Online Appendix to: "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks"

Lance J. Lochner
University of Western Ontario

Youngki Shin
University of Western Ontario

May 12, 2014

This Online Appendix provides additional details regarding estimation. We first define notation for a general version of our model with an $ARMA(1,q)$ process for transitory shocks. We then define moments used in estimation and provide details for our minimum distance estimator.

1 General Model for Residuals

Assuming an $ARMA(1,q)$ process for transitory shocks, we can write the model as:

$$
\begin{aligned}
W_{i,t} &= \mu_t(\theta_i) + \kappa_{i,t} + \nu_{i,t} \\
\kappa_{i,t} &= \kappa_{i,t-1} + \eta_{i,t} \\
\nu_{i,t} &= \rho \nu_{i,t-1} + \xi_{i,t} + \sum_{j=1}^{q} \beta_j \xi_{i,t-j}
\end{aligned}
$$

where $E[\mu_t(\theta_i)] = E[\eta_{i,t}] = E[\xi_{i,t}] = 0$ for all t.

Let c_i reflect the cohort (i.e. year of birth) for individual i, so individual i is $a_{i,t} = t - c_i$ years old in year t.

Assuming individuals do not receive any shocks prior to age 20, we write:

$$\kappa_{i,t} = \sum_{j=0}^{a_{i,t}-20} \eta_{i,t-j}$$
$$\nu_{i,t} = \sum_{j=0}^{a_{i,t}-20} \gamma_j \xi_{i,t-j},$$

where

$$\gamma_j = \begin{cases} 1 & \text{for } j = 0 \\ \beta_j + \rho\gamma_{j-1} & \text{for } 1 \leq j \leq q \\ \rho\gamma_{j-1} & \text{for } j > q. \end{cases}$$

Individuals older than 20 years old in the initial sample period (1970) are left-censored. We assume the distributions of permanent (η_t) and transitory shocks (ξ_t) prior to 1970 are identical to those in 1970.

Unknown parameters in the model to be estimated include ρ (in models with an autoregressive process), $\{\beta_j\}_{j=1}^{q}$, $\{\mu_t(\cdot)\}_{t=1970}^{2008}$, and various moments of θ and $\{\eta_t, \xi_t\}_{t=1970}^{2008}$. In general, we are only able to identify and estimate distributional parameters for (η_t, ξ_t) up through 2002 for reasons discussed in Section 2 of the paper and due to biennial observations. Since we do not have any earnings data in odd-numbered years after 1996, we assume moments for η_t and ξ_t in those missing years are the midpoint between adjacent years.

2 Calculating Model Moments

Define 3 age groups $\{30, 31, \ldots, 38, 39\}$, $\{40, 41, \ldots, 48, 49\}$, and $\{50, 51, \ldots, 58, 59\}$ indexed by A. If $\mu_t(\theta_i)$, $\kappa_{i,t}$, and $\nu_{i,t}$ are all independent of each other, all moments of $W_{i,t}$ can be written as the sum of the moments for each component.[1] Second moments for all period t, age group A, and lag length k are

$$E[W_{i,t}W_{i,t-k}|a_{i,t} \in A] = E[\mu_t(\theta_i)\mu_{t-k}(\theta_i)] + E[\kappa_{i,t}\kappa_{i,t-k}|a_{i,t} \in A]. + E[\nu_{i,t}\nu_{i,t-k}|a_{i,t} \in A].$$

The maximum lag length is reached when $t - k = 1970$ or the youngest cohort in each age group is 30 years old in year $t - k$. Similarly, third moments for all period t, age group A,

[1] Later, we consider the case where the distribution of $\eta_{i,t}$ depends on θ_i. In this case, there are interaction terms between $\mu_t(\theta_i)$ and $\kappa_{i,t}$.

and lag lengths k and l are

$$\begin{aligned} E[W_{i,t}W_{i,t-k}W_{i,t-k-l}|a_{i,t} \in A] &= E[\mu_t(\theta_i)\mu_{t-k}(\theta_i)\mu_{t-k-l}(\theta_i)] + E[\kappa_{i,t}\kappa_{i,t-k}\kappa_{i,t-k-l}|a_{i,t} \in A] \\ &\quad +E[\nu_{i,t}\nu_{i,t-k}\nu_{i,t-k-l}|a_{i,t} \in A]. \end{aligned}$$

The maximum lag length is reached when $t-k-l = 1970$ or the youngest cohort in each age group is 30 years old in year $t-k-l$.

Altogether, our residual observations from 1970 to 2008 (annual observations through 1996, biennial thereafter) in the PSID generate 783 second moments and 5,718 third moments for all years, age groups, and lag lengths.

To simplify notation, the individual subscript i will be omitted throughout the rest of this appendix. We next consider each set of moments corresponding to $\mu_t(\cdot)$, κ_t, and η_t.

2.1 Moments of $\mu_t(\theta)$

We approximate $\mu_t(\theta)$ by a polynomial of degree p:

$$\mu_t(\theta) = \sum_{d=0}^{p} m_{d,t}\theta^d.$$

Empirically, we use $p = 1$ or 3. The normalization $E[\mu_t(\theta)] = 0$ implies the following restriction: $m_{0,t} = -\sum_{d=1}^{p} m_{d,t}E[\theta^d]$.

When $\mu_t(\theta)$ is linear in θ, we only need to know the variance of θ in order to calculate the second moments of $\mu_t(\theta)$. However, when $p > 1$, we need to know higher moments of θ as well. In this case, we assume θ is a mixture of two normal random variables with $E[\theta] = 0$.

Second moments of $\mu_t(\theta)$ are

$$\begin{aligned} E\Big[\mu_t(\theta)\mu_{t-k}(\theta)\Big] &= E\left[\left(\sum_{d=0}^{p} m_{d,t}\theta^d\right)\left(\sum_{d'=0}^{p} m_{d',t-k}\theta^{d'}\right)\right] \\ &= \sum_{d=0}^{p}\sum_{d'=0}^{p} m_{d,t}m_{d',t-k}E[\theta^{d+d'}], \end{aligned}$$

and third moments are

$$E\Big[\mu_t(\theta)\mu_{t-k}(\theta)\mu_{t-k-l}(\theta)\Big] = \sum_{d=0}^{p}\sum_{d'=0}^{p}\sum_{d''=0}^{p} m_{d,t}m_{d',t-k}m_{d'',t-k-l}E[\theta^{d+d'+d''}].$$

2.2 Permanent Component

Permanent shocks are accumulated with age, so we calculate the moments of the permanent component for each age and take average for each age group in each year. Consider the second moments:

$$E\big[\kappa_t \kappa_{t-k} | a_t \in A\big] = \sum_{a_t \in A} \text{Prob}(a_t | t, A) E\big[\kappa_t \kappa_{t-k} | a_t\big],$$

where the sample age distribution (within each age group) in each year is used and $E\big[\kappa_t \kappa_{t-k}|a_t\big] = E\big[\kappa_{t-k}^2 | a_t\big] = \sum_{j=0}^{a_t-20-k} \sigma_{\eta_{t-k-j}}^2$. Third moments are calculated analogously.

2.3 Transitory Component

Transitory shocks are also serially correlated, so we have the following second moments:

$$E\big[\nu_t \nu_{t-k} | a_t \in A\big] = \sum_{a_t \in A} \text{Prob}(a_t | t, A) E\big[\nu_t \nu_{t-k} | a_t\big],$$

where the sample age distribution in each year is used, $E\big[\nu_t \nu_{t-k}|a_t\big] = \sum_{j=0}^{a_t-20-k} \gamma_j \gamma_{j+k} \sigma_{\xi_{t-k-j}}^2$ and $\sigma_{\xi_t}^2 = E[\xi_t^2]$. Third moments are calculated analogously.

2.4 Heteroskedasticity in Permanent Shocks

We relax the assumption that η_t and θ are independent by allowing the following form of heteroskedasticity:

$$\eta_t = \sigma_t(\theta)\zeta_t,$$

where $E[\zeta_t] = 0$, $E[\zeta_t^2] = 1$, $\sigma_t(\theta) > 0$, and $\sigma_t(\theta)$ is approximated by a polynomial of degree D_σ:

$$\sigma_t(\theta) = \sum_{d=0}^{D_\sigma} \delta_{d,t} \theta^d.$$

Empirically, we consider $D_\sigma = 1$.

Second moments for permanent shocks are now given by

$$E\big[\kappa_t(\theta)\kappa_{t-k}(\theta)|a_t\big] = \sum_{j=0}^{a_t-20-k} E\big[(\sigma_{t-k-j}(\theta))^2\big]$$

where

$$E\big[(\sigma_\tau(\theta))^2\big] = \sum_{d=0}^{D_\sigma} \sum_{d'=0}^{D_\sigma} \delta_{d,\tau}\delta_{d',\tau} E[\theta^{d+d'}].$$

Third moments for the permanent component are

$$E\Big[\kappa_t(\theta)\kappa_{t-k}(\theta)\kappa_{t-k-l}(\theta)|a_t\Big] = \sum_{j=0}^{a_t-20-k} E\Big[\big(\sigma_{t-k-l-j}(\theta)\big)^3\Big]\sigma^3_{\zeta_{t-k-l-j}},$$

where

$$E\Big[\big(\sigma_\tau(\theta)\big)^3\Big] = \sum_{d=0}^{D_\sigma}\sum_{d'=0}^{D_\sigma}\sum_{d''=0}^{D_\sigma} \delta_{d,\tau}\delta_{d',\tau}\delta_{d'',\tau}E[\theta^{d+d'+d''}].$$

Unlike the case with full independence between θ and κ_t, third moments for residuals include the following three interaction terms:

$$\begin{aligned}
E\Big[\mu_t(\theta)\kappa_{t-k}(\theta)\kappa_{t-k-l}(\theta)|a_t\Big] &= E\left[\mu_t(\theta)\left\{\sum_{j=0}^{a_t-20-k-l}\big(\sigma_{t-k-l-j}(\theta)\big)^2\right\}\right] \\
&= \sum_{j=0}^{a_t-20-k-l} E\Big[\mu_t(\theta)\big(\sigma_{t-k-l-j}(\theta)\big)^2\Big] \\
E\Big[\kappa_t(\theta)\mu_{t-k}(\theta)\kappa_{t-k-l}(\theta)|a_t\Big] &= \sum_{j=0}^{a_t-20-k-l} E\Big[\mu_{t-k}(\theta)\big(\sigma_{t-k-l-j}(\theta)\big)^2\Big] \\
E\Big[\kappa_t(\theta)\kappa_{t-k}(\theta)\mu_{t-k-l}(\theta)|a_t\Big] &= \sum_{j=0}^{a_t-20-k} E\Big[\mu_{t-k-l}(\theta)\big(\sigma_{t-k-j}(\theta)\big)^2\Big],
\end{aligned}$$

where

$$E\Big[\mu_t(\theta)\big(\sigma_\tau(\theta)\big)^2\Big] = \sum_{d=0}^{p}\sum_{d'=0}^{D_\sigma}\sum_{d''=0}^{D_\sigma} m_{d,t}\delta_{d',\tau}\delta_{d'',\tau}E[\theta^{d+d'+d''}].$$

3 Minimum Distance Estimation

There are $i = 1, \dots, N$ individuals with $t = 1, \dots, T$ periods of data. Let $\boldsymbol{W}_i$ be the vector of earnings residuals for individual i and $\boldsymbol{W}$ be the matrix of residuals for all individuals:

$$\boldsymbol{W}_i = \begin{bmatrix} W_{i,1} \\ \vdots \\ W_{i,T} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 & \dots & \boldsymbol{W}_N \end{bmatrix}$$

We estimate the model by choosing the parameters that minimize the distance between the moments implied by the model and the corresponding moments from the data. For each moment $m = 1, \dots, M$, there are N_m non-missing observations. The sample moment is given by

$$\bar{s}_m(\boldsymbol{W}) = \frac{1}{N_m}\sum_{i=1}^{N_m} s_{i,m}(\boldsymbol{W}_i),$$

where $s_{i,m}(\boldsymbol{W}_i)$ returns the m^{th} moment from individual i's earnings residual vector. For example, if the first moment is the variance of $W_{i,1}$, then $s_{i,1}(\boldsymbol{W}_i) = W_{i,1}^2$.

Given a parameter vector $\boldsymbol{\lambda}$, we can calculate M theoretical moments $\boldsymbol{f}(\boldsymbol{\lambda})$. Let $\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda})$ be the distance between the theoretical moments and the sample moments for individual i and $\bar{\boldsymbol{g}}(\boldsymbol{W}, \boldsymbol{\lambda})$ be the average of $\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda})$ across individuals:

$$\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda}) = \begin{bmatrix} s_{i,1}(\boldsymbol{W}_i) - f_1(\boldsymbol{\lambda}) \\ \vdots \\ s_{i,M}(\boldsymbol{W}_i) - f_M(\boldsymbol{\lambda}) \end{bmatrix}, \quad \bar{\boldsymbol{g}}(\boldsymbol{W}, \boldsymbol{\lambda}) = \begin{bmatrix} \bar{s}_1(\boldsymbol{W}) - f_1(\boldsymbol{\lambda}) \\ \vdots \\ \bar{s}_M(\boldsymbol{W}) - f_M(\boldsymbol{\lambda}) \end{bmatrix}.$$

Then, for the true parameter vector $\boldsymbol{\lambda}_0$, the following moment condition holds:

$$E\big[\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda}_0)\big] = 0.$$

We use the minimum distance estimator $\hat{\boldsymbol{\lambda}}$ that solves

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \Big\{ N \bar{\boldsymbol{g}}(\boldsymbol{W}, \boldsymbol{\lambda})' \, \boldsymbol{I_N} \, \bar{\boldsymbol{g}}(\boldsymbol{W}, \boldsymbol{\lambda}) \Big\},$$

where $\boldsymbol{I_N} = \operatorname{diag}\left(\frac{N_1}{N}, \ldots, \frac{N_M}{N}\right)$. Thus, moments are weighted by the number of observations used to compute that moment.

3.1 Calculating Standard Errors

Under standard regularity conditions,[2]

$$\begin{aligned} \hat{\boldsymbol{\lambda}} &\xrightarrow{p} \boldsymbol{\lambda}_0 \\ \sqrt{N}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0) &\xrightarrow{d} N\Big(0, \operatorname{Avar}(\hat{\boldsymbol{\lambda}})\Big), \end{aligned}$$

where

$$\begin{aligned} \operatorname{Avar}(\hat{\boldsymbol{\lambda}}) =& (\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{\Omega}\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{G})^{-1} \\ \boldsymbol{G} =& E\left[\frac{\partial \boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}'} \Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0} \right] \\ \boldsymbol{\Omega} =& E\left[\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda}_0)\boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda}_0)'\right]. \end{aligned}$$

A consistent estimator of $\operatorname{Avar}(\hat{\boldsymbol{\lambda}})$ is $\widehat{\operatorname{Avar}}(\hat{\boldsymbol{\lambda}}) = (\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}$, where

$$\hat{\boldsymbol{G}} = \frac{\partial \boldsymbol{g}(\boldsymbol{W}_i, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}'}\Big|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}} = \frac{\partial \boldsymbol{f}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}'}\Big|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}} \quad \text{and} \quad \hat{\boldsymbol{\Omega}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{g}(\boldsymbol{W}_i, \hat{\boldsymbol{\lambda}})\boldsymbol{g}(\boldsymbol{W}_i, \hat{\boldsymbol{\lambda}})'.$$

[2] In estimating the asymptotic variance for $\hat{\boldsymbol{\lambda}}$, we also assume $\boldsymbol{I_N}$ converges in probability to the identity matrix.