Understanding Income Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks

Lance Lochner (University of Western Ontario)
Youngki Shin (University of Western Ontario)

April 2013
There is considerable interest in the evolution of inequality and the returns to ability/skill over time.

Widespread agreement that returns to observed skills (education, experience) have risen since the early 1980s.

Less agreement on role of unobserved skills:

More generally, there is interest in understanding the factors driving the evolution of residual inequality.
Earnings and Weekly Wage Inequality in the US

Source: 1970-2008 PSID
Residuals and Unobserve Ability/Skill

- CPS-based literature interprets all changes in residual inequality as changes in the ‘pricing’ of unobserved skills
  - e.g., Katz & Murphy (1992), Juhn, Murphy & Pierce (1993), Autor, Katz & Kearney (2008)
  - increased residual inequality reflects an increase in the ‘returns’ to unobserved skill
- Coincidence of rising returns to observed skills and increasing variance of residuals has motivated theories of SBTC
- Card & DiNardo (2002) and Lemieux (2006) raise a number of objections to SBTC, arguing that changes in institutional factors and minimum wages are important
- Most recently, Acemoglu & Autor (2011) and Autor & Dorn (2012) argue that mechanization of routine tasks has led to polarization of skill demand
Interquantile Comparisons for Log Earnings Residuals, 1970-2008 PSID

Year

Interquantile Range of Log Annual Earnings Residual

90%−10%
90%−50%
50%−10%

Lochner & Shin

Understanding Income Dynamics
What about Idiosyncratic Shocks?

- CPS-based literature largely ignores parallel research on earnings dynamics using PSID
  - macro: (Guvenen 2007, Heathcote, Perri & Violante 2010, Heathcote, Storesletten & Violante 2010)

- Decomposes variance of log wages/earnings residuals into permanent and transitory shocks over time
- Important for understanding consumption and savings behavior/inequality
- Estimates suggest similar increases in variance of permanent and transitory shocks
- Transitory component unlikely to be related to unobs. skill
- No accounting for changes in pricing of unobs. skills
Our Goal: Incorporating All Three Components

We consider a very general log earnings/wage residual decomposition:

\[
\begin{align*}
W_{it} &= \mu_t(\theta_i) + \kappa_{i,t} + \nu_{i,t} \\
\kappa_{i,t} &= \kappa_{i,t-1} + \eta_{i,t} \\
\nu_{i,t} &= \rho \nu_{i,t-1} + \xi_{i,t} + \beta_1 t \xi_{i,t-1} + \beta_2 t \xi_{i,t-2} + \ldots + \beta_q t \xi_{i,t-q}
\end{align*}
\]

- ‘Unobserved Ability’ literature effectively ignores any changes in distributions of \( \kappa_{i,t} \) and \( \nu_{i,t} \)
- ‘Earnings Dynamics’ literature effectively ignores \( \mu_t(\theta_i) \) or assumes \( \mu_t(\theta_i) \) is time invariant
Earnings Components

\[ W_{it} = \mu_t(\theta_i) + \kappa_{i,t} + \nu_{i,t} \]
\[ \kappa_{i,t} = \kappa_{i,t-1} + \eta_{i,t} \]
\[ \nu_{i,t} = \rho \nu_{i,t-1} + \xi_{i,t} + \beta_{1t} \xi_{i,t-1} + \beta_{2t} \xi_{i,t-2} + \ldots + \beta_{qt} \xi_{i,t-q} \]

- \( \mu_t(\cdot) \) reflects the pricing of unobserved skills \( \theta \), which may change over time due to technological change or institutional factors (e.g. unions, minimum wage)
- \( \eta_t \) reflects permanent idiosyncratic shocks like job displacement, switching employers, disability
- \( \nu_t \) reflects persistent but transitory shocks like temporary illness, family disruption, temporary demand shocks for employers
What We Do – Outline

- We first consider nonparametric identification and estimation
  - show conditions for nonparametric identification
  - requires independence between far away observations, so need $\rho = 0$ but can have persistent $MA(q)$ process for $\nu_t$
  - based on identification strategy, straightforward to estimate $f_{\theta}(\cdot)$ and $\mu_t(\cdot)$ over time to identify role of unobserved skills
- Consider a moment-based approach with polynomial $\mu_t(\cdot)$ functions and $ARMA(1, q)$ process for $\nu_t$
  - discuss necessary conditions for identification
  - provide Minimum Distance estimates assuming $\mu_t(\cdot)$ are linear or cubic polynomials
- Focus on log earnings for men in PSID, 1970-2008
Consider non-parametric identification, beginning with a simple instructive case:

\[ W_{it} = \mu_t(\theta_i) + \varepsilon_{it} \]

where \( \varepsilon_{it} \) are independent over time and \( \mu_t(\cdot) \) are strictly increasing.

Some normalizations:

- \( E(\theta) = E(\varepsilon_t) = 0 \)
- \( \mu_1(\theta) = \theta \)

Problem is very similar to that of measurement error literature.
Assumption 1

The following conditions hold for $T = 3$:

(i) The joint density of $\theta$, $W_1$, $W_2$, and $W_3$ is bounded and continuous, and so are all their marginal and conditional densities.

(ii) $\theta \perp \perp \varepsilon_t$ for all $t$ and $\varepsilon_t \perp \perp \varepsilon_s$ for $t \neq s$.

(iii) $f_{W_1|W_2}(w_1|w_2)$ and $f_{\theta|W_1}(\theta|w_1)$ form a bounded complete family of distributions indexed by $W_2$ and $W_1$, respectively.

(iv) The functions $\mu_t(\cdot)$ are strictly monotone.
Lemma 1: Identification in the Simple Case

Lemma 1

Under Assumption 1, $f_\theta(\cdot)$, $f_\varepsilon_t(\cdot)$, and $\mu_t(\cdot)$ are identified $\forall t$.

Proof:

- Thm 1 of Hu and Schennach (2008) gives identification of $f_{W_1|\theta}(\cdot|\cdot)$, $f_{W_2|\theta}(\cdot|\cdot)$, and $f_{W_3,\theta}(\cdot,\cdot)$ from $f_{W_1,W_2,W_3}(\cdot,\cdot,\cdot)$
- $f_\theta(\cdot)$ can be recovered from $f_{W_3,\theta}(\cdot,\cdot)$ by integrating out $W_3$ (Cunha, Heckman and Schennach 2010)
- Identify $\mu_2(\cdot)$ and $\mu_3(\cdot)$ from $E[W_t|\theta] = \mu_t(\theta)$ given $f_{W_t|\theta}(\cdot|\cdot)$
- $f_\varepsilon_t(\cdot)$ is identified from $f_{\varepsilon_t}(\varepsilon) = f_{W_t|\theta}(\mu_t(\theta) + \varepsilon)$, since $\mu_t(\cdot)$ and $f_{W_t|\theta}(\cdot|\cdot)$ are already known.
Some Intuition on identifying $\mu_t(\theta)$ and $f_\theta(\cdot)$

- Notice $\text{Cov}(W_{i1}, W_{it}) = \text{Cov}(\theta_i, \mu_t(\theta_i))$ is positive if $\sigma^2_{\theta} > 0$ and $\mu_t'(\theta) > 0$
- If $\mu_t(\theta) = m_{0,t} + m_{1,t}\theta$, then the problem is just like a standard measurement error problem with three measurements
- For example, IV regression of $W_2$ on $W_1$ using $W_3$ as an instrument yields

$$\frac{\text{Cov}(W_{i2}, W_{i3})}{\text{Cov}(W_{i1}, W_{i3})} = m_{1,2} \frac{m_{1,3} \sigma^2_{\theta}}{m_{1,3} \sigma^2_{\theta}} = m_{1,2}$$

- Can identify/estimate $m_{1,3}$ similarly, and $\sigma^2_{\theta}$ from $\text{Cov}(W_{i1}, W_{it})$
- General case is like nonparametric IV in context of measurement error
Now, consider permanent shocks and an $MA(1)$ process for $\varepsilon_{it}$:

$$W_{it} = \mu_t(\theta_i) + \kappa_{i,t} + \nu_{i,t}$$

$$\kappa_{i,t} = \kappa_{i,t-1} + \eta_{i,t}$$

$$\nu_{i,t} = \xi_{i,t} + \beta_t \xi_{i,t-1}$$

- Identification for most parameters/densities/functions requires $T \geq 9$
- We show identification for $T = 9$ but all results extend naturally to $T > 9$
- Use differences to eliminate correlations in shocks
Assumption 2

The following conditions hold for $T = 9$:

(i) The joint density of $\theta, W_1, W_2, W_3, \Delta W_4, \ldots, \Delta W_9$ is bounded and continuous, and so are all their marginal and conditional densities. $f_\theta(\cdot)$ is non-vanishing on $\mathbb{R}$.

(ii) All unobserved components $\eta_t, \xi_t,$ and $\theta$ are mutually independent for all $t$.

(iii) $f_{W_t|\Delta W_{t+3}}(w_t|\Delta w_{t+3})$ and $f_{\theta|W_t}(\theta|w_t)$ form a bounded complete family of distributions indexed by $\Delta W_{t+3}$ and $W_t$, respectively, for $t = 1, 2, 3$.

(iv) The functions $\mu_t(\cdot)$ are strictly monotone. For $t = 7, 8, 9$, the function $\Delta \mu_t(\theta)$ is continuously differentiable and $\Delta \mu_t'(\theta^*) = 0$ for at most a finite number of $\theta^*$.

(v) The characteristic functions of $\{W_t\}_{t=1}^9$ and $\{\Delta W_t\}_{t=4}^9$ are non-vanishing.
Theorem 1: Identification with Serially Correlated Errors

Under Assumption 2, \( f_\theta(\cdot) \), \( \{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot), \beta_t\}_{t=1}^7 \), and \( \{\mu_t(\cdot)\}_{t=1}^9 \) are identified.

Proof has three steps:

- Identify \( f_\theta(\cdot) \) and \( \mu_t(\cdot) \) for all \( t \).
- Identify \( f_{\eta_t}(\cdot) \) and \( f_{\nu_t}(\cdot) \) for \( t = 1, \ldots, 7 \).
- Identify \( f_{\xi_t}(\cdot) \) and \( \beta_t \) for \( t = 1, \ldots, 7 \).
Step 1: Identifying $f_{\theta}(\cdot)$ and $\mu_t(\cdot)$

Consider the following subset of equations:

\[
W_1 = \theta + \varepsilon_1 = \theta + \{\eta_1 + \nu_1\}
\]
\[
\Delta W_4 = \Delta \mu_4(\theta) + \Delta \varepsilon_4 = \Delta \mu_4(\theta) + \{\eta_4 + \Delta \nu_4\}
\]
\[
\Delta W_7 = \Delta \mu_7(\theta) + \Delta \varepsilon_7 = \Delta \mu_7(\theta) + \{\eta_7 + \Delta \nu_7\}
\]

- Now, identification of $f_{\theta}$ and $\mu_t(\cdot)$ functions must come from a relationship between $W_1$, $\Delta W_4$ and $\Delta W_7$
- Without monotonicity in $\Delta \mu_7$, need to generalize Lemma 1 a bit to identify $f_{\theta}(\cdot)$, $\Delta \mu_4(\cdot)$, and $\Delta \mu_7(\cdot)$
- Use a similar approach for $(W_2, \Delta W_5, \Delta W_8)$ and $(W_3, \Delta W_6, \Delta W_9)$ identifies $\mu_2(\cdot)$, $\mu_3(\cdot)$, $\Delta \mu_5(\cdot)$, $\Delta \mu_6(\cdot)$, $\Delta \mu_8(\cdot)$, and $\Delta \mu_9(\cdot)$
- Can recover all $\mu_t(\cdot)$ sequentially from $\mu_t(\cdot) = \Delta \mu_t(\cdot) + \mu_{t-1}(\cdot)$ for $t = 4, \ldots, 9$
Step 2: Identifying $f_{\eta_t}(\cdot)$ and $f_{\nu_t}(\cdot)$

Notice:

\[
W_1 - \theta = \epsilon_1 = \eta_1 + \nu_1 \\
W_3 - \mu_3(\theta) = \epsilon_3 = \eta_1 + \nu'_3
\]

where $\nu'_3 = \eta_2 + \eta_3 + \nu_3$.

- $\eta_1, \nu_1,$ and $\nu'_3$ are all independent
- We first show the density $f(\epsilon_1, \epsilon_3)$ is identified
- Then, apply Lemma 1 of Kotlarski (1967) to identify $f_{\eta_1}(\cdot)$, $f_{\nu_1}(\cdot)$, and $f_{\nu'_3}(\cdot)$
- Repeat this for $(W_2, W_4), \ldots, (W_7, W_9)$ sequentially to identify $f_{\eta_t}(\cdot)$ and $f_{\nu_t}(\cdot)$ for all $t = 1, \ldots, 7$
Step 3: Identifying $f_{\xi_t}(\cdot)$ and $\beta_t$

- We already know density of $\nu_t = \xi_{i,t} + \beta_t \xi_{i,t-1}$ for all $t = 1, \ldots, 7$ from Step 2.
- Using these densities and observed $Cov(W_{t-1}, W_t)$, it is straightforward to identify $\beta_t$ and $f_{\xi_t}(\cdot)$ sequentially for $t = 1, \ldots, 7$. 

Details
For $T \geq 9$, this general strategy can be used to identify $f_\theta(\cdot), \{\mu_t(\cdot)\}_{t=1}^T$ and $\{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot), \beta_t\}_{t=1}^{T-2}$.

Assumption 2 (iv) rules out flat regions in $\Delta\mu_t(\cdot)$ for $t = 7, \ldots, 9$.

Identification approach rules out an autoregressive process where transitory shocks never die out.

Can handle arbitrarily long $MA(q)$ process, but may require a long panel.

$MA(q)$ requires $T \geq 6 + 3q$ time periods.
Now, consider a moment-based approach

- We assume past shocks begin at age 20, accumulating to date $t$
- $f_{\xi_t} (\cdot)$ and $f_{\eta_t} (\cdot)$ are time-specific
- Assume $\theta$, $\eta_t$ and $\xi_t$ are mean zero and mutually independent with $\eta_t$ and $\xi_t$ independent over time
- Normalize $\mu_1 (\theta) = \theta$
- Let $\mu_t (\theta) = m_{0,t} + m_{1,t} \theta + ... + m_{p,t} \theta^p$ with $E[\mu_t (\theta)] = 0$ for $t = 2, ..., T$
- Allow for an autoregressive shock
Using Covariances

For ARMA(1,1), $\nu_{i,t} = \rho \nu_{i,t-1} + \xi_i,t + \beta_t \xi_i,t-1$, we have

$$E[W_{i,a,t}^2] = \sum_{i=1}^{p} \sum_{j=0}^{p} m_{i,t} m_{j,t} E[\theta^{i+j}] + \sum_{j=0}^{a-1} \sigma_{\eta_{t-j}}^2$$

$$+ \sigma_{\xi_t}^2 + \sum_{j=0}^{a-2} \rho^{2j} (\rho + \beta_{t-j})^2 \sigma_{\xi_{t-j-1}}^2$$

$$E[W_{i,a,t}W_{i,a+l,t+l}] = \sum_{i=1}^{p} \sum_{j=0}^{p} m_{i,t} m_{j,t+l} E[\theta^{i+j}] + \sum_{j=0}^{a-1} \sigma_{\eta_{t-j}}^2$$

$$+ \rho^{l-1} (\rho + \beta_{t+1}) \sigma_{\xi_t}^2 + \rho^l \sum_{j=0}^{a-2} \rho^{2j} (\rho + \beta_{t-j})^2 \sigma_{\xi_{t-j-1}}^2$$

for $l \geq 1$
Consider the number of moments & parameters for one cohort

Number of parameters:
- $2p - 1$ parameters for $E[\theta^{2p}], E[\theta^{2p-1}],...,E[\theta^2]$
- $(p + 1)(T - 1)$ parameters for $\mu_t(\theta)$ polynomials, $t = 2, ..., T$
- $2T$ parameters for $\sigma^2_{\eta_t}$ and $\sigma^2_{\xi_t}$, $t = 1, ..., T$
- $T - 1$ parameters for $\beta_t$, $t = 2, ..., T$
- 1 parameter for $\rho$
- Total number of parameters: $(4 + p)T + p - 2$

Number of moments:
- $\frac{T(T+1)}{2}$ variance/covariance terms
- $T - 1$ moments coming from $E[\mu_t(\theta)] = 0$, $t = 2, ..., T$
- Total moments: $\frac{T(T+1)}{2} + T - 1$
Identification

Necessary condition for identification: \( p \leq \frac{T^2 - 5T + 2}{2(T+1)} \)

- \( T \) must be at least 7 for \( p \geq 1 \)
- Higher order polynomial for \( \mu_t(\theta) \) requires longer panel
  - cubic \( \mu_t(\cdot) \) requires \( T \geq 12 \)
- Identification requires changes in \( \mu_t(\theta) \) function over time
- Adding higher residual moments can be helpful
  - Hausman, et al. (1991) suggest additional moments \( E[W_{i,t}W_{i,1}^{j}] \) for \( j = 2, \ldots, p, \ t = 2, \ldots, T \) to provide more direct measures related to \( m_{2,t}, \ldots, m_{p,t} \)
  - Higher moments necessary to identify higher moments of shock distributions \( f_{\eta_t}(\cdot) \) and \( f_{\xi_t}(\cdot) \)
Multiple Cohorts

- With changing distribution of cohorts over time (aging in and out of panel), it is important to account for the fact that older cohorts have accumulated a longer history of shocks.
- Additional cohorts can aid in identification, since $\mu_t(\cdot)$ does not vary across cohorts.
PSID Data: Overview

- PSID is a longitudinal survey of a representative sample of US individuals and their families
- Collected annually through 1997, biennially starting in 1999
- We use data from interview years 1971 through 2009
- Earnings are collected for the previous year, so data cover calendar years 1970-2008
- Earnings: household head’s total wages and salaries (excluding farm and business income)
- Earnings reported in 1996 dollars using CPI-U-RS
Sample Restrictions

- Core (SRC) sample with nonzero weights
  - exclude oversamples (SEO, Latino) and nonsample persons
- Male heads of households
- Ages 30-59
- Positive annual wages and weeks worked
- Non-students
- Trim top and bottom 1% of wages within each age-year cells (ten-year age group used)
- Resulting data set has 3,042 men and 32,543 person-year observations
Sample Statistics

- Race: 92% White, 6% black, 1% hispanic
- Age: mean age is 47
- Educational Attainment

<table>
<thead>
<tr>
<th>Education (years)</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elementary (1-5)</td>
<td>1.2</td>
</tr>
<tr>
<td>Middle (6-8)</td>
<td>5.0</td>
</tr>
<tr>
<td>Some High (9-11)</td>
<td>9.9</td>
</tr>
<tr>
<td>Completed High (12)</td>
<td>33.7</td>
</tr>
<tr>
<td>Some College (13-15)</td>
<td>20.0</td>
</tr>
<tr>
<td>Completed College (16)</td>
<td>20.6</td>
</tr>
<tr>
<td>Advanced Degrees (17+)</td>
<td>9.8</td>
</tr>
</tbody>
</table>
Obtaining Residuals

- We focus on the distribution of residual earnings, controlling for differences in education, race, and age.
- Run a cross-sectional regression of log earnings for each year on:
  - age dummies
  - race dummies
  - education dummies
  - race dummies $\times$ cubic polynomial in age
  - education dummies $\times$ cubic polynomial in age
Residual Earnings Inequality in the US, 1970-2008

Quantile of Log Annual Earnings Residual

Interquantile Range of Log Annual Earnings Residual

Lochner & Shin
Understanding Income Dynamics
We use second- and third-order moments to estimate the model

Assume $\beta_{j,t} = \beta_j$ for all $j = 1, ..., q$ and $t = 1, ..., T$

We assume $\sigma^2_{\eta_\tau} = \sigma^2_{\eta_1}$ and $\sigma^2_{\xi_\tau} = \sigma^2_{\xi_1}$ for all $\tau$ years prior to our data (other assumptions yield similar results)

Use minimum distance for estimation
  - weight moments by share of observations used for that moment
We begin by assuming $\mu_t(\theta)$ is linear
  - only use variances & covariances in estimation

Focus on decomposing variance of earnings residuals into
  - pricing of unobserved skills: $\text{Var}[\mu_t(\theta)]$
  - permanent shocks: $\sigma^2_{\kappa_t} = \sum_{j=0}^{a-1} \sigma^2_{\eta_t-j}$
  - transitory shocks: $\sigma^2_{\nu_t} = \sigma^2_{\xi_t} + \sum_{j=0}^{a-2} \rho^{2j} (\rho + \beta_{t-j})^2 \sigma^2_{\xi_{t-j-1}}$
### MD estimation (general time patterns)

<table>
<thead>
<tr>
<th></th>
<th>ARMA(1,1), $\mu$ const.</th>
<th>ARMA(1,1)</th>
<th>ARMA(1,5)</th>
<th>MA(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obj. Fun.</td>
<td>144.2</td>
<td><strong>114.4</strong></td>
<td>114.1</td>
<td>116.9</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.861</td>
<td>0.804</td>
<td>0.837</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.034)</td>
<td>(0.077)</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.529</td>
<td>-0.496</td>
<td>-0.504</td>
<td>0.299</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.046)</td>
<td>(0.079)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>.</td>
<td>.</td>
<td>-0.035</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>(0.051)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>.</td>
<td>.</td>
<td>-0.020</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>(0.050)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>.</td>
<td>.</td>
<td>0.022</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>(0.045)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>.</td>
<td>.</td>
<td>-0.017</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>(0.041)</td>
<td>(0.033)</td>
</tr>
</tbody>
</table>
Variance of $\theta$ and shocks

ARMA(1,1) model with time-varying $\mu_t(\theta)$
**Variance Decomposition**

**ARMA(1,1) shocks with time-varying** $\mu_t(\theta)$

---

Variance Decomposition

<table>
<thead>
<tr>
<th>Year</th>
<th>Total (Data)</th>
<th>Total (Fitted)</th>
<th>Permanent ($\kappa_t$)</th>
<th>Transitory ($\nu_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Share of Each Component

<table>
<thead>
<tr>
<th>Year</th>
<th>$\mu_t(\theta)$</th>
<th>Permanent ($\kappa_t$)</th>
<th>Transitory ($\nu_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison with time-invariant $\mu$ model

ARMA(1,1) shocks with time-varying vs. time-invariant $\mu_t(\theta)$

Time-Varying $\mu_t(\theta)$

Time-Invariant $\mu(\theta)$

Variance Decomposition

- Total (Data)
- Total (Fitted)
- $\mu_t(\theta)$ and Permanent ($\kappa_t$)
- Transitory ($\nu_t$)

Lochner & Shin
Understanding Income Dynamics
Comparison with MA(5) model

ARMA(1,1) vs. MA(5) shocks with time-varying $\mu_t(\theta)$

ARMA(1,1) vs. MA(5) shocks with time-varying $\mu_t(\theta)$

Variance Decomposition

Lochner & Shin

Understanding Income Dynamics
Comparison with MA(1) model

ARMA(1,1) vs. MA(1) shocks with time-varying $\mu_t(\theta)$
Now, assume

- $\mu_t(\theta)$ are time-varying cubic functions with $\mu_{1985}(\theta) = \theta$
- $f_\theta(\cdot)$ is a mixture of two normals
- permanent and ARMA(1,1) transitory shocks

We now use all second- and third-order moments of log earnings residuals
Variance Decomposition

Comparison with linear $\mu_t(\theta)$

**Linear $\mu_t(\theta)$**

**Cubic $\mu_t(\theta)$**
Distribution of $\theta$ (1985) and $\mu_t(\theta)$ functions

Density of $\theta$

$\mu_t(\cdot)$

- $\theta$ (1985)
- $\mu_t(\theta)$
Evolution of $\mu_t(\theta)$ distribution vs. residual distribution

Residuals

$\mu_t(\theta)$

Interquantile Range of Log Annual Earnings Residual

Interquantile Range of $\mu_t(\cdot)$
Variances and Skewness Over Time

Comparison with linear $\mu_t(\theta)$

Variance Decomposition

Skewness of Each Component
We consider identification and estimation for a model with:
- unobserved skill differences with varying ‘pricing’ functions
- permanent shocks with varying distributions
- transitory shocks with varying distributions

**Identification**
- prove nonparametric identification (without AR shocks)
- discuss minimum identification requirements for a moment-based approach

**Estimation**
- nonparametric identification strategy motivates a straightforward (reasonably simple) estimation strategy
- estimation using Minimum Distance with second- and third-order residual moments
Use log earnings residuals in PSID to estimate the model

Moment-based estimation results suggest that all components of earnings have played an important role since 1970

- ‘returns’ to unobserved skill increased broadly in 1970s and early 1980s but fell (especially for lower skill levels) in late 1980s/early 1990s (polarization?)
- variance of permanent shocks rose consistently over 1980s and 1990s
- variance of transitory shocks rose sharply in early 1980s and fluctuated thereafter

In general, inequality in unobserved skills evolves quite differently from overall residual inequality
Some Future Applications

- Cohort quality differences
  - allow $f_\theta(\cdot)$ to vary by cohort

- Differences by education
  - allow $f_\theta(\cdot)$ and $\mu_t(\theta)$ to vary by education
  - what role does change in pricing of unobservable skills play in earnings gaps by education?

- Differences by race (e.g. Chay and Lee 2000)
  - allow $\mu_t(\theta)$ to vary by race and $f_\theta(\cdot)$ by race and cohort
  - with some location restriction (e.g. $\mu_t^b(\tilde{\theta}) = \mu_t^w(\tilde{\theta})$ for some $t$ and $\tilde{\theta}$) we can identify role of differences in unobserved skill vs. unobserved skill pricing

- Incorporate changes in unobserved skill based on observables to move from residuals to earnings/wages

- Allow for multiple unobserved skills?
Recovering \( f_{W_2, \theta}(\cdot, \cdot) \)

Note that

\[
\phi_{W_1, W_2}(t_1, t_2) = E \left[ e^{-i(t_1 W_1 + t_2 W_2)} \right] \\
= E \left[ e^{-it_1 W_1} e^{-it_2 W_2} \right] \\
= E \left[ e^{-it_1 (\theta + \varepsilon_1)} e^{-it_2 W_2} \right] \\
= E \left[ e^{-it_1 \varepsilon_1} e^{-i(t_1 \theta + t_2 W_2)} \right] \\
= E \left[ e^{-it_1 \varepsilon_1} \right] E \left[ e^{-i(t_1 \theta + t_2 W_2)} \right] \\
= \phi_{\varepsilon_1}(t_1) \cdot \phi_{\theta, W_2}(t_1, t_2)
\]

where \( \phi_{W_1, W_2}(t_1, t_2) \) and \( \phi_{\varepsilon_1}(t_1) \) are already known.
Bounded complete family of distributions

$f_{\theta|W}(\theta|W)$ forms a bounded complete family of distributions indexed by $W$ if $g(\theta) = 0$ is the only bounded function that solves:

$$\int g(\theta) f_{\theta|W}(\theta|W) d\theta = 0, \quad \forall W$$

- Standard assumption in nonparametric identification literature related to invertability of conditional expectation integral function
- E.g. violated if
  - $\theta \perp \perp W$, since $g(\theta) = \theta - E(\theta)$ solves the equation above
  - $f_{\theta|W}$ is symmetric about 0 for all $W$, e.g. $W$ only affects variance of $\theta$
Step 1 continued

Consider the second subset of equations:

\[ W_2 = \mu_2(\theta) + \varepsilon_2 = \theta_2 + \{\eta_1 + \eta_2 + \nu_2\} \]
\[ \Delta W_5 = \Delta \mu_5(\theta) + \Delta \varepsilon_5 = g_5(\theta_2) + \{\eta_5 + \Delta \nu_5\} \]
\[ \Delta W_8 = \Delta \mu_8(\theta) + \Delta \varepsilon_8 = g_8(\theta_2) + \{\eta_8 + \Delta \nu_8\} \]

where \( g_t(\theta_2) \) is implicitly defined by \( \Delta \mu_t(\theta) = g_t(\mu_2(\theta)) \).

- Can identify \( f_{\theta_2}(\cdot), g_5(\cdot), \) and \( g_8(\cdot) \) using same approach
- Recover the function \( \mu_2(\cdot) \) by \( \mu_2(\theta) = F_{\theta_2}^{-1}(F_{\theta}(\theta)) \)
- Once we identify \( \mu_2(\cdot), \Delta \mu_5(\cdot) \) and \( \Delta \mu_8(\cdot) \) are identified from \( \Delta \mu_t(\theta) = g_t(\mu_2(\theta)) \)
Identifying \( f(\varepsilon_1, \varepsilon_3) \)

\[
\phi_{W_1,W_3}(\tau_1, \tau_3) = E \left[ e^{-i(\tau_1 W_1 + \tau_3 W_3)} \right] \\
= E \left[ e^{-i(\tau_1 (\theta + \varepsilon_1) + \tau_3 (\mu_3(\theta) + \varepsilon_3))} \right] \\
= E \left[ e^{-i(\tau_1 \varepsilon_1 + \tau_3 \varepsilon_3)} e^{-i(\tau_1 \theta + \tau_3 \mu_3(\theta))} \right] \\
= E \left[ e^{-i(\tau_1 \varepsilon_1 + \tau_3 \varepsilon_3)} \right] E \left[ e^{-i(\tau_1 \theta + \tau_3 \mu_3(\theta))} \right] \\
= \phi_{\varepsilon_1, \varepsilon_3}(\tau_1, \tau_3) \phi_{\theta, \mu_3(\theta)}(\tau_1, \tau_3)
\]

- 4th equality exploits independence between \((\varepsilon_1, \varepsilon_3)\) and \(\theta\)
- We can now identify \( f(\varepsilon_1, \varepsilon_3) \) from

\[
\phi_{\varepsilon_1, \varepsilon_3}(\tau_1, \tau_3) = \frac{\phi_{W_1,W_3}(\tau_1, \tau_3)}{\phi_{\theta, \mu_3(\theta)}(\tau_1, \tau_3)}
\]
Step 3: Identifying $f_{\xi_t}(\cdot)$ and $\beta_t$

- We already know density of $\nu_t = \xi_{i,t} + \beta_t \xi_{i,t-1}$ for all $t = 1, \ldots, 7$ from Step 2
- Normalizing $\xi_0 = 0$, $f_{\xi_1}(\cdot)$ is identified by $f_{\xi_1}(\cdot) = f_{\nu_1}(\cdot)$
- We can identify $\beta_2$ from

$$\text{Cov}(W_1, W_2) = \text{Cov}(\theta, \mu_2(\theta)) + \text{Var}(\eta_1) + \beta_2 \text{Var}(\xi_1),$$

since we know all other terms
- We next identify the distribution of $\xi_2$ using standard deconvolution methods, since $f_{\nu_2}(\cdot)$ and $f_{\xi_1}(\cdot)$ are already known
- In the same way, we expand $\text{Cov}(W_{t-1}, W_t)$ and identify $\beta_t$ and $f_{\xi_t}(\cdot)$ sequentially for $t = 3, \ldots, 7$