0.1 The micro "wage process"

References

For estimation of micro wage models:

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.

David Card and Thomas Lemieux (1994). "Changing Wage Structure and Black-White Wage Differentials. AER May 1994.

Note: many of the same issues arise in the closely related literature on estimating micro "earnings models". Some useful references in this literature are:

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

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

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 $\omega_i + u_{it} + e_{it}$. (There is a more sophisticated approach which we may discuss briefly in class). 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 (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", $cov(\Delta u_{it},\ \Delta u_{is}) \rightarrow 0$. 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?

0.2 Estimation Method

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. 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.