-2$ back to some earliest possible period ($e.g., R-40$).

2) We could introduce tastes in one of several ways. One way is to allow the marginal utilities of consumption or leisure to change with age in some way, e.g.,

$$d_t(h_t) = f(t) \frac{1}{1+1/\eta} h_t^{1+1/\eta}$$

where $f(t)$ is a simple function like $f(t) = \exp(\upsilon t)$. For a given value of υ it is possible to solve for the optimal consumption and hours functions in each period, and then search for a best fitting choice. Another way is to assume there are discrete types $\upsilon \in \{\upsilon_1, \upsilon_2, ...\upsilon_K\}$, and assume

$$d_k(h) = \exp(\upsilon_k) \frac{1}{1+1/\eta} h^{1+1/\eta}$$

Then we have to solve the problem for each type,and think of how to map the behavior we see into an average across the types.

3) How do we get the π_{ij} elements?

Suppose that we want to approximate a first order serially correlated continuous process by a 1st order Markov process. G. Tauchen (1986 Economics Letters) described a simple algorithm. For example, suppose we want to approximate an AR-1 wage process:

$$w_t = a + \rho w_{t-1} + \epsilon_t$$

[2]The standard method is Newton-Raphson. Recall that if you are trying to find a c such that $g(c) = 0$ you can normally start with an initial guess c_1 and iterate: $c_j = c_{j-1} - g(c_{j-1})/g_c(c_{j-1})$. In the case where we are approximating the marginal utility of income with polynomials, the analytical derviative is easy.

where $\epsilon_t \sim N(0, \sigma^2)$. Note that for this process $E[w_t] = \mu_w = a/(1-\rho)$, and $var[w_t] = \sigma_w^2 = \sigma^2/(1-\rho^2)$. To approximate this with a discrete 1st order markov model with N points of support, first find $N-1$ cut points k_j $(j = 1, ..N-1)$ such that

$$\Phi[\frac{k_{j+1} - \mu_w}{\sigma_w}] - \Phi[\frac{k_j - \mu_w}{\sigma_w}] = \frac{1}{N}$$

with $k_0 = -\infty$, and $k_N = \infty$. (This defines the boundaries so that the probability a draw from $N(\mu_w, \sigma_w^2)$ falls in each bin is 1/N). Next, find the mean value of a $N(\mu_w, \sigma_w^2)$ within each bin. These values will be the points of support for the discrete process. If $\rho = 0$ we can stop. Otherwise, the last step is to define transition probabilities π_{ij} such that

$$\pi_{ij} = P(k_i < w_t < k_{i+1} | k_j < w_{t-1} < k_{j+1})$$

assuming that

$$\begin{pmatrix} w_{t-1} \\ w_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_w \\ \mu_w \end{pmatrix}, \sigma_w^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

This can be computed using the usual formulas (e.g. in Johnson and Kotz) (or using simple simulation methods).

4) How do we use the optimal consumption and hours functions, $c_t^*(A_t, w_t), h_t^*(A_t, w_t)$?

A huge obstacle to micro research on consumption and labor supply is the absence of reliable data on assets. For example, the well known structural study of retirement by Rust and Phelan, "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," Econometrica 65(July 1997), **assumes no savings**, in part because of the low quality of the asset information in their data set. As a result, almost no studies have tried to estimate structural labor supply models that are directly based on observed data on consumption, hours, wages, and assets. One of the few is Imai and Keane, IER 2004, which solves the problem by evaluating the value function at a discrete number of points and interpolating (rather than interpolating the marginal utility of wealth function). Imai and Keane allow for mismeasurement in assets and hours.

Models of Earnings/Wage Dynamics

Some references

John Abowd and David Card (1989). "On the Covariance Structure of Earnings and Hours Changes". Econometrica 57 (2): 411-445.

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.

Bhashkar Mazumder (2005). "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." RESTAT 87(2): 235-255.

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.

Orley Ashenfelter and David Card (1985). "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." RESTAT 67(4): 648-660.

Introduction

As discussed in lecture 4, an important question for interpreting the reaction of hours to wage changes is to what extent wage innovations are expected to persist. Pistaferri assumes that innovations are permanent, i.e., 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

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

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:

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

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 (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 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). 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, $cov(\Delta u_{it}, \Delta u_{is}) \to 0$. 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

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

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).

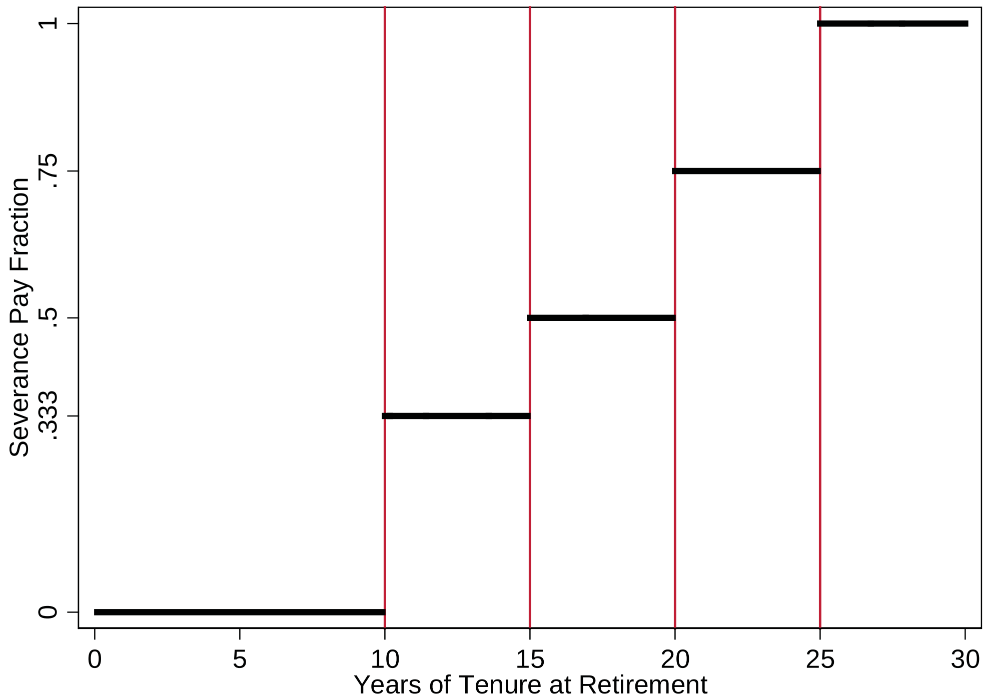
As a practical matter the optimal choice for the weighting matrix can lead to substantial problems in small samples. This was not well understood at the time of Abowd-Card, but was pointed out in the paper by Altonji and Segel. It is generally agreed that when the moments of interest are all (roughly)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

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**		

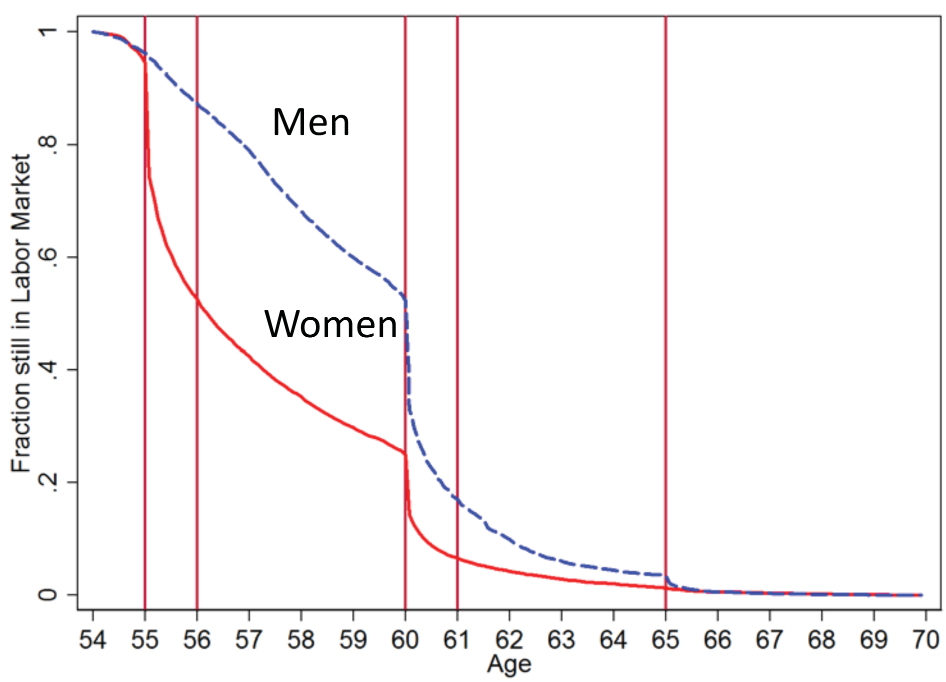
Notes: This table reports elasticities of employment rates with respect to wages, defined as the log change in employment rates divided by the log change in net-of-tax wages. Where possible, we report elasticities from the authors' preferred specification. When estimates are available for multiple populations 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



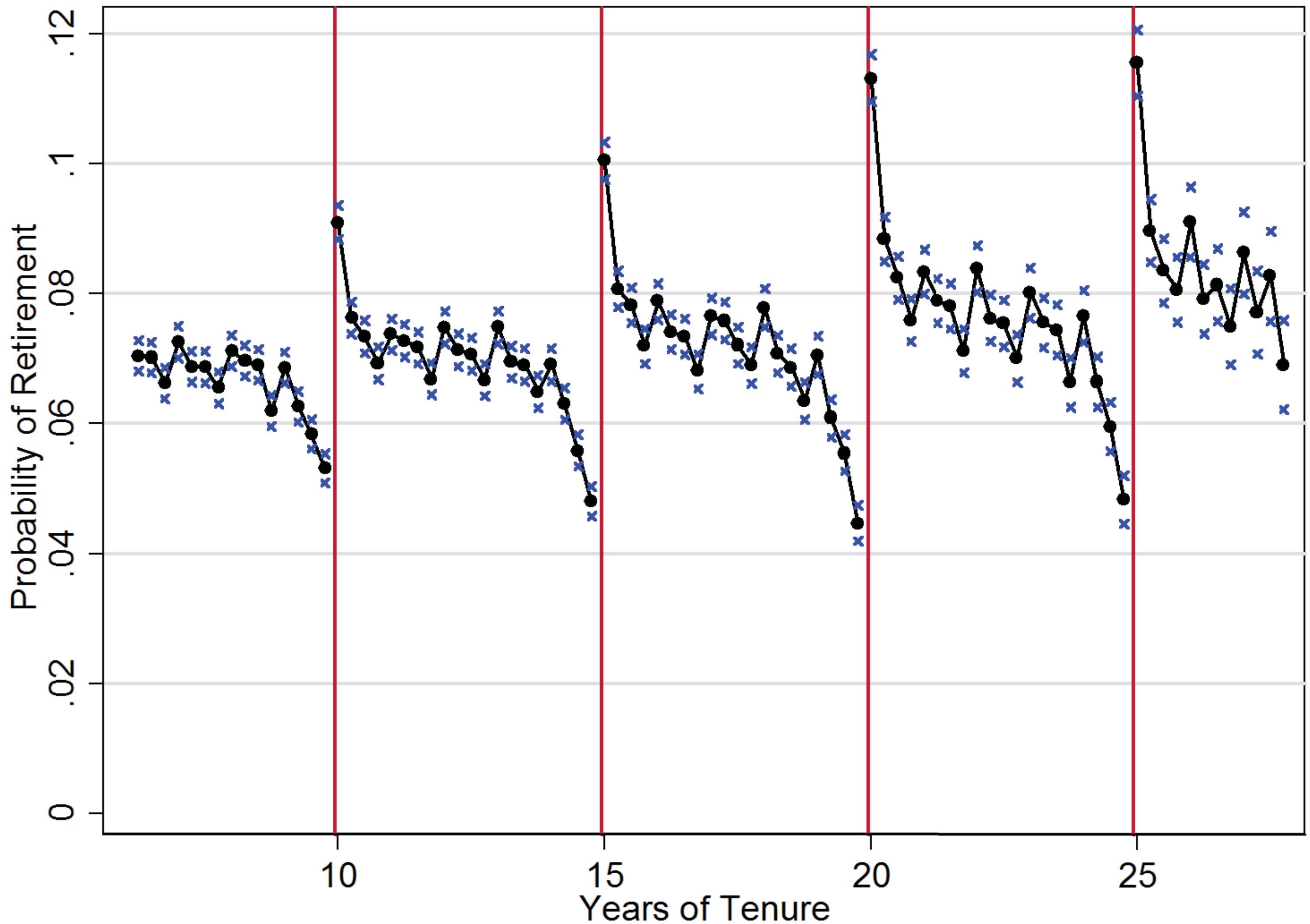
Notes: The employer-provided severance payments are made to private sector employees who have accumulated sufficient years of tenure by the time of their retirement. Tenure is defined as uninterrupted employment time with a given employer and retirement is based on claiming a government-provided pension.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



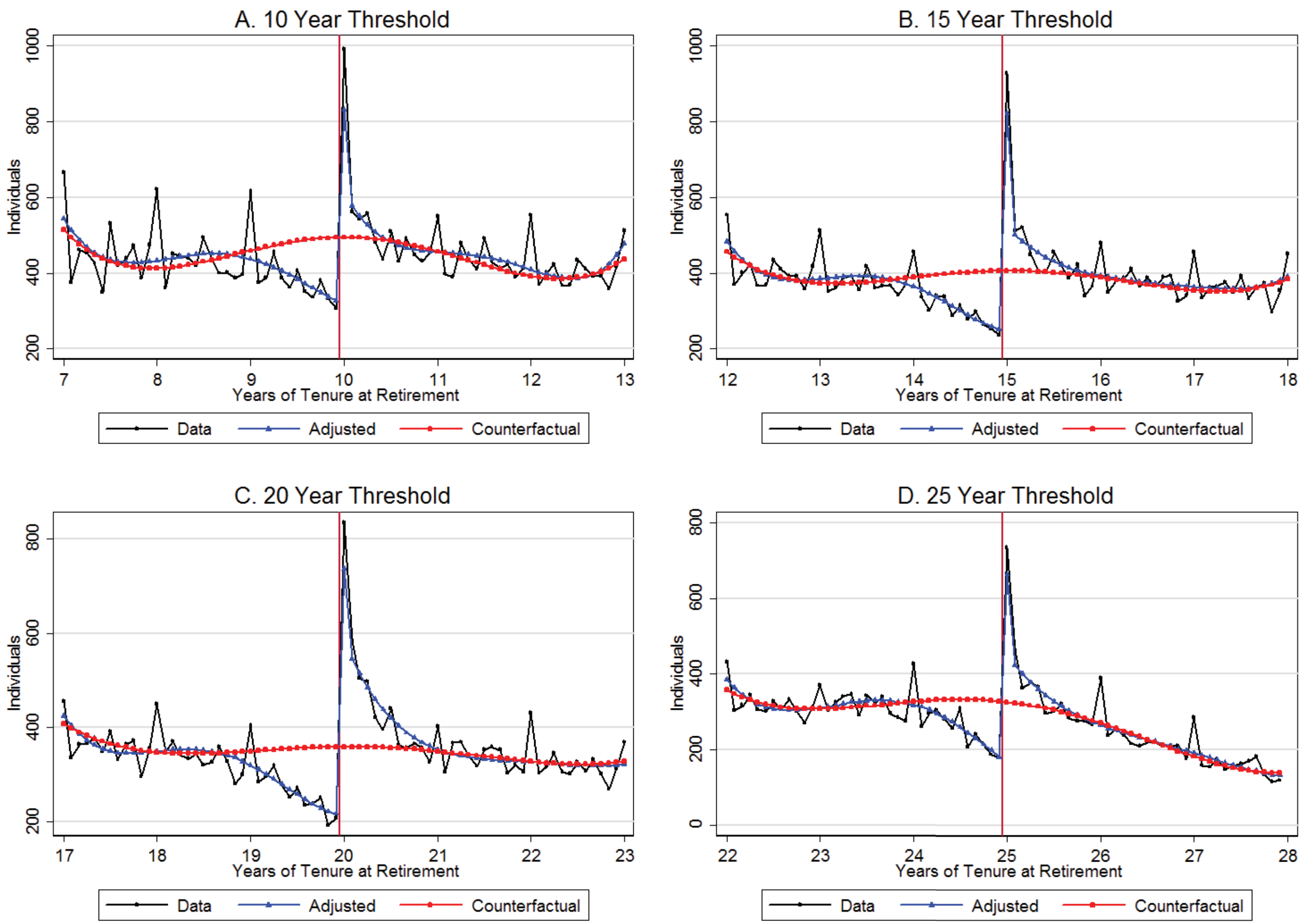
Notes: The survival functions are computed at a monthly frequency using birthdates and last observed job ending dates. The solid red line is the survival function for women; the Early Retirement Age and Normal Retirement Age for women are respectively 55 and 60. The dashed blue line is the survival curve 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

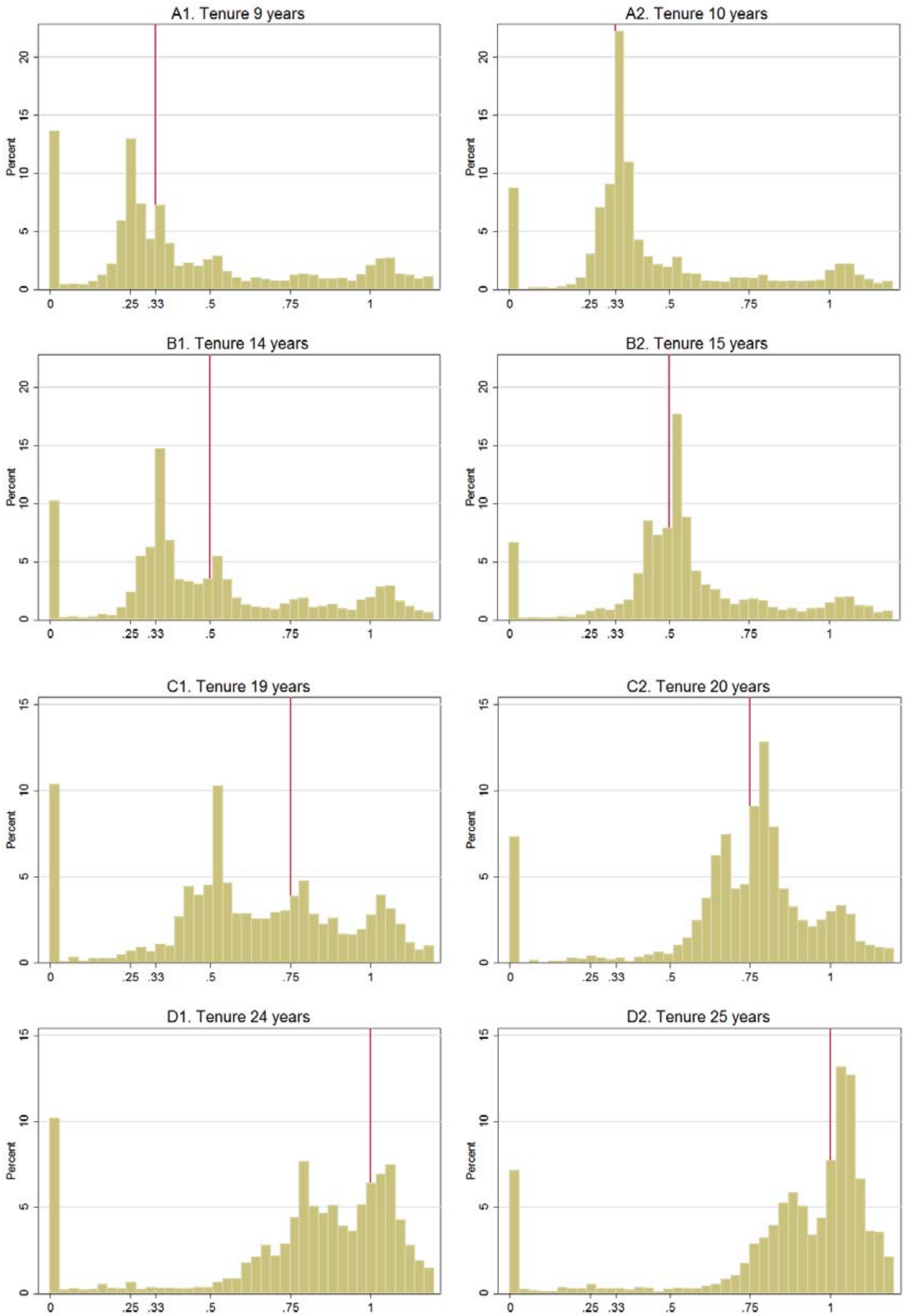
	10 Year Threshold N=21,729	15 Year Threshold N=19,724	20 Year Threshold N=15,588	25 Year Threshold N=18,461	Average
Change in Retirement Probabilities	0.1414	0.2424	0.3777	0.2123	0.2434
	(0.0224)	(0.0273)	(0.0330)	(0.0251)	(0.0146)
Δ Sev Pay Fraction	0.3333	0.1667	0.2500	0.2500	0.2500
	(0.0000)	(0.0000)	(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.