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

Lance Lochner
University of Western Ontario
NBER

Youngki Shin
University of Western Ontario

October 3, 2013

1 Introduction

Sustained growth in earnings and wage inequality over the past few decades (most notably in the U.S. but also in many other developed countries) has generated widespread interest in both its causes and consequences, spurring large bodies of research in labor economics, macroeconomics, and growth economics.[1] Perhaps, the greatest efforts have been devoted to understanding the role of skills, observed and unobserved, in accounting for the evolution of wage and earnings inequality. There is widespread agreement that the returns to observable measures of skill like education and labor market experience have increased dramatically since the early 1980s (Card, 1999; Katz and Autor, 1999; Heckman, Lochner, and Todd, 2006, 2008). There is greater disagreement about the evolution of returns to unobserved abilities and skills (e.g., see Card and DiNardo (2002); Lemieux (2006); Autor, Katz, and Kearney (2008)).[2] More generally, the broader literature has yet to reach a consensus on the factors underlying changes in residual wage and earnings inequality (i.e. inequality conditional on observable measures of skill like educational attainment and age/experience). Figure 1 shows that the evolution of total inequality in log weekly wages and earnings for 30-59 year-old American men is closely mirrored by the evolution of residual inequality, and while the

[1]For example, see surveys by Katz and Autor (1999), Acemoglu (2002), and Aghion (2002).

[2]Taber (2001) argues that increasing returns to unobserved skill in recent decades may be the main driver for the increase in measured returns to college, since individuals with higher unobserved skills are more likely to attend college.

1

Figure 1: Variance of Log Earnings and Weekly Wages and Residuals (Men Ages 30-59, PSID)
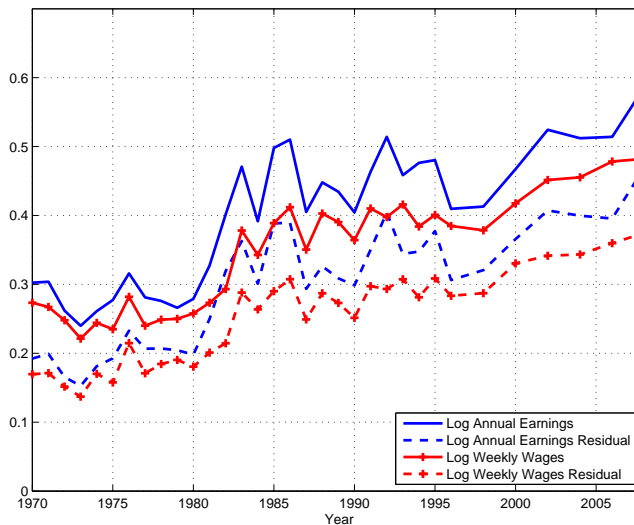
variance of log earnings is always greater than that for weekly wages, both sets of variances follow nearly identical time patterns.[3] This paper focuses on the evolution of residual earnings inequality in the U.S. from 1970 to 2008.

Most studies on the evolution of residual wage and earnings inequality in the U.S. use data from either the Current Population Surveys (CPS) or the Panel Study of Income Dynamics (PSID); however, these two literatures look at residuals through very different economic lenses. Beginning with Katz and Murphy (1992) and Juhn, Murphy, and Pierce (1993), the CPS-based literature has largely interpreted changes in residual inequality as changes in returns to unobserved abilities or skills.[4] According to this literature, the increase in residual inequality beginning in the early 1980s reflects an increase in the 'returns' to unobserved skill. Due to the general coincidence of increasing returns to measured skills and increasing residual inequality, many studies argue that skill-biased technical change (SBTC) likely explains both phenomena; however, Card and DiNardo (2002) and Lemieux (2006) raise a number of objections to this interpretation, arguing instead that institutional factors like the declining minimum wage and de-unionization may help explain rising

---

[3]Residuals are based on year-specific regressions of log earnings on age, education, race and their interactions. See Section 4 for a detailed description of the residual regressions and sample used in creating this figure.

[4]More recently, see Card and DiNardo (2002); Lemieux (2006); Autor, Katz, and Kearney (2008).

residual inequality. Most recently, Acemoglu and Autor (2011) and Autor and Dorn (2012) offer a more nuanced view of technological change, arguing that the mechanization of many routine tasks in recent decades has led to polarization in both employment and wages by skill.

In attributing changes in residual wage distributions to changes in the returns to unobserved skill, this literature largely ignores a parallel line of research based on the PSID. This literature includes studies in labor economics (Lillard and Weiss, 1979; MaCurdy, 1982; Abowd and Card, 1989; Baker, 1997; Haider, 2001; Meghir and Pistaferri, 2004; Gottschalk and Moffitt, 1994, 2009; Moffitt and Gottschalk, 2002, 2012; Bonhomme and Robin, 2010) as well as macroeconomics (Guvenen, 2007; Heathcote, Perri, and Violante, 2010; Heathcote, Storesletten, and Violante, 2010) and aims to quantify the relative importance of transitory and permanent shocks to earnings inequality over time. These studies examine the same type of wage or (usually) earnings residuals of the CPS-based literature, only they decompose these residuals into different idiosyncratic stochastic shocks (typically, permanent martingale shocks, autoregressive and moving average processes), estimating the variances of these shocks over time. Decompositions of this type are of economic interest, because the persistence of income shocks has important implications for consumption and savings behavior at the individual and aggregate levels. PSID-based studies typically estimate that increases in the variance of transitory shocks account for about one-third to one-half of the increase in total residual variance in the U.S. since the early 1970s. As Gottschalk and Moffitt (1994) point out, it is unlikely that the increasing variance of transitory shocks reflects an increase in the returns to unobserved abilities or skills, since these are likely to be fixed or, at least, changing very slowly over time for any given individual (especially older workers). While this literature often considers a rich structure for stochastic shocks, it does not explicitly account for potential changes in the pricing of unobserved skills as emphasized by the CPS-based literature.

In this paper, we consider a general framework for wage and earnings residuals that incorporates the features highlighted in both of these literatures: unobserved skills with changing non-linear pricing functions and idiosyncratic shocks that follow a rich stochastic process with permanent and transitory components. Specifically, we consider log wage and earnings residuals for individual i in period t of the form:[5]

$$W_{i,t} = \mu_t(\theta_i) + \varepsilon_{i,t} \tag{1}$$

where θ_i represents an unobserved permanent ability or skill, $\mu_t(\cdot)$ a pricing function for unobserved skills, and $\varepsilon_{i,t}$ idiosyncratic shocks. We assume that the error components $(\theta_i, \varepsilon_{i,t})$ are mutually

---

[5]This is consistent with wage/earnings functions that are multiplicatively separable in observable factors (like education and experience), unobserved skills, and idiosyncratic wage/earnings shocks.

independent and allow for a rich stochastic process for $\varepsilon_{i,t}$ as in much of the literature on earnings dynamics.

Economic shifts in the demand for unobserved skills or changes in minimum wages or union premiums are likely to be reflected in $\mu_t(\cdot)$. To the extent that unobserved skills are important, the wages and earnings of workers at similar points in the wage distribution are likely to co-move over time as the labor market rewards their skill set more or less. The recent literature on 'polarization' in the U.S. labor market (Acemoglu and Autor, 2011; Autor and Dorn, 2012) suggests that these skill pricing functions have become more convex in recent years, rewarding skill more and more at the top of the distribution but not at the bottom. This motivates our emphasis on general non-linear $\mu_t(\cdot)$ pricing functions.

By contrast, permanent and transitory shocks embodied in $\varepsilon_{i,t}$ are idiosyncratic and unrelated across workers regardless of how close they may be within the skill distribution. We consider a process for $\varepsilon_{i,t}$ that is general enough to account for shocks that produce lasting changes in a worker's earnings (e.g. job displacement, moving from a low- to high-paying firm, or a permanent disability) and as well as those more short-term in nature (e.g. temporary illness, family disruption, or a good/bad year for the worker's employer).[6] The latter may be persistent, but their importance depreciates with time.

Both permanent and transitory shocks are likely to be unpredictable, so their relative importance has implications for the dynamics of consumption and savings behavior. Large increases in the variance of permanent (or even very persistent transitory) shocks should lead to increases in consumption inequality over time. Changes in skill prices, $\mu_t(\cdot)$, are likely to be more predictable and smooth over time, since they are largely driven by economic changes in the supply and demand for skills or major policy changes. To the extent that they are well-anticipated, changes in $\mu_t(\cdot)$ may have little effect on consumption inequality over time for a given cohort; however, an increase in the returns to unobserved skills over time should raise consumption inequality across successive cohorts.

An important question is whether these earnings components can be separately identified (using standard panel data sets) without strong distributional or functional form assumptions. We, therefore, begin with an analysis of nonparametric identification, drawing on insights from the mea-

---

[6]A growing literature studies the implications for wage dynamics of different assumptions about wage setting in markets with search frictions and worker productivity shocks (Flinn, 1986; Postel-Vinay and Turon, 2010; Yamaguchi, 2010; Burdett et al., 2011; Bagger et al., 2011). This literature incorporates unobserved worker human capital or productivity differences but does not account for changing market demands for human capital, implicitly assuming $\mu_t(\theta) = \mu(\theta)$.

surement error literature (especially, Hu and Schennach (2008); Schennach and Hu (2013)) and the analysis of Cunha et al. (2010). We focus on the case where ε_{it} contains a permanent (Martingale) shock κ_{it} and a shock characterized by a moving average process ν_{it}, allowing the distributions for these shocks to vary over time.[7] We derive conditions necessary for nonparametric identification of the distribution for θ and all $\mu_t(\cdot)$ pricing functions, as well as (nearly) all distributions and parameters characterizing the stochastic process for idiosyncratic shocks.

Assuming ν_t follows an $MA(q)$ process, our main result establishes that a panel of length $T \geq 6 + 3q$ time periods is needed for full nonparametric identification. For modest q, this is easily satisfied with many common panel data sets like the PSID or National Longitudinal Surveys of Youth (NLSY). Intuitively, identification of the distribution of θ and the $\mu_t(\cdot)$ pricing functions derives from the fact that correlations in wage residual changes far enough apart in time are due the unobserved component θ and not idiosyncratic permanent or transitory shocks. Once these are identified, the distributions of permanent and transitory shocks can be identified from joint distributions of residuals closer together in time.

We next consider identification and estimation using a moment-based approach standard in the literature on earnings dynamics. Here, we restrict $\mu_t(\cdot)$ to be polynomial functions of arbitrary order p. We discuss minimal data and moment requirements for identification of $\mu_t(\cdot)$; order $2, ..., 2p$ moments on the θ distribution; order $2, ..., k$ moments on distributions for both permanent and transitory shocks; and all other parameters characterizing the stochastic process for ε_{it}. To identify moments of order k for the distributions of idiosyncratic shocks, residual moments of the same order are needed.

Much of the PSID-based literature focuses on variance decompositions over time. We show that when the transitory shock ν_t follows an $MA(1)$ process and $\mu_t(\cdot)$ functions are assumed to be cubic polynomials, using only variances and covariances of log earnings residuals over time to estimate the contributions of unobserved skill prices, permanent shocks, and transitory shocks to the total residual variance over time requires panel data with at least ten periods of observation. Higher order polynomials require longer panels. Alternatively, one can secure identification with shorter panels if higher order residual moments are incorporated in estimation.[8]

Using minimum distance estimation and second- and third-order residual moments, we estimate

---

[7]Assumptions on ε_{it} of this nature are often employed in the literature (Abowd and Card, 1989; Blundell and Preston, 1998; Gourinchas and Parker, 2002; Meghir and Pistaferri, 2004; Blundell, Pistaferri, and Preston, 2008; Heathcote, Perri, and Violante, 2010).

[8]For example, Hausman et al. (1991) show the value of incorporating moments of the form $E(W_{i,1}^j W_{i,t})$ for $j = 2, ..., p - 1$.

our model with linear and cubic $\mu_t(\cdot)$ functions and different assumptions about the process for transitory shocks ν_t. We first show that allowing for time-varying unobserved skill pricing functions significantly improves the model's fit to the data relative to the standard assumption that $\mu_t(\theta)$ are fixed. Accounting for unobserved skills is important for understanding the evolution of log earnings residuals. We next decompose the variance of residuals into components for (i) unobserved skill prices, (ii) permanent shocks, and (iii) transitory shocks. Our decompositions are not very sensitive to assumptions about the order of polynomial for $\mu_t(\cdot)$ or the process for ν_t. The returns to unobserved skills rose over the 1970s and early 1980s, fell over the late 1980s and early 1990s, and then remained quite stable through the end of our sample period. The variance of permanent shocks declined slightly over the 1970s, then rose systematically through 2002, whereas the variance of transitory shocks jumped up considerably in the early 1980s but shows little long-run trend otherwise over the more than thirty year period we study.

Our estimates assuming cubic $\mu_t(\cdot)$ allow us to examine more general changes in the distribution of returns to unobserved skills over time. These estimates suggest that the returns to skill evolved similarly – first rising then falling – throughout the skill distribution over the 1970s, 1980s and early 1990s. From 1995 onward, there is little change in the return to unobserved skill at the top of the distribution, but the return continues to decline at the bottom until there is very little difference in expected earnings from those at the bottom and middle of the unobserved skill distribution. The declining return to skill at the bottom (over the late 1990s and 2000s) is consistent with the polarization story emphasized by Autor, Katz, and Kearney (2008) and Autor and Dorn (2012); however, we do not observe the rising returns to unobserved skills at the top that is evident in log earnings residuals over this period. The time patterns for the distribution of $\mu_t(\theta)$ are notably different from those for log earnings residuals, suggesting that the latter are not necessarily very informative about changes in the pricing of unobserved skills.

This paper proceeds as follows. In Section 2, we provide nonparametric identification results for our model with unobserved ability/skills, permanent shocks, and transitory shocks following an $MA(q)$ process. Section 3 briefly discusses estimation and identification using a moment-based approach, assuming polynomial $\mu_t(\cdot)$ pricing functions. We describe the PSID data used to estimate earnings dynamics for American men in Section 4 and report our empirical findings in Section 5. We offer concluding thoughts in Section 6.

2 Nonparametric Identification

In this section, we provide nonparametric identification results for our model.

The baseline model is the following factor model:

$$W_{it} = \mu_t(\theta_i) + \varepsilon_{it} \ \text{ for } \ t = 1, \ldots, T, \text{ and } i = 1, \ldots, n, \tag{2}$$

where the distributions of unobserved factors θ_i and ε_{it} are unspecified, the functional form of $\mu_t(\cdot)$ is also unspecified but strictly increasing, and we only observe $\{W_{it}\}$ (earnings or wage residuals in our empirical context). We consider a short time panel, i.e. n is large and T is (relatively) small and fixed. For notational simplicity, we will drop the cross-sectional subscript i except where there may otherwise be some confusion. First, we derive our identification result when ε_t is independent over t. We next generalize this result when there is some serial correlation in ε_t.

2.1 Case 1: Serially Independent ε_t

From the model in Equation (2), we want to identify the following objects: (i) the distribution of θ, (ii) the distributions of ε_t for all t, and (iii) the functions $\mu_t(\cdot)$ for all t. We first define some notation. For generic random variables A and B, let $f_A(\cdot)$ and $f_{A|B}(\cdot|\cdot)$ denote the probability density function of A and the conditional probability density function of A given B, respectively. Similarly, $F_A(\cdot)$ and $F_{A|B}(\cdot|\cdot)$ denote their cumulative distribution functions, and $\phi_{A,B}(\cdot)$ denotes the joint characteristic function. We want to identify $f_\theta(\cdot)$, $f_{\varepsilon_t}(\cdot)$, and $\mu_t(\cdot)$ for all t in the model.

Since all components are nonparametric, we need some normalization. Throughout the paper, we impose $\mu_1(\theta) = \theta$. The following regularity conditions ensure identification:

Assumption 1. *The following conditions hold in equation* (2) *for $T = 3$:*

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

(ii) *All factors are independent, i.e. $\theta \perp\!\!\!\perp \varepsilon_t$ 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 function $\mu_3(\cdot)$ is strictly monotone.*

(v) *We normalize $\mu_1(\theta) = \theta$ and $E[\varepsilon_t] = 0$.*

Condition (i) assumes a well-defined joint density of the persistent factor θ and observed residuals. In our empirical setup, they are all continuous random variables, and this condition holds naturally. Condition (ii) is the mutual independence assumption commonly imposed in linear factor models. Conditions (iii) and (iv) are the key requirements for identification. Heuristically speaking, condition (iii) requires enough variation for each conditional density given different values of the conditioning variable. For example, exponential families satisfy this condition. Newey and Powell (2003) apply it to identification of nonparametric models with instrumental variables, and it is standard in many nonparametric analyses. Strict monotonicity of $\mu_3(\cdot)$ in condition (iv) is natural in our empirical context. It assures distinct eigenvalues in the spectral decomposition. Condition (v) is a standard location and scale normalization.

The following lemma establishes identification for $T = 3$, so a panel with 3 (or more) periods is necessary for nonparametric identification of the full model. This result is a special case of Theorem 1 in Hu and Schennach (2008)

Lemma 1. *Under Assumption 1, $f_\theta(\cdot), f_{\varepsilon_t}(\cdot)$, and $\mu_t(\cdot)$ are identified for all t.*

The proof for this (and subsequent results) is provided in Appendix A. Here, we sketch its key steps. The spectral decomposition result in Theorem 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 $f_{W_1,W_2,W_3}(\cdot,\cdot,\cdot)$ which is already known from observations. Then, $f_\theta(\cdot)$ can be recovered from $f_{W_3,\theta}(\cdot,\cdot)$ by integrating out θ, which is the result of Theorem 2 in Cunha, Heckman, and Schennach (2010). Next, for $t = 2,3$, we can identify $\mu_t(\cdot)$ from the conditional density $f_{W_t|\theta}(\cdot,\cdot)$ since we know that $E[W_t|\theta] = \mu_t(\theta)$ from $E[\varepsilon_t|\theta] = E[\varepsilon_t] = 0$. Finally, we can identify $f_{\varepsilon_t|\theta}(\cdot|\cdot)$ applying the standard variable transformation given $W_t = \mu_t(\theta) + \varepsilon_t$ with known $f_{W_t|\theta}(\cdot|\cdot)$ and $\mu_t(\cdot)$, from which $f_{\varepsilon_t}(\cdot)$ is recovered immediately.

2.2 Case 2: Serially Correlated ε_t

We now generalize the model to allow for serial correlation in ε_t. Specifically, we decompose the idiosyncratic error ε_t into two components: a persistent shock, κ_t, and a transitory shock, ν_t. The persistent shock follows a martingale/unit root process and the transitory shock follows a moving average of order q. For simplicity, we analyze the case of $q = 1$ in detail; however, the results can be readily extended to any finite q as we discuss below. This model assumes $\varepsilon_t = \kappa_t + \nu_t$, so

$$W_t = \mu_t(\theta) + \kappa_t + \nu_t \tag{3}$$

where $\kappa_t = \kappa_{t-1} + \eta_t$ and $\nu_t = \xi_t + \beta_t \xi_{t-1}$. The unobserved components η_t, ξ_t, and θ are mutually independent. The distribution of (η_t, ξ_t) and β_t can vary over time, but $E[\eta_t]$ and $E[\xi_t]$ are

normalized to zero. We also normalize $\mu_1(\theta) = \theta$ and $\kappa_0 = \xi_0 = 0$. For additional notation, let $\Delta A_t \equiv A_t - A_{t-1}$ and $\theta_t \equiv \mu_t(\theta)$. The following assumption ensures identification for $T \geq 9$.

Assumption 2. *The following conditions hold in equation* (3) *for $T = 9$:*

(i) *The joint density of θ, 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. The density of θ, $f_\theta(\cdot)$, is non-vanishing on $\mathbb{R}$.*

(ii) *All unobserved components η_t, ξ_t, and θ are mutually independent for all t.*

(iii) *$f_{W_1|\Delta W_4}(W_1|\Delta W_4)$ and $f_{\theta|W_1}(\theta|W_1)$ form a bounded complete family of distributions indexed by ΔW_4 and W_1, respectively. The same condition holds for $(W_2, \Delta W_5, \theta_2)$ and $(W_3, \Delta W_6, \theta_3)$.*

(iv) *The functions $\mu_t(\cdot)$ are strictly monotone. In addition, the function $\Delta\mu_7(\theta)$ is continuously differentiable, and $\Delta\mu_7'(\theta^*) = 0$ for at most a finite number of θ^*. The same condition holds for $\Delta\mu_8(\cdot)$ and $\Delta\mu_9(\cdot)$.*

(v) *We normalize that $\mu_1(\theta) = \theta$, $E[\eta_t] = E[\xi_t] = 0$, $\kappa_0 = \xi_0 = 0$.*

(vi) *The characteristic functions of $\{W_t\}_{t=1}^9$ and $\{\Delta W_t\}_{t=4}^9$ do not vanish.*

The conditions in Assumption 2 mainly extend those in Assumption 1 to a longer time period and to include some differences in $\{W_t\}$. Differencing is required to cancel out the persistent shock κ_t, which helps to map this problem into that of Lemma 1. Unless ΔW_t is degenerate on an interval, these conditions hold in a similar situation to that described for Assumption 1. In Condition (i), we additionally assume that θ is a continuous random variable on $\mathbb{R}$. Condition (ii) imposes similar mutual independence between underlying unobserved components. Notice that condition (iv) allows $\Delta\mu_t(\cdot)$ to be non-monotone, which is crucially different from its analogue in Assumption 1 (iv). Condition (vi) is a standard regularity condition imposed in the deconvolution literature (see Schennach (2004) and the references therein). Combined with the independence assumption in (ii), this condition implies that the characteristic functions for θ_t, κ_t, and ν_t are all non-vanishing.[9] This allows for non-monotone $\Delta\mu_t(\cdot)$. We now introduce the main identification theorem.

Theorem 1. *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.*

---

[9]Common examples of vanishing characteristic functions include the uniform, triangular, and symmetrically truncated normal distributions. None of these are likely to be relevant for our empirical case.

The proof provided in Appendix A proceeds in three main steps. First, we jointly consider distributions of $(W_t, \Delta W_{t+3}, \Delta W_{t+6})$ for $t = 1, 2, 3$. For example, consider:

$$W_1 = \theta + \varepsilon_1 = \theta + \eta_1 + \nu_1$$

$$\Delta W_4 = \Delta \mu_4(\theta) + \Delta \varepsilon_4 = \Delta \mu_4(\theta) + \eta_4 + \Delta \nu_4$$

$$\Delta W_7 = \Delta \mu_7(\theta) + \Delta \varepsilon_7 = \Delta \mu_7(\theta) + \eta_7 + \Delta \nu_7.$$

Note that $(\varepsilon_1, \Delta \varepsilon_4, \Delta \varepsilon_7)$ are mutually independent. The differences ΔW_4 and ΔW_7 are analogous to W_2 and W_3 in the simpler model with serially independent ε_t. Ideally, we would simply apply Lemma 1 to identify $f_\theta(\cdot)$ and all $\mu_t(\cdot)$ functions, then establish identification of β_t, $f_{\eta_t}(\cdot)$ and $f_{\xi_t}(\cdot)$. Unfortunately, the analogue to Assumption 1 (iv) requires monotonicity in $\Delta \mu_t(\cdot)$ functions for $t = 4, ..., 9$ in this model. To relax this, we establish identification for two distinct cases depending on the functional form of $\Delta \mu_t(\cdot)$ for $t = 7, 8, 9$.

Consider two cases for $\Delta \mu_7(\cdot)$: (i) $\Delta \mu_7(\theta) = a + b \ln(e^{c\theta} + d)$ and (ii) $\Delta \mu_7(\theta) \neq a + b \ln(e^{c\theta} + d)$ for $a, b, c (\neq 0)$, and $d \in \mathbb{R}$. In the first case, the function $\Delta \mu_7(\theta)$ is one-to-one in θ, so we can apply Lemma 1 to establish identification of $f_\theta(\cdot)$, $\Delta \mu_4(\cdot)$, and $\Delta \mu_7(\cdot)$.[10] In the second case when $\Delta \mu_7(\theta) \neq a + b \ln(e^{c\theta} + d)$, we apply Theorem 1 in Schennach and Hu (2013) to the pair $(W_1, \Delta W_7)$ in order to identify the function $\Delta \mu_7(\cdot)$ and the densities f_θ, f_{ε_1}, and $f_{\Delta \varepsilon_7}$. Note that f_{ε_1} and the independence between ε_1 and θ make it possible to identify $f_{W_1|\theta}$. Then, the function $\Delta \mu_4(\cdot)$ is identified from the conditional density $f_{\Delta W_4|\theta}$, which is recovered from the completeness of $f_{\theta|W_1}$ and $f_{W_1|\Delta W_4}$.

A similar approach can be taken for triplets $(W_2, \Delta W_5, \Delta W_8)$ and $(W_3, \Delta W_6, \Delta W_9)$; however, these cases are slightly more complicated since W_2 and W_3 depend on $\mu_2(\cdot)$ and $\mu_3(\cdot)$, respectively. Still, monotonicity of the functions $\mu_t(\cdot)$ for $t = 2, 3$ and knowledge of $f_\theta(\cdot)$ enables identification of $\mu_t(\cdot)$, $\Delta \mu_{t+3}(\cdot)$, and $\Delta \mu_{t+6}(\cdot)$ for $t = 2, 3$. Altogether, we identify $f_\theta(\cdot)$ and all $\mu_t(\cdot)$ functions from these three sets of triplets. It is worth emphasizing that this identification result relies on two practical assumptions about the $\mu_t(\cdot)$ functions: (i) $\mu_2(\cdot)$ and $\mu_3(\cdot)$ are strictly increasing functions, and (ii) $\Delta \mu_t(\cdot)$ is only constant on a finite number of θ's for $t = 7, 8, 9$.

In a second step, we establish identification of $f_{\eta_t}(\cdot)$ and $f_{\nu_t}(\cdot)$ for $t = 1, \ldots, 7$. This step builds on the fact that $\varepsilon_t = \kappa_t + \nu_t$ and $\varepsilon_{t+2} = \kappa_t + \nu'_{t+2}$, where $\nu'_{t+2} = \eta_{t+1} + \eta_{t+2} + \nu_{t+2}$. We first show that the joint density for $(\varepsilon_t, \varepsilon_{t+2})$ can be recovered from the joint density for (W_t, W_{t+2}) and

---

[10]This condition is also a sufficient condition to ensure that $f_{\Delta W_7|\theta}(w|\bar{\theta}) \neq f_{\Delta W_7|\theta}(w|\tilde{\theta})$ for all $\bar{\theta} \neq \tilde{\theta}$ with positive probability, which is required for identification (see the proof of Lemma 1). Another sufficient condition for different conditional density functions is the monotonicity utilized in Lemma 1.

knowledge of $f_\theta(\cdot)$, $\mu_t(\cdot)$ and $\mu_{t+2}(\cdot)$. With this joint density, it is straightforward to identify $f_{\kappa_t}(\cdot)$ and $f_{\nu_t}(\cdot)$ for $t = 1, \ldots, 7$ using Lemma 1 in Kotlarski (1967), since κ_t, ν_t, and ν'_{t+2} are mutually independent. From $f_{\kappa_t}(\cdot)$, we can easily obtain $f_{\eta_t}(\cdot)$ for $t = 1, \ldots, 7$. Importantly, we are unable to identify distributions for η_t and ν_t for $t = 8, 9$ unless we have additional periods of data. In a third step, we use the joint distributions of (W_t, W_{t+1}) and knowledge of $f_\theta(\cdot)$, all $\mu_t(\cdot)$, and f_{η_t} and f_{ν_t} for $t = 1, \ldots, 7$ to identify $f_{\xi_t}(\cdot)$ and β_t for $t = 1, \ldots, 7$.

2.3 Some General Comments on Identification

For panels of length $T \geq 9$, the same general strategy as above can be used to identify $f_\theta(\cdot)$, $\{\mu_t(\cdot)\}_{t=1}^T$ and $\{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot), \beta_t\}_{t=1}^{T-2}$. The fact that distributions of transitory and permanent shocks cannot be identified for the final two periods is useful to keep in mind below when we estimate our models using data from the PSID.

Our identification strategy can also be used for more general $MA(q)$ processes. In this case, we need to consider triplets of the form $(W_t, \Delta W_{t+q+2}, \Delta W_{t+2q+4})$ to ensure independence across the three observations. As q increases, we also need to include additional sets of triplets to 'roll' over in step 1 of our proof. Thus, a panel of length $T \geq 6 + 3q$ is needed, so the required panel length increases with persistence in the moving average shock at the rate of $3q$. We can still identify $\mu_t(\cdot)$ for all T periods; however, we can only identify $f_{\eta_t}(\cdot)$, $f_{\xi_t}(\cdot)$, and β_t up through period $T - q - 1$.

Finally, our approach rules out any stochastic process that does not eventually die out, including the commonly employed $AR(1)$ process. Independence across some subsets of observations is crucial at a number of points in our identification proof, so accommodating these type of errors would require a very different approach. Still, our results generalize to an arbitrarily long $MA(q)$ process provided q is finite. Of course, the data demands grow quickly with q making it impractical to estimate models with q much larger than five in typical panel survey data sets.[11]

3 A Moment-Based Approach

We next consider a moment-based estimation approach that simultaneously uses data from all time periods; however, we restrict our analysis to the case where

$$\mu_t(\theta) = m_{0,t} + m_{1,t}\theta + \ldots + m_{p,t}\theta^p \qquad (4)$$

---

[11]Models with q as large as 10 might feasibly be estimated in administrative data sets that effectively contain lifetime earnings records.

are pth order polynomials. Because our data contain multiple cohorts with the cohort distribution changing over time, it is useful to explicitly incorporate age, a:

$$W_{i,a,t} = \mu_t(\theta_i) + \kappa_{i,a,t} + \nu_{i,a,t} \tag{5}$$

$$\kappa_{i,a,t} = \kappa_{i,a-1,t-1} + \eta_{i,a,t} \tag{6}$$

$$\nu_{i,a,t} = \xi_{i,a,t} + \beta_{1,t}\xi_{i,a-1,t-1} + \beta_{2,t}\xi_{i,a-2,t-2} + ... + \beta_{q,t}\xi_{i,a-q,t-q}. \tag{7}$$

As above, we normalize $\mu_1(\theta) = \theta$ and continue to assume all residual components are mean zero in each period (for each age group/cohort): $E[\mu_t(\theta_i)] = E[\kappa_{i,a,t}] = E[\nu_{i,a,t}] = 0$ for all a, t.

In this section, we describe residual moment conditions as well as minimum data and moment requirements for identification. In particular, we discuss the moments and data needed to identify various moments of shock and unobserved skill distributions as well as different order polynomials for the $\mu_t(\cdot)$ functions. In Section 5, we use these moments to estimate our model. We focus on the evolution of skill pricing functions over time and decompose the variance in log earnings residuals over time into components related to: (i) the pricing of unobserved skills $\mu_t(\theta)$, (ii) permanent shocks κ_{it}, and (iii) transitory shocks ν_{it}. We further extend our analysis to examine the evolution of higher moments of distributions related to these three components.

3.1 Moments, Parameters and Identification

We assume individuals begin receiving shocks $\eta_{i,a,t}$ and $\xi_{i,a,t}$ when they enter the labor market at age $a = 1$. Thus, $\kappa_{i,0,t} = \nu_{i,0,t} = 0$ and $\eta_{i,a,t} = \xi_{i,a,t} = 0$ for all $a \leq 0$. We also assume that the distributions of $\eta_{i,a,t}$ and $\xi_{i,a,t}$ shocks are age-invariant, changing only with time. Importantly, the distributions of $\kappa_{i,a,t}$ will depend on age as older individuals will have experienced a longer history of shocks over their working lives. Define the following moments: $\sigma_{\eta_t}^k \equiv E[\eta_{i,a,t}^k]$ and $\sigma_{\xi_t}^k \equiv E[\xi_{i,a,t}^k]$ for all a, t. Due to mutual independence between $\eta_{i,a,t}$ and $\xi_{i,a,t}$ and their independence across time, these assumptions imply that

$$\sigma_{\kappa_{a,t}}^k \equiv E[\kappa_{i,a,t}^k] = \sum_{j=0}^{a-1} \sigma_{\eta_{t-j}}^k$$

$$\sigma_{\nu_{a,t}}^k \equiv E[\nu_{i,a,t}^k] = \sigma_{\xi_t}^k + \sum_{j=1}^{\min\{q,a-1\}} \beta_{j,t}^k \sigma_{\xi_{t-j}}^k.$$

Because all shocks are assumed to be mean zero, these moments are equivalent to central moments.

With these assumptions, we have the following residual variances for individuals age a in time t:

$$E[W_{i,a,t}^2|a,t] = \sum_{j=0}^{p}\sum_{j'=0}^{p} m_{j,t}m_{j',t}E[\theta^{j+j'}] + \sum_{j=0}^{a-1}\sigma_{\eta_{t-j}}^2 + \sigma_{\xi_t}^2 + \sum_{j=1}^{\min\{q,a-1\}}\beta_{j,t}^2\sigma_{\xi_{t-j}}^2 \tag{8}$$

and covariances:

$$E[W_{i,a,t}W_{i,a+l,t+l}|a,t,l] = \sum_{j=0}^{p}\sum_{j'=0}^{p} m_{j,t}m_{j',t+l}E[\theta^{j+j'}] + \sum_{j=0}^{a-1}\sigma^2_{\eta_{t-j}} + E(\nu_{i,a,t}\nu_{i,a+l,t+l}) \text{ for } l \geq 1. \quad (9)$$

The last term, $E(\nu_{i,a,t}\nu_{i,a+l,t+l})$, reflects the covariance of transitory shocks and is straightforward to determine for any q. These covariances are generally non-zero for $l \leq q$ and zero otherwise. For $q = 1$ (i.e. an $MA(1)$ process for ν_t), we have $E(\nu_{i,a,t}\nu_{i,a+1,t+1}) = \beta_{1,t+1}\sigma^2_{\xi_t}$ and $E(\nu_{i,a,t}\nu_{i,a+l,t+l}) = 0, \forall l \geq 2$.

As equation (8) makes clear, the variance of earnings residuals can change over time for three reasons: (i) unobserved skills may become more or less valuable (i.e. $\mu_t(\cdot)$ may change), (ii) permanent shocks accumulate, and (iii) transitory shocks may become more or less variable. Covariances across time are key to sorting out these three potential factors. Holding t constant in equation (9) but varying l, the second term due to permanent shocks remains constant, while the final term disappears altogether for $l > q$. Analogous to our nonparametric identification results above, we can learn about the $\mu_t(\cdot)$ functions and the distribution of θ by looking at covariances between residuals in some period t and changes in residuals more than q periods later.

3.1.1 A Single Cohort

How many periods of data do we need if residual variances and covariances are used in estimation? It is useful to begin our analysis with a single cohort (normalizing $a = t$), following them over time for $t = 1, ..., T$ where $T \geq 3$. For simplicity, consider an $MA(1)$ process for ν_t. In this case, we need to identify/estimate a total of $(4 + p)T + p - 9$ parameters: $2p - 1$ parameters for $E[\theta^2], ..., E[\theta^{2p}]$; $(T-1)(p+1)$ parameters for $\mu_t(\theta)$ polynomials for $t = 2, ..., T$; $2(T-2)$ parameters for $\sigma^2_{\eta_t}$ and $\sigma^2_{\xi_t}$ for $t = 1, ..., T-2$; and $T - 3$ parameters for $\beta_{1,t}$, $t = 2, ..., T - 2$.[12] For T periods of data, we have a total of $T(T+1)/2 + T - 1$ moments, which includes $T(T+1)/2$ unique variance/covariance terms and $T - 1$ moments coming from $E[\mu_t(\theta_i)] = 0$ for $t = 2, ..., T$. So, a necessary condition for identification is that $(4 + p)T + p - 9 \leq T(T+1)/2 + T - 1$. Re-arranging this inequality, identification requires

$$p \leq \frac{T^2 - 5T + 16}{2(T+1)}$$

as well as $T \geq 3$. Using only variances and covariances in estimation, identification with cubic $\mu_t(\cdot)$ functions requires a panel of length $T \geq 10$, while quadratic $\mu_t(\cdot)$ functions require $T \geq 8$. Despite

---

[12]These parameter counts incorporate the normalization $\mu_1(\theta) = \theta$. The normalization $\xi_{i,0,0} = 0$ means that we are unable to identify $\beta_{1,1}$. Furthermore, as in Theorem 1, $(\beta_{1,t}, \sigma^2_{\eta_t}, \sigma^2_{\xi_t})$ are unidentified for $t = T - 1, T$ for $q = 1$.

the data requirements implied by Theorem 1, we require more than nine periods of data when $\mu_t(\cdot)$ is a high order polynomial if we only use covariances and variances to estimate the model. For linear $\mu_t(\cdot)$, $T \geq 3$ is necessary and sufficient for identification.

Using higher order moments can reduce the required panel length. For example, Hausman, et al. (1991) show the value of adding the moments $E[W_1^k W_t]$ for $k = 2, ..., p$. While incorporating these moments would add $2(p-1)$ additional parameters $\left(\sigma_{\eta_1}^3, ..., \sigma_{\eta_1}^{p+1}, \sigma_{\xi_1}^3, ..., \sigma_{\xi_1}^{p+1} \right)$, the number of additional parameters is independent of T. Of course, the number of moments increases by $(T-1)(p-1)$. These extra moments provide relatively direct information about $m_{j,t}$ parameters (and higher moments of the distribution of θ) as t varies.

More generally, we could incorporate a broader set of higher moments in estimation. Indeed, if we want to identify moments for the distribution of shocks up to order k (e.g. $\sigma_{\eta_t}^k$ and $\sigma_{\xi_t}^k$) for all $t = 1, ..., T$, we need to incorporate up to kth order moments for residuals in all periods. Including $E[W_t^j]$ for $j = 3, ..., k$ adds $(k-2)T$ new moments (relative to variance/covariances only), but it adds $2(k-2)(T-2)$ new parameters for higher moments of η_t and ξ_t as well as $(k-2)p$ new parameters for higher moments of θ. Thus, higher order cross-product terms should also be incorporated.[13] In practice, it may be difficult to precisely estimate higher order residual moments given the size of standard panel data sets.

3.1.2 Multiple Cohorts

In many applications, it is common to follow multiple cohorts at once. If the distribution of cohorts changes over time (e.g. new cohorts enter the data at later dates while older ones age out of the sample), then it is important to account for this directly as in equations (8) and (9). While we have assumed that the distributions for η_{it} and ξ_{it} depend only on time and not age or cohort, the distribution of $\kappa_{i,a,t}$ varies with age, since older cohorts have accumulated a longer history of permanent shocks.[14]

Differences in the variance and covariance terms across age/cohort for the same time periods in equations (8) and (9) can be used to help identify the effects of permanent and transitory

---

[13]Including all cross-product moments from order $2, ..., k$ yields a total of

$$\sum_{j=2}^{k} \binom{T+j-1}{j} \quad \text{moments}$$

plus $T - 1$ moments for $E[\mu_t(\theta)] = 0$.

[14]This variation adds new parameters to be identified/estimated for each additional cohort. In particular, we must identify/estimate separate $\sigma_{\kappa_{a(0),0}}^k$ for each cohort, where $a(0)$ reflects their age at date $t = 0$. At very young ages (i.e. $a < q$), the distribution of $\nu_{i,a,t}$ also varies with age.

shocks, since terms related to unobserved skills are the same for all cohorts/ages. (Recall, $\mu_t(\cdot)$ are assumed to be market pricing functions that depend only on time and not age or cohort.) The value of additional cohorts is most evident when looking at equation (9) for $l > q$. In this case, the final term due to transitory shocks disappears, while the first term is the same for all cohorts. By comparing these covariances across cohorts for different t, we can easily identify the variances of permanent shocks over time.

Allowing the distributions of transitory and permanent shocks to vary more freely with age/cohort would complicate the problem slightly; however, it is still possible to use the same general approach to identify the variances of permanent shocks by age/cohort and time. Allowing for cohort differences in the distribution of θ would raise more fundamental problems, since any terms related to unobserved skills would become cohort-specific. Still, as long as skill pricing functions are independent of age/cohort, the inclusion of additional cohorts should provide an additional source of identification.

4 PSID Data

The PSID is a longitudinal survey of a representative sample of individuals and families in the U.S. beginning in 1968. The survey was conducted annually through 1997 and biennially since. We use data collected from 1971 through 2009. Since earnings and weeks of work were collected for the year prior to each survey, our analysis considers earnings and weekly wages from 1970-2008.

Our sample is restricted to male heads of households from the core (SRC) sample.[15] We use earnings from any year these men were ages 30-59, had positive wage and salary income, worked at least one week, and were not enrolled as a student. Our earnings measure reflects total wage and salary earnings (excluding farm and business income) and is denominated in 1996 dollars using the CPI-U-RS. We trim the top and bottom 1% of all earnings measures within year by ten-year age cells. The resulting sample contains 3,302 men and 33,207 person-year observations – roughly ten observations for each individual.

Our sample is composed of 92% whites, 6% blacks and 1% hispanics with an average age of 47 years old. We create seven education categories based on current years of completed schooling: 1-5 years, 6-8 years, 9-11 years, 12 years, 13-15 years, 16 years, and 17 or more years. In our sample,

---

[15]We exclude those from any PSID oversamples (SEO, Latino) as well as those with non-zero individual weights. The earnings questions we use are asked only of household heads. We also restrict our sample to those who were heads of household and not students during the survey year of the observation of interest as well as two years earlier. Our sampling scheme is very similar to that of Gottschalk and Moffitt (2012), except that we do not include earnings measures before age 30.

16% of respondents finished less than 12 years of schooling, 34% had exactly 12 years of completed schooling, 20% completed some college (13-15 years), 21% completed college (16 years), and 10% had more than 16 years of schooling.

Our analysis focuses on residual earnings and weekly wages, controlling for differences in educational attainment, race, and age. Specifically, we use residuals from year-specific regressions of log earnings or weekly wages on age, race, and education indicators, along with interactions between race and education indicators and a third order polynomial in age. Figure 2 shows selected quantiles of the log earnings residual distribution from 1970 through 2008 for our sample, while Figure 3 displays changes in the commonly reported ratio of log earnings residuals at the 90th percentile over residuals at the 10th percentile (the '90-10 ratio'), as well as analogous figures for the 90-50 and 50-10 ratios. This figure reports changes in these ratios from 1970 to the reported year. The 90-10 ratio exhibits a modest increase over the 1970s, a sharp increase in the early 1980s, followed by ten years of modest decline from 1985-95, and then an increase from 1995 through 2008. Over the full time period, the 90-10 log earnings residual ratio increased more than 0.5, with nearly two-thirds of that increase coming between 1980 and 1985. The figure further shows that changes in inequality were quite different at the top and bottom of the residual distribution. While the 90-50 ratio shows a steady increase of about 0.25 over the 38 years of our sample, changes in the 50-10 ratio largely mirror changes in the 90-10 ratio. Thus, the sharp increase in residual inequality in the early 1980s is largely driven by sharp declines in log earnings at the bottom of the distribution. Similarly, declines in residual inequality over the late 1980s and early 1990s come from increases in earnings at the bottom relative to the middle of the residual distribution.

5 Minimum Distance Estimation

We use minimum distance estimation and the residual moments described above to estimate the model for log earnings residuals for men using the PSID. Because some age cells have few observations when calculating residual variances and covariances (or higher moments), we aggregate within three broad age groupings corresponding to ages 30-39, 40-49 and 50-59. Specifically, for variances/covariances we use the following moments:

$$\frac{1}{n_{A,t,l}} \sum_{i:a\in A} W_{i,a,t} W_{i,a-l,t-l} \overset{p}{\to} E[W_{i,a,t} W_{i,a-l,t-l}|a \in A, t, l]$$

$$= \sum_{a\in A} \omega_{a,t,l} E[W_{i,a,t} W_{i,a-l,t-l}|a, t, l]$$

Figure 2: Selected Log Earnings Residual Quantiles from 1970 to 2008
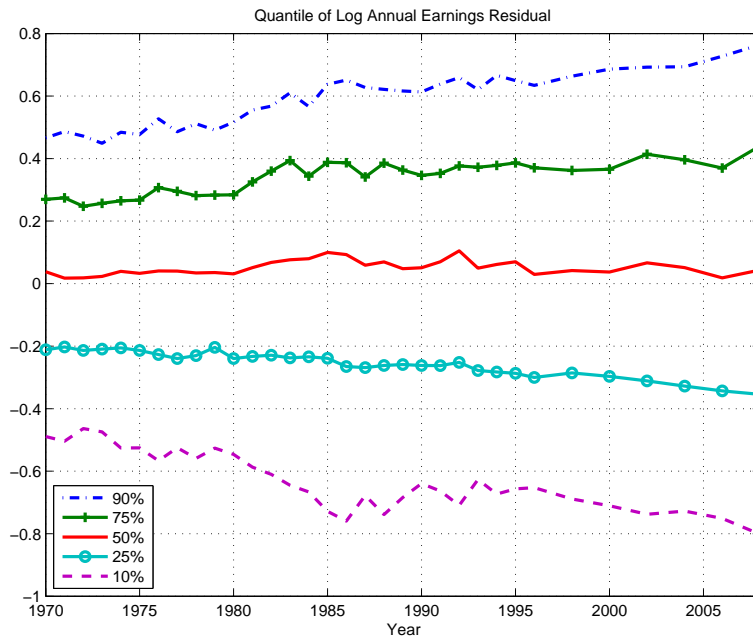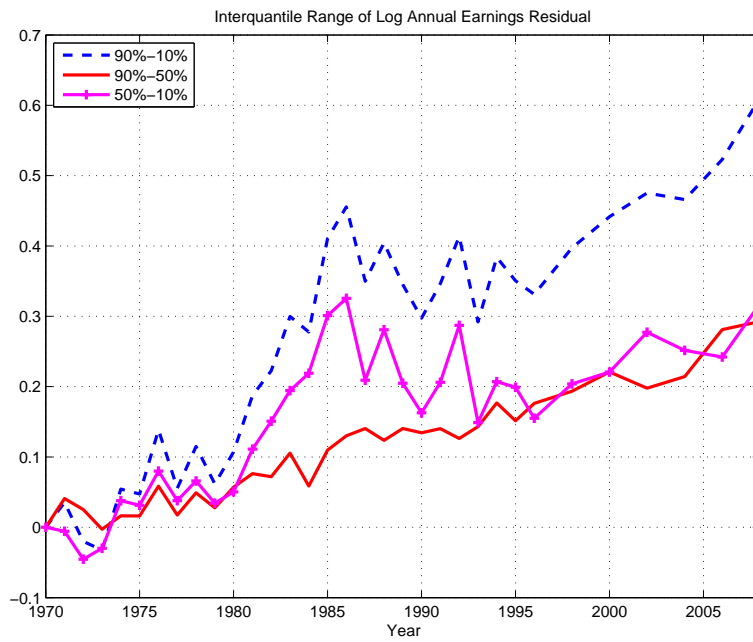

Figure 3: Changes in 90-10, 90-50, and 50-10 Ratios for Log Earnings Residuals from 1970 through 2008

where A reflects one of our three age categories, $n_{A,t,l}$ is the total number of observations used in calculating this moment, and $\omega_{a,t,l}$ is the fraction of observations used in calculating this moment that are of age a in period t. We weight each moment by the share of observations used for that sample moment (i.e. $n_{A,t,l}/\sum_A \sum_t \sum_l n_{A,t,l}$). Higher moments are treated analogously.

We impose a few restrictions to reduce the dimension of the problem given our modest sample sizes. First, we assume that the $MA(q)$ stochastic process remains the same over our sample period, so $\beta_{j,t} = \beta_j$ for all $j = 1, ..., q$ and $t = 2, ..., T$. Second, we assume that $\sigma^2_{\eta_\tau} = \sigma^2_{\eta_0}$ and $\sigma^2_{\xi_\tau} = \sigma^2_{\xi_0}$ for all τ years prior to our sample period. These assumptions are useful in accounting for differences in residual variances and covariances across cohorts observed in our initial survey year without substantially increasing the number of parameters to be estimated.[16]

We decompose the variance of log residual earnings into three components:

1. pricing of unobserved skills: $Var[\mu_t(\theta)]$

2. permanent shocks: $\sigma^2_{\eta_t}$ and $\sigma^2_{\kappa_t} = \sum_{j=0}^{a-1} \sigma^2_{\eta_{t-j}}$

3. transitory shocks: $\sigma^2_{\xi_t}$ and $\sigma^2_{\nu_t} = \sigma^2_{\xi_t} + \sum_{j=1}^{q} \beta^2_{jt} \sigma^2_{\xi_{t-j}}$.

We also discuss the evolution of different quantiles in distribution of $\mu_t(\theta)$ over time when we consider cubic $\mu_t(\cdot)$ pricing functions.

5.1   Linear $\mu_t(\cdot)$

We begin with the case of linear $\mu_t(\cdot)$. We do not need to make any distributional assumptions on θ or the permanent and transitory shocks to decompose the residual variances. Table 1 reports the minimum values of the objective function and key parameter estimates determining the process for ν_t under different assumptions about $\mu_t(\cdot)$ and the stochastic process for ν_t. The first three columns report results when $\mu_t(\theta)$ is restricted to be time invariant. This is equivalent to including individual fixed effects, as is standard in the PSID-based literature. The remaining four columns allow $\mu_t(\cdot)$ to vary over time. A few lessons emerge from this table. First, comparing columns 1-3 with their counterparts in columns 4-6 shows that allowing for changes in the pricing of unobserved skills significantly improves the fit to the data. This is generally true for any $MA(q)$ specification

---

[16]More generally, we could estimate separate variances for these shocks going back to the year of labor market entry for the oldest cohort in our initial sample period. Without observing earnings in those earlier years, these variances would need to be identified from cross-cohort differences in the variances and covariances we do observe. We have explored different assumptions about these pre-survey year variances (e.g. linear time trends); however, the results we discuss are robust across all assumptions.

Table 1: Estimates Assuming $\nu_t \sim MA(q)$ using Variances/Covariances (Linear $\mu_t(\cdot)$)

| | Constant $\mu_t(\cdot)$ | | | Time-Varying $\mu_t(\cdot)$ | | | |
	$MA(1)$	$MA(2)$	$MA(3)$	$MA(1)$	$MA(2)$	$MA(3)$	$MA(5)$
Min. Obj. Function	194.89	179.74	168.27	130.73	124.16	121.10	116.93
β_1	0.361	0.320	0.326	0.297	0.281	0.288	0.299
	(0.032)	(0.023)	(0.025)	(0.030)	(0.024)	(0.025)	(0.024)
β_2	·	0.257	0.222	·	0.186	0.172	0.194
	·	(0.029)	(0.023)	·	(0.022)	(0.019)	(0.018)
β_3	·	·	0.246	·	·	0.141	0.137
	·	·	(0.035)	·	·	(0.026)	(0.024)
β_4	·	·	·	·	·	·	0.126
	·	·	·	·	·	·	(0.020)
β_5	·	·	·	·	·	·	0.084
	·	·	·	·	·	·	(0.023)

for ν_t.[17] We strongly reject the restriction of constant $\mu_t(\cdot)$ functions at 5% significance levels. Second, the stochastic process for ν_t has a modest degree of persistence. We can reject $q = 1$ in favor of $q = 2$; however, we cannot reject that $q = 2$ and $q = 3$ fit equally well at 5% significance levels.[18] In addition to these cases, we report results for an $MA(5)$ for comparison.[19]

Unless otherwise noted, the rest of our analysis focuses on the case with time-varying $\mu_t(\cdot)$ and an $MA(3)$ process for ν_t; however, other $MA(q)$ processes yield very similar conclusions as we show below. Figure 4 reports the estimated variances (and standard errors) for $\mu_t(\theta)$, η_t, and ξ_t over time.[20] Figure 5 decomposes the total residual variance into its three components: unobserved skills $\mu_t(\theta)$, permanent shocks κ_t, and transitory shocks ν_t. All three components are important for understanding the evolution of earnings inequality in the PSID; however, they contribute in very different ways over time. Initially quite low, the variance of returns to unobserved ability/skills rises over the 1970s and early 1980s, then falls back to its original level by the late 1990s. It remains fairly constant thereafter. The variance of the transitory component rises sharply in the early 1980s, then fluctuates up and down for about ten years before it stabilizes in the late 1990s.

---

[17]As shown in Appendix B, allowing for time-varying $\mu_t(\cdot)$ functions also significantly improves the fit when ν_t is assumed to be $ARMA(1,1)$.

[18]These tests are based on a comparison of the minimized objective functions (reported in the first row of the table), which are distributed $\chi^2(1)$.

[19]We cannot reject an $MA(5)$ in favor of an $ARMA(1,5)$ at a 5% significance level.

[20]In some years, the variance of η_t is estimated to be zero; we do not report its standard errors for these years. As shown in Section 2, distributions for transitory and permanent shocks are not identified for the last few years of our panel; however, $\mu_t(\cdot)$ is identified for all periods. Our figures report variances for permanent and transitory components through 2002 and variances for $\mu_t(\theta)$ through 2008.

The variance of the permanent component declines slightly over the 1970s, then rises continuously and at roughly the same rate over the rest of the sample period. As a share of the total variance in log earnings residuals, the transitory component plays the largest role until the mid-1990s, after which the permanent shocks dominate. Inequality due to variation in the returns to unobserved skills reaches its peak of more than 40% of the total residual variance around 1980.

Unlike much of the PSID-based literature on earnings dynamics, our estimates suggest that the variance of permanent shocks increased much more than the variance of transitory shocks over the entire sample period. While there is a strong secular increase in the variance of permanent shocks starting in 1980, the variance of transitory shocks is much more modest and episodic, with nearly all of the lasting increase occurring in the early 1980s. These patterns imply quite different patterns for the total variance of log earnings residuals compared to the variance of returns to unobserved skills. While the total variance mainly increases in the early 1980s and the late 1990s and 2000s, variation in the returns to unobserved skills increases smoothly for fifteen years before declining for the next fifteen years. Variation in the returns to skill is remarkably stable from 1995 onward, in sharp contrast to the rising total residual variance.

In Figure 6, we compare our baseline model with the more standard case assuming $\mu_t(\cdot)$ is time-invariant. Here, the variance decomposition combines the component related to unobserved skills with the permanent shock component in our baseline model. The standard model under-predicts the impact of transitory shocks in the 1970s but over-predicts their importance in the 1980s. When variation in skill prices is ignored, the variance of transitory shocks appears to increase substantially more in the early 1980s than our baseline model suggests.

Figure 7 decomposes the residual variance into its three main components assuming $MA(1)$ or $MA(5)$ processes for ν_t. The dynamics and relative importance of each component are quite similar regardless of our assumptions about ν_t. However, the $MA(1)$ model slightly over-predicts the relative importance of permanent shocks, since some of the persistence attributed to 'transitory' shocks in the more general $MA(5)$ model effectively gets allocated to the permanent shock in the $MA(1)$ specification. Estimated patterns for the variance in unobserved skills prices are remarkably robust to assumptions about ν_t.

In Appendix B, we estimate the model assuming ν_t follows an $ARMA(1,1)$ process as is often assumed in the earnings dynamics literature. The variance decomposition is nearly identical to that of the $MA(3)$ and $MA(5)$ models (see Figure 13).[21]

---

[21]Compared to our $MA(3)$ baseline specification, the $ARMA(1,1)$ specification fits the data slightly better. However, we cannot reject that adding an autoregressive component to the $MA(5)$ model improves the fit (at the 5%

Figure 4: Variance of $\mu_t(\theta)$ η_t and ξ_t $(MA(3)$ model$)$

Variance of $\mu_t(\theta)$

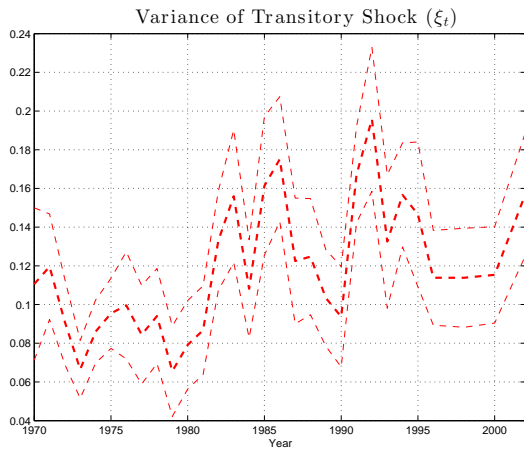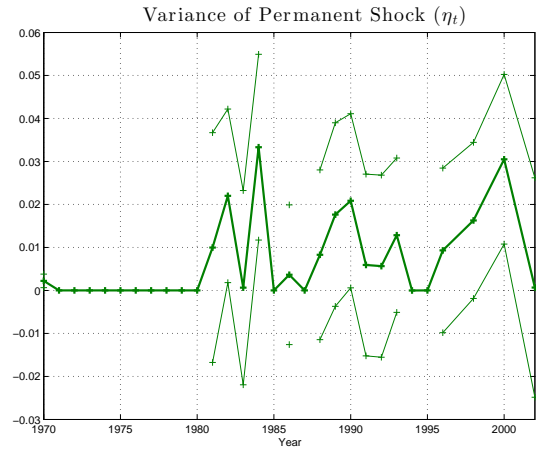Variance of Permanent Shock (η_t)

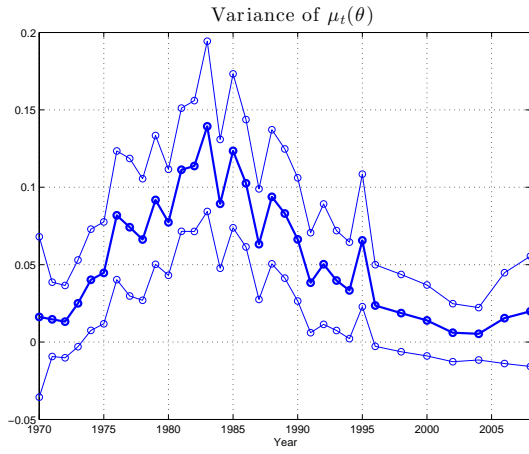Variance of Transitory Shock (ξ_t)
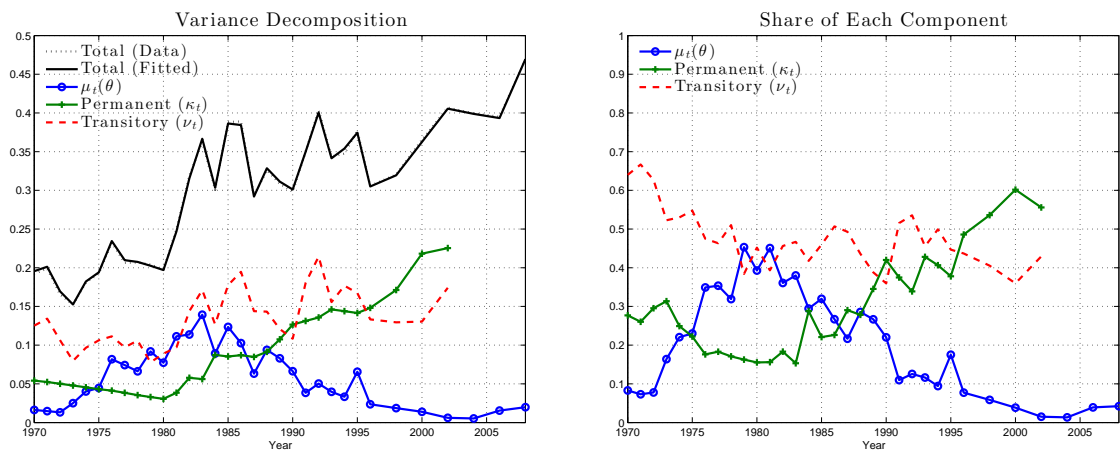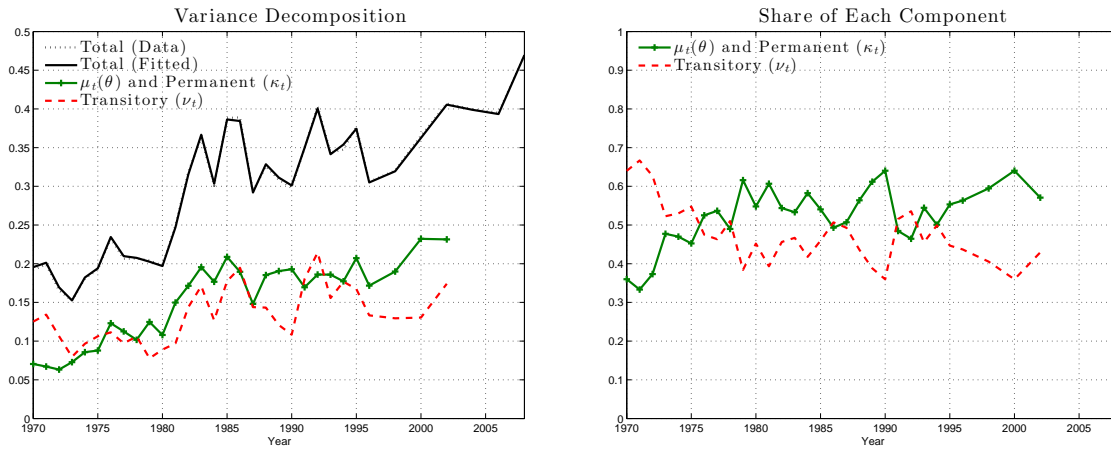
Figure 5: Variance Decomposition ($MA(3)$ model)

Figure 6: Comparison of time-varying and time-invariant $\mu_t(\theta)$ ($MA(3)$ model)

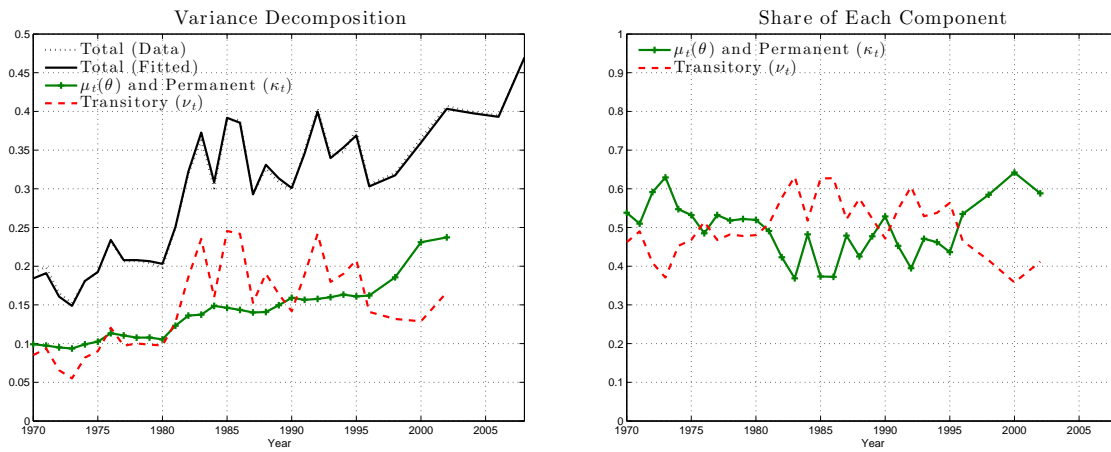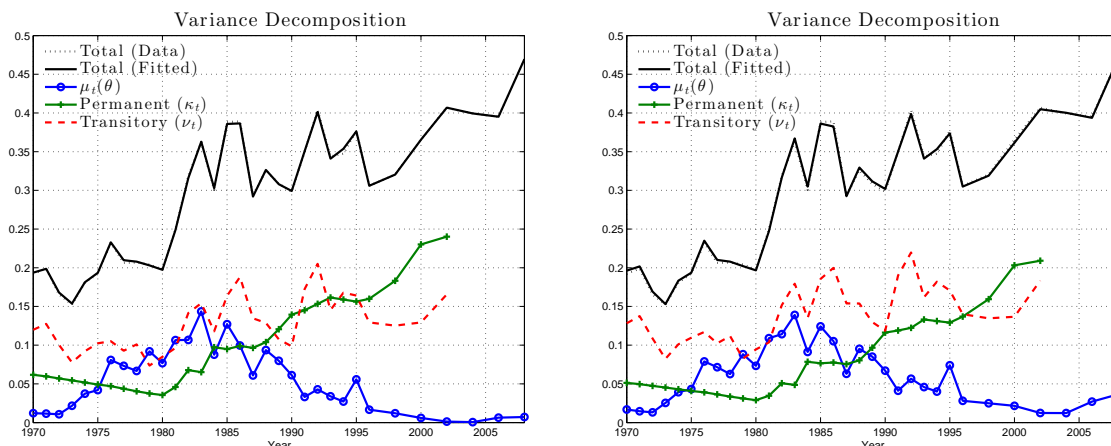a. Time-Varying $\mu_t(\theta)$

b. Time-Invariant $\mu(\theta)$

Figure 7: Variance Decomposition Assuming ν_t is $MA(1)$ or $MA(5)$

a. $MA(1)$ b. $MA(5)$

5.2   Cubic $\mu_t(\cdot)$ and Third Order Moments

We next estimate the $MA(3)$ model assuming $\mu_t(\theta)$ is a cubic function (normalizing $\mu_{1985}(\theta) = \theta$). We do not impose monotonicity on the $\mu_t(\cdot)$ functions; however, the results are quite similar if we do. In addition to variances and covariances, we also incorporate third-order residual moments in estimation.[22] This aids in identification of $\mu_t(\cdot)$ pricing functions and allows for estimation of third-order moments for permanent and transitory shocks.

In this case, our third-order moment conditions contain moments of θ up to $E(\theta^9)$. While we could estimate these higher moments directly along with all other parameters of the model, we instead assume that f_θ is a mixture of two normal distributions. Figure 8 shows the estimated distribution for θ.[23] Figure 9 performs the same type of variance decomposition as above. The results are quite similar to those assuming linear $\mu_t(\cdot)$ functions.

---

significance level).

[22]Specifically, we include all $E[W_{i,a,t}, W_{i,a-l,t-l}, W_{i,a-l-k,t-l-k}]$ moments along with all variances/covariances. In aggregating across cohorts, we calculate these third-order moments in the same we calculate variance/covariance terms.

[23]The first mixture component has a mean of .013 and standard deviation of 0.366, while the second has a mean of -1.290 and standard deviation of 1.870. The mixing probability places a weight of 99% on the first distribution and 1% on the second.

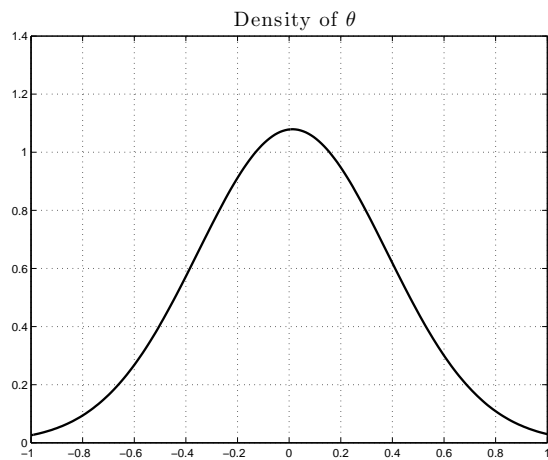Figure 8: Distribution of θ (Cubic $\mu_t(\cdot)$ functions)

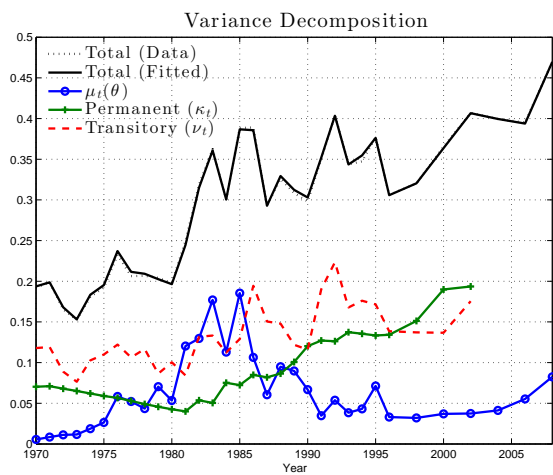Figure 9: Variance Decomposition (Cubic $\mu_t(\cdot)$ functions)

25

Figure 10 shows the evolution of estimated $\mu_t(\cdot)$ pricing functions for each decade. These functions are quite flat in the early 1970s, consistent with very low variance of $\mu_t(\theta)$ in that period. The increased importance of unobserved skills throughout the 1970s and early 1980s is reflected in the steepening of the $\mu_t(\theta)$ functions over this period. This is followed by declining inequality and a flattening in the $\mu_t(\theta)$ pricing functions over the late 1980s and early 1990s. Beginning in the mid-1990s, the $\mu_t(\theta)$ functions start to flatten at the bottom of the θ distribution, such that there is little difference in the reward to skill at the low end. The last few $\mu_t(\theta)$ functions actually appear to decline slightly in θ for very low values.[24] At the same time, the $\mu_t(\theta)$ functions are quite stable or even steepening slightly at the top of the distribution.

These patterns are more simply summarized in Figure 11, which shows the evolution of selected quantiles and the 90-10, 90-50, and 50-10 ratios for the distribution for $\mu_t(\theta)$ (the latter are relative to their 1970 values). The 90-10 ratio follows a similar pattern to that observed for the variance of $\mu_t(\theta)$ reported in Figure 9. The 50-10 ratio evolves much like the 90-10 ratio, increasing from 1970-1985, then falling fairly systematically ever since (except for a brief but sharp increase in the early 1990s). The 90-50 ratio shows a similar pattern through the mid-1990s, rising and falling over that period. Interestingly, while the 50-10 ratio falls rapidly over the late 1990s and early 2000s, the 90-50 ratio is relatively flat over that period. Since the mid-1990s, unobserved skill prices have become more compressed over the bottom of the distribution while they have remained stable at the top. These patterns differ markedly from those observed for total log earnings residuals as reported in Figures 2 and 3.

The patterns for $\mu_t(\cdot)$ imply an increasing (more positive) skewness over the 1990s. Figure 12 shows the skewness of total log earnings residuals along with the skewness for the permanent and transitory components of earnings over time. All are typically negatively skewed, in contrast to the skewness for $\mu_t(\theta)$, which is positive in all years except a few in the mid-1980s. The skewness of permanent shocks is generally declining except for a dramatic one-year jump in the early 1980s. Over the entire period, the skewness goes from slightly less than zero to around -2. There is no obvious trend to the skewness of transitory shocks, which hovers between -2 and -3 in most years.

6   Conclusions

Studies that estimate the changing role of unobserved skills generally abstract from the changing dynamics of earnings shocks, attributing changes in log earnings/wage residual distributions to the
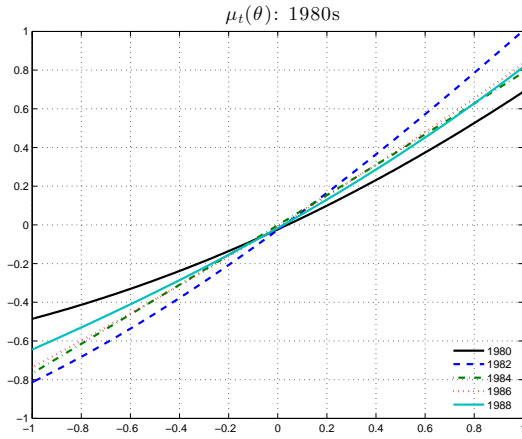
---

[24]It should be noted that standard errors are sizeable for $\mu_t(\theta)$ at very low and high values of θ, especially in the last few years.

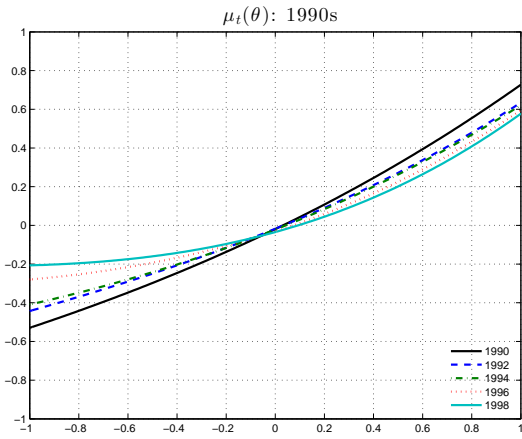Figure 10: Estimated $\mu_t(\theta)$ functions

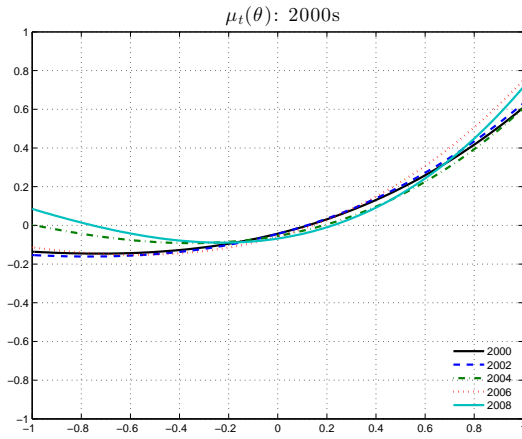a. 1970s
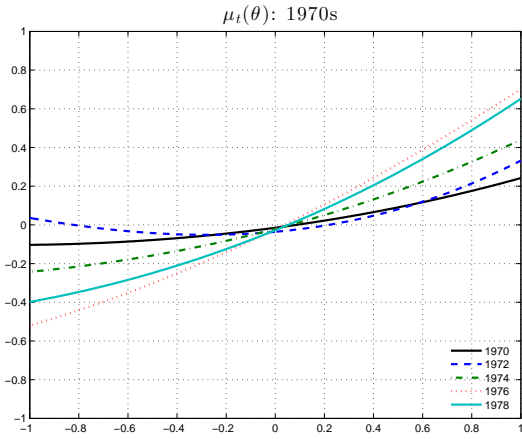
b. 1980s

c. 1990s

d. 2000s

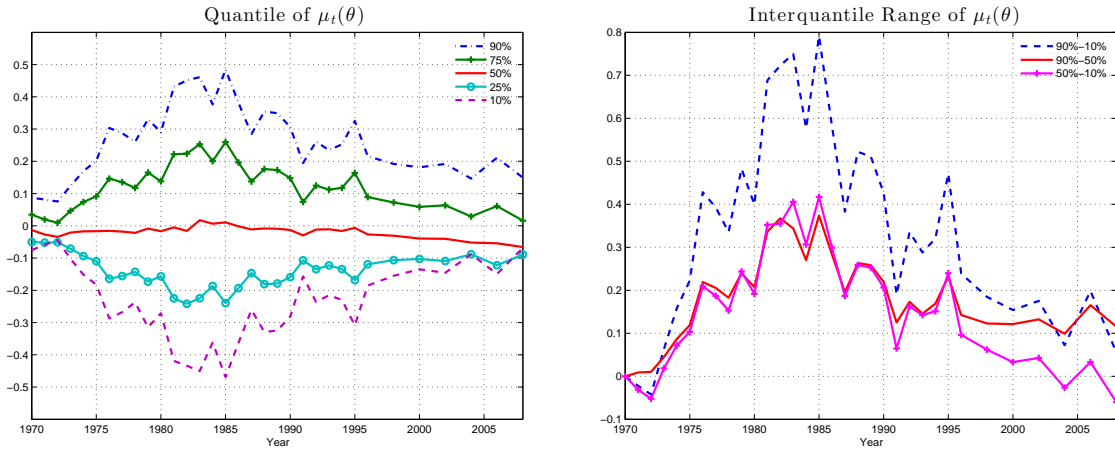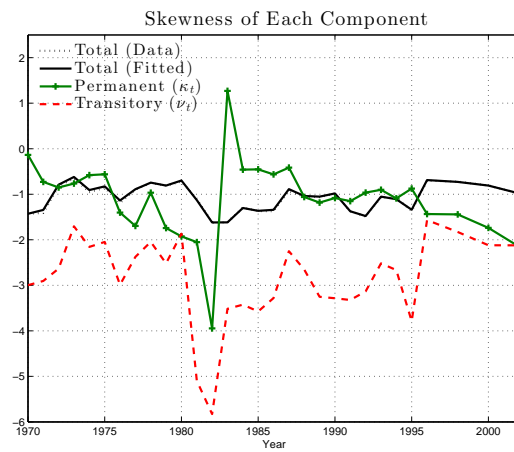Figure 11: Evolution of $\mu_t(\theta)$ distribution (Cubic $\mu_t(\cdot)$ functions)

Figure 12: Skewness of Each Component over Time (Cubic $\mu_t(\cdot)$ functions)

evolution of unobserved skill pricing over time. A separate literature in labor and macroeconomics estimates important changes in the variance of transitory and permanent shocks over the past few decades; however, this literature typically neglects changes in unobserved skill prices.

We show that the distribution of unobserved skills and the evolution of skill pricing functions can be separately identified from changing distributions of idiosyncratic permanent and transitory shocks using panel data. Specifically, a panel of length $T \geq 6 + 3q$ is needed for full nonparametric identification in the presence of permanent Martingale shocks and transitory shocks that follow an $MA(q)$ process. We then discuss a moment-based approach to estimating the distribution of unobserved skills, changes in unobserved skill pricing functions, and the changing nature of permanent and transitory shocks over time.

Using panel data from the PSID on male earnings in the U.S. from 1970-2008, we show that accounting for time-varying unobserved skill prices is important for explaining the variances and autocovariances of log earnings residuals over this period. With our estimates, we decompose the variance of log earnings residuals over time into three components: the pricing of unobserved skills, permanent and transitory (but persistent) shocks. Our results suggest that there was a sizeable increase in the returns to unobserved skill over the 1970s and early 1980s. But, this pattern reversed in the late 1980s and 1990s, with the pricing of unobserved skills largely returning to what it was in 1970. From 1995 onward, unobserved skill prices were much more stable (especially at the top of the skill distribution). These patterns contrast sharply with time trends for the variance of log earnings residuals, which rose sharply in the early 1980s, remained relatively stable over the late 1980s and early 1990s, and then began rising again. The differences are due to important changes in the variance of permanent and transitory shocks. In particular, the variance of earnings residuals rose much more sharply in the early 1980s than the variance of unobserved skill prices due to sizeable increases in the variance of both transitory and permanent shocks. While the variance of transitory skills fluctuated up and down afterwards (without any obvious long-run trend), the variance of permanent shocks continued to rise at a steady pace through the end of our sample. Over the late 1980s and early 1990s, this increase largely offset declines in the price of unobserved skills, leaving the total residual variance relatively unchanged over a ten to fifteen year period. When the pricing of unobserved skills stabilized (at least at the top of the skill distribution) in the mid-1990s, residual inequality rose along with the variance of permanent shocks.

Our estimates of flexible skill pricing functions allow us to identify changes in the returns to unobserved skill at different points in the distribution. Over the 1970s, 1980s and early 1990s,

the returns to unobserved skill rose and fell by similar amounts throughout the skill distribution; however, this was no longer true beginning in the mid-1990s. From 1995 on, we estimate very little change in unobserved skill prices over the top half of the distribution, while we estimate significant declines in the value of unobserved skill over the bottom half. By the mid-2000s, skill pricing functions were essentially flat over the entire bottom half of the distribution. The latter is broadly consistent with the skill polarization phenomenon emphasized by Acemoglu and Autor (2011) and Autor and Dorn (2012); however, we find no evidence of an increase in the returns to skill at the top of the distribution (as one would infer from looking at residuals themselves). An important lesson from these findings is that changes in the distribution of log earnings residuals are not particularly informative about the evolution of unobserved skill pricing functions, especially in recent decades.

Appendix A   Technical Results

A.1   Proof of Lemma 1

Without loss of generality, we set $T = 3$. Assumption 1 implies Assumptions 1–5 in Hu and Schennach (2008). Especially, the strict monotonicity of $\mu_3(\cdot)$ implies their Assumption 4, which is, for any $\bar{\theta} \neq \tilde{\theta}$, the set $\{w : f_{W_3|\theta}(w_3|\bar{\theta}) \neq f_{W_3|\theta}(w_3|\tilde{\theta})\}$ has a positive probability measure. Therefore, we can apply their Theorem 1 by setting $x^* = \theta$, $x = W_1$, $z = W_2$, and $y = W_3$. The same strategy is also adopted in Cunha, Heckman, and Schennach (2010) for identifying $f_\theta(\cdot)$. Given additive separability in the model, we show that $f_{\varepsilon_t}(\cdot)$ and $\mu_t(\cdot)$ are also identified.

Theorem 1 in Hu and Schennach (2008) implies that when we have the joint density of (W_1, W_2, W_3), the equation

$$f_{W_3,W_1,W_2}(w_3, w_1, w_2) = \int_{\Theta} f_{W_1|\theta}(w_1|\theta) f_{W_3,\theta}(w_3, \theta) f_{W_2|\theta}(w_2|\theta)\, d\theta \quad \text{for all}\ \ w_t \in \mathcal{W}_t \qquad (10)$$

admits a unique solution $(f_{W_1|\theta}, f_{W_3,\theta}, f_{W_2|\theta})$. Since we already know the marginal distribution f_{W_3}, we can first identify f_θ by integrating $f_{W_3,\theta}$ over W_3. Next, the functions $\mu_t(\cdot)$ for $t = 2, 3$ are identified from conditional densities $f_{W_t|\theta}$ since we know that $E[W_t|\theta] = \mu_t(\theta)$ from $E[\varepsilon_t|\theta] = E[\varepsilon_t] = 0$. Finally, f_{ε_t} is identified from $f_{\varepsilon_t}(\varepsilon) = f_{\varepsilon_t|\theta}(\varepsilon|\theta) = f_{W_t|\theta}(\mu_t(\theta) + \varepsilon)$ since both $f_{W_t|\theta}$ and $\mu_t(\cdot)$ are already known. $\square$

A.2   Proof of Theorem 1

We prove this identification result in three steps.

Step 1: Identification of $f_\theta(\cdot)$ and $\mu_t(\cdot)$ for all t.

In this step, we jointly consider distributions of $(W_t, \Delta W_{t+3}, \Delta W_{t+6})$ for $t = 1, 2, 3$. These triplets are mutually independent under Assumption 2 (ii). Begin with the following subset of equations:

$$W_1 = \theta + \varepsilon_1 = \theta + \eta_1 + \nu_1$$

$$\Delta W_4 = \Delta\mu_4(\theta) + \Delta\varepsilon_4 = \Delta\mu_4(\theta) + \eta_4 + \Delta\nu_4$$

$$\Delta W_7 = \Delta\mu_7(\theta) + \Delta\varepsilon_7 = \Delta\mu_7(\theta) + \eta_7 + \Delta\nu_7.$$

Different from Lemma 1, Assumption 4 in Hu and Schennach (2008) described above may not hold since $\Delta\mu_7(\cdot)$ can be a non-monotone function. We consider two cases depending on the functional form of $\Delta\mu_7(\cdot)$.

First, consider the case of $\Delta\mu_7(\theta) = a + b\ln(e^{c\theta} + d)$ for $a, b, c(\neq 0), d \in \mathbb{R}$. Then, for any $\bar{\theta} \neq \tilde{\theta}$, $\Delta\mu_7(\bar{\theta}) \neq \Delta\mu_7(\tilde{\theta})$ because of the strict monotonicity of the exponential function. Therefore,

$f_{\Delta W_7|\theta}(w|\bar{\theta}) \neq f_{\Delta W_7|\theta}(w|\tilde{\theta})$ for some w with positive probability. We can now apply Lemma 1 to identify f_θ, $\Delta\mu_4(\cdot)$ and $\Delta\mu_7(\cdot)$ as before.

Second, consider the case of $\Delta\mu_7(\theta) \neq a + b\ln(e^{c\theta} + d)$. Then, we first apply Theorem 1 in Schennach and Hu (2013) to the following pair of equations:

$$W_1 = \theta + \varepsilon_1 \tag{11}$$

$$\Delta W_7 = \Delta\mu_7(\theta) + \Delta\varepsilon_7. \tag{12}$$

Notice that Assumptions 1–6 in their paper are implied by Assumption 2 (i), (ii), (iv), and (v). Therefore, we can identify the function $\Delta\mu_7(\cdot)$, and the densities $f_\theta(\cdot), f_{\varepsilon_1}(\cdot), f_{\Delta\varepsilon_7}(\cdot)$, and it only remains to identify $\Delta\mu_4(\cdot)$. To do that, we need some additional notation. Let $L_{A|B}$ be a linear operator defined as

$$L_{A|B} : \mathcal{G}(\mathcal{B}) \mapsto \mathcal{G}(\mathcal{A}) \text{ with } [L_{A|B}g](\cdot) \equiv \int f_{A|B}(\cdot|b)g(b)db \tag{13}$$

where $\mathcal{A}$ is support of a random variable A, and $\mathcal{G}(\mathcal{A})$ is the space of all bounded and absolutely integrable functions supported on $\mathcal{A}$. Simlarly, $\mathcal{B}$ and $\mathcal{G}(\mathcal{B})$ are defined. For any given $\Delta W_7 = w_7$, we also define

$$L_{\Delta W_7;W_1|\Delta W_4} : \mathcal{G}(\mathcal{W}_4) \mapsto \mathcal{G}(\mathcal{W}_1) \text{ with } \left[L_{\Delta W_7;W_1|\Delta W_4}g\right](\cdot) \equiv \int f_{\Delta W_7,W_1|\Delta W_4}(w_7,\cdot|w_4)g(w_4)dw_4 \tag{14}$$

$$\Lambda_{\Delta W_7;\theta} : \mathcal{G}(\Theta) \mapsto \mathcal{G}(\Theta) \text{ with } \left[\Lambda_{\Delta W_7;\theta}g\right](\cdot) \equiv f_{\Delta W_7|\theta}(w_7,\cdot)g(\cdot). \tag{15}$$

Using Assumption 2 (ii), we can rewrite the conditional density $f_{\Delta W_7,W_1|\Delta W_4}$ as follows

$$f_{\Delta W_7,W_1|\Delta W_4}(w_7,w_1|w_4) = \int f_{W_1|\theta}(w_1|\theta)f_{\Delta W_7|\theta}(w_7|\theta)f_{\theta|\Delta W_4}(\theta|w_4)d\theta, \tag{16}$$

which is equivalent to

$$L_{\Delta W_7;W_1|\Delta W_4} = L_{W_1|\theta}\Lambda_{\Delta W_7;\theta}L_{\theta|\Delta W_4}. \tag{17}$$

By integrating over w_7, we have

$$L_{W_1|\Delta W_4} = L_{W_1|\theta}L_{\theta|\Delta W_4} \tag{18}$$

$$L_{\theta|\Delta W_4} = L_{W_1|\theta}^{-1}L_{W_1|\Delta W_4}, \tag{19}$$

where Equation (19) is made possible from Assumption 2 (iii). Since we have already identified f_{ε_1} and ε_1 is independent of θ, the conditional density $f_{W_1|\theta}$ is identified from $f_{W_1|\theta}(w_1|\theta) = f_{\varepsilon_1}(w_1-\theta)$.

Therefore, we know both terms on the right hand side of Equation (19), and identify the density $f_{\theta|\Delta W_4}$. Applying the Bayes theorem with known f_θ and $f_{\Delta W_4}$, we can identify $f_{\Delta W_4|\theta}$. Finally, $\Delta\mu_4(\cdot)$ is recovered from $f_{\Delta W_4|\theta}$ and $E[\Delta\varepsilon_4|\theta] = E[\Delta\varepsilon_4] = 0$.

Next we 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 as $\Delta\mu_t(\theta) = g_t(\mu_2(\theta))$. We apply the same method described above to identify $f_{\theta_2}(\cdot)$, $g_5(\cdot)$, and $g_8(\cdot)$. Then, we can recover the function $\mu_2(\cdot)$ by $\mu_2(\theta) = F_{\theta_2}^{-1}(F_\theta(\theta))$. Once we identify $\mu_2(\cdot)$, $\Delta\mu_t(\cdot)$ for $t = 5, 8$ are identified from $\Delta\mu_t(\theta) = g_t(\mu_2(\theta))$. We apply the same argument to the set of equations composed of $(W_3, \Delta W_6, \Delta W_9)$ and identify $\mu_3(\cdot), \Delta\mu_6(\cdot)$, and $\Delta\mu_9(\cdot)$. Finally, we 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: Identification of $f_{\eta_t}(\cdot)$ and $f_{\nu_t}(\cdot)$ for $t = 1, \ldots, 7$.

Consider the following two equations:

$$W_1 = \theta + \varepsilon_1 = \theta + \eta_1 + \nu_1$$

$$W_3 = \mu_3(\theta) + \varepsilon_3 = \mu_3(\theta) + \eta_1 + \eta_2 + \eta_3 + \nu_3 \equiv \mu_3(\theta) + \eta_1 + \nu_3'$$

where $\nu_3' = \eta_2 + \eta_3 + \nu_3$. We rearrange these equations as follows

$$W_1 - \theta = \varepsilon_1 = \eta_1 + \nu_1$$

$$W_3 - \mu_3(\theta) = \varepsilon_3 = \eta_1 + \nu_3'.$$

We first show that the joint density of $(\varepsilon_1, \varepsilon_3)$ is identified. Note that

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

The second to the last equality exploits the independence between $(\varepsilon_1, \varepsilon_3)$ and θ. Since both $\phi_{W_1,W_3}(\tau_1, \tau_3)$ and $\phi_{\theta,\mu_3(\theta)}(\tau_1, \tau_3)$ are already identified, we can identify the joint density of $(\varepsilon_1, \varepsilon_3)$

from

$$\phi_{\varepsilon_1,\varepsilon_3}(\tau_1,\tau_3) = \frac{\phi_{W_1,W_3}(\tau_1,\tau_3)}{\phi_{\theta,\mu_3(\theta)}(\tau_1,\tau_3)}.$$

Next, η_1, ν_1, and ν'_3 are mutually independent. Therefore, we can identify $f_{\eta_1}(\cdot)$ and $f_{\nu_1}(\cdot)$, and $f_{\nu'_3}(\cdot)$ by applying Lemma 1 in Kotlarski (1967). Applying this argument to $(W_2, W_4), \ldots, (W_7, W_9)$ sequentially, we can identify $f_{\eta_t}(\cdot)$ and $f_{\nu_t}(\cdot)$ for all $t = 1, \ldots, 7$. However, for $t = 8$ and 9, we cannot decompose η_t from ν_t unless we have additional observations W_{10} and W_{11}.

Step 3: Identification of $f_{\xi_t}(\cdot)$ and β_t for $t = 1, \ldots, 7$.

Finally, we identify all components of the transitory shock, $\nu_t = \xi_t + \beta_t \xi_{t-1}$. Because of the normalization $\nu_1 = \xi_1$, the distribution $f_{\xi_1}(\cdot)$ is identified by $f_{\xi_1}(\cdot) = f_{\nu_1}(\cdot)$ in Step 2. Next, we identify β_2 from

$$Cov(W_1, W_2) = Cov\left(\theta, \mu_2(\theta)\right) + Var(\eta_1) + \beta_2 Var(\xi_1),$$

since we know all other terms except β_2. Therefore, unless $Var(\xi_1) = 0$, we can identify β_2. In the above equation, note that all cross moments between unobservables are zero because of the conditional mean zero assumption. For example,

$$Cov(\mu_2(\theta), \eta_1) = E[\mu_2(\theta)\eta_1] \tag{20}$$

$$= E_\theta\left[E[\mu_2(\theta)\eta_1|\theta]\right] \tag{21}$$

$$= E_\theta\left[\mu_2(\theta)E[\eta_1|\theta]\right] \tag{22}$$

$$= 0 \tag{23}$$

We next identify the distribution of ξ_2 using the standard deconvolution method (conditional on θ) as

$$\phi_{\xi_2}(\tau) = \frac{\phi_{\nu_2}(\tau)}{\phi_{\beta_2\xi_1}(\tau)}$$

where $\phi_{\nu_2}(\cdot)$ and $\phi_{\beta_2\xi_1}(\cdot)$ are identified. In the same way, we expand $Cov(W_t, W_{t+1})$ and identify β_{t+1} and $f_{\xi_{t+1}}(\cdot)$ sequentially for $t = 2, \ldots, 7$. Again, we cannot identify the components of ν_8 and ν_9 unless we have additional observations.

Combining results from Step 1–3, we establish the identification of $f_\theta(\cdot), \{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot), \beta_t\}_{t=1}^7$, and $\{\mu_t(\cdot)\}_{t=1}^9$. $\square$

Appendix B   Moments and Estimates when $\nu_{i,a,t}$ Follows an $ARMA(1,1)$ Process

In this appendix, we briefly describe residual moments when $\nu_{i,a,t}$ follows an $ARMA(1,1)$ stochastic process: $\nu_{i,a,t} = \rho\nu_{i,a-1,t-1} + \xi_{i,a,t} + \beta_t\xi_{i,a-1,t-1}$. All other assumptions and notation are the same as in Section 3, including the assumption that individuals begin receiving shocks $\eta_{i,a,t}$ and $\xi_{i,a,t}$ when they enter the labor market at age $a = 1$ (i.e. $\kappa_{i,0,t} = \nu_{i,0,t} = 0$ and $\eta_{i,a,t} = \xi_{i,a,t} = 0$ for all $a \leq 0$).

Due to mutual independence of $\xi_{i,a,t}$ across time,

$$E[\nu_{i,a,t}^k] = \sigma_{\xi_t}^k + \sum_{j=0}^{a-2} \rho^{kj}(\rho + \beta_{t-j})^k \sigma_{\xi_{t-j-1}}^k.$$

For $k = 2$, this expression defines the variance of the 'transitory' component in our variance decompositions. Other variance components are unchanged, so

$$E[W_{i,a,t}^2|a,t] = \sum_{j=0}^{p}\sum_{j'=0}^{p} m_{j,t}m_{j',t}E[\theta^{j+j'}] + \sum_{j=0}^{a-1}\sigma_{\eta_{t-j}}^2 + \sigma_{\xi_t}^2 + \sum_{j=0}^{a-2}\rho^{2j}(\rho + \beta_{t-j})^2\sigma_{\xi_{t-j-1}}^2$$

and

$$E[W_{i,a,t}W_{i,a+l,t+l}|a,t,l] = \sum_{j=0}^{p}\sum_{j'=0}^{p} m_{j,t}m_{j',t+l}E[\theta^{j+j'}] + \sum_{j=0}^{a-1}\sigma_{\eta_{t-j}}^2 + \rho^{l-1}(\rho+\beta_{t+1})\sigma_{\xi_t}^2 + \rho^l\sum_{j=0}^{a-2}\rho^{2j}(\rho+\beta_{t-j})^2\sigma_{\xi_{t-j-1}}^2$$

for $l \geq 1$.

Using variance/covariance moments and assuming linear $\mu_t(\cdot)$, we estimate this model assuming $\beta_t = \beta$ for all t. Results from this model are reported in Table 2 along with analogous results assuming time invariant pricing functions (i.e. $\mu_t(\theta) = \theta$ for all t).[25] Figure 13 shows the variance decomposition associated with these estimates (for time-varying $\mu_t(\cdot)$), which is quite similar to that for the $MA(5)$ model as reported in Figure 4.

---

[25]Generalizing the process for $\nu_{i,a,t}$ to an $ARMA(1,5)$ produces a negligible improvement in fit over the $ARMA(1,1)$ with a minimized objective function of 114.13.

Table 2: Estimates for $\nu_t \sim ARMA(1,1)$ using Variances/Covariances (Linear $\mu_t(\cdot)$)

	$\mu_t(\cdot)$ constant	μ_t varying
Min. Obj. Function	144.2	114.4
ρ	0.861	0.804
	(0.017)	(0.034)
β	-0.529	-0.496
	(0.030)	(0.046)

Figure 13: Variance Decomposition Assuming $\nu_t \sim ARMA(1)$ (Linear $\mu_t(\cdot)$)

36
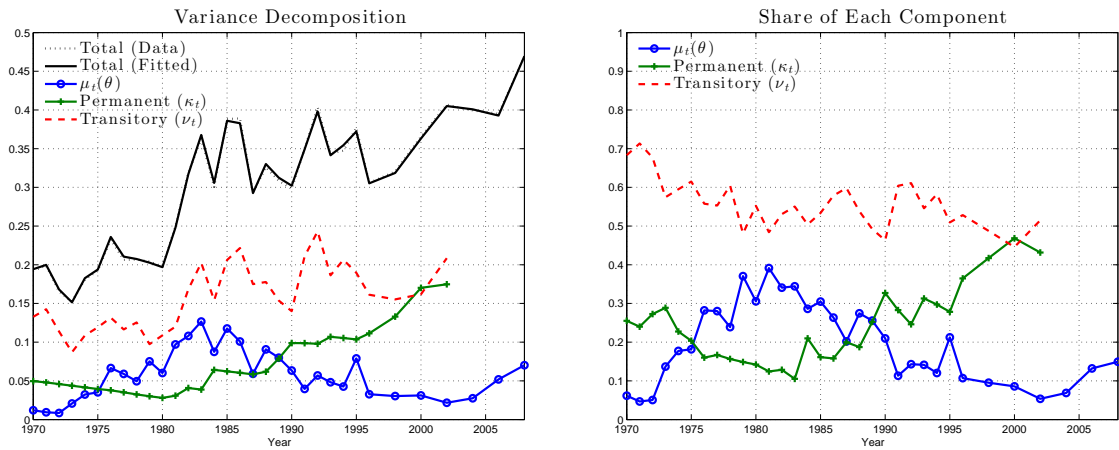
References

Abowd, J. M. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica 57*(2), 411–45.

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature 40*(1), 7–72.

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4, Chapter 12, pp. 1043–1171. Elsevier.

Aghion, P. (2002). Schumpeterian growth theory and the dynamics of income inequality. *Econbometrica 70*(3), 855–82.

Autor, D. and D. Dorn (2012). The growth of low skill service jobs and the polarization of the u.s. labor market. IZA Discussion Papers 7068, Institute for the Study of Labor (IZA).

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in u.s. wage inequality: Revising the revisionists. *The Review of Economics and Statistics 90*(2), 300–323.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2011, August). Tenure, experience, human capital and wages: A tractable equilibrium search model of wage dynamics. Working Paper.

Baker, M. (1997). Growth-rate heterogeneity and the covariance structure of life-cycle earnings. *Journal of Labor Economics 15*(2), 338–75.

Blundell, R., L. Pistaferri, and I. Preston (2008). Consumption inequality and partial insurance. *The American Economic Review 98*(5), pp. 1887–1921.

Blundell, R. and I. Preston (1998). Consumption inequality and income uncertainty. *The Quarterly Journal of Economics 113*(2), 603–640.

Bonhomme, S. and J.-M. Robin (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies 77*(2), 491–533.

Burdett, K., C. Carrillo-Tudela, and M. G. Coles (2011). Human capital accumulation and labor market equilibrium. *International Economic Review 52*(3), 657–677.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 30, pp. 1801–1863. Elsevier.

Card, D. and J. E. DiNardo (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics 20*(4), 733–783.

Cunha, F., J. Heckman, and S. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Flinn, C. J. (1986). Wages and job mobility of young workers. *Journal of Political Economy 94*(3), S88–S110.

Gottschalk, P. and R. Moffitt (1994). The growth of earnings instability in the u.s. labor market. *Brookings Papers on Economic Activity 25*(2), 217–272.

Gottschalk, P. and R. Moffitt (2009). The rising instability of u.s. earnings. *Journal of Economic Perspectives 23*(4), 3–24.

Gourinchas, P.-O. and J. A. Parker (2002). Consumption over the life cycle. *Econometrica 70*(1), 47–89.

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*(3), 687–712.

Haider, S. J. (2001). Earnings instability and earnings inequality of males in the united states: 1967–1991. *Journal of Labor Economics 19*(4), 799–836.

Hausman, J. A., H. Ichimura, W. K. Newey, and J. L. Powell (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics 50*(3), 273–295.

Heathcote, J., F. Perri, and G. L. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the united states: 1967-2006. *Review of Economic Dynamics 13*(1), 15–51.

Heathcote, J., K. Storesletten, and G. L. Violante (2010). The macroeconomic implications of rising wage inequality in the united states. *Journal of Political Economy 118*(4), 681–722.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, Chapter 7, pp. 307–458. Elsevier.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2008). Earnings functions and rates of return. *Journal of Human Capital 2*(1), 1–31.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy 101*(3), 410–42.

Katz, L. F. and D. H. Autor (1999). Changes in the wage structure and earnings inequality. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 26, pp. 1463–1555. Elsevier.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics 20*(1), 69–76.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review 96*(3), 461–498.

Lillard, L. A. and Y. Weiss (1979). Components of variation in panel earnings data: American scientists, 1960-70. *Econometrica 47*(2), 437–54.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics 18*(1), 83–114.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*(1), 1–32.

Moffitt, R. A. and P. Gottschalk (2002). Trends in the transitory variance of earnings in the united states. *Economic Journal 112*(478), C68–C73.

Moffitt, R. A. and P. Gottschalk (2012). Trends in the transitory variance of male earnings: Methods and evidence. *Journal of Human Resources 47*(1), 204–236.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Postel-Vinay, F. and H. Turon (2010). On-the-job search, productivity shocks, and the individual earnings process. *International Economic Review 51*(3), 599–629.

Schennach, S. and Y. Hu (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association 108*(501), 177–186.

Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica 72*(1), 33–75.

Taber, C. R. (2001). The rising college premium in the eighties: Return to college or return to unobserved ability? *Review of Economic Studies 68*(3), 665–691.

Yamaguchi, S. (2010). Job search, bargaining, and wage dynamics. *Journal of Labor Economics 28*(3), 595–631.