2014-2 Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks

Lance J. Lochner

Youngki Shin

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

by

Lance Lochner and Youngki Shin

Working Paper # 2014-2 **May 2014**

CIBC Working Paper Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

Understanding Earnings 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

May 12, 2014

Abstract

We consider a general framework to study the evolution of wage and earnings residuals that incorporates features highlighted by two influential but distinct literatures in economics: (i) unobserved skills with changing non-linear pricing functions and (ii) idiosyncratic shocks that follow a rich stochastic process. Specifically, we consider residuals for individual i in period t of the form: $W_{i,t} = \mu_t(\theta_i) + \varepsilon_{i,t}$, 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 with both permanent and transitory components.

We first provide nonparametric identification conditions for the distribution of unobserved skills, all $\mu_t(\cdot)$ skill pricing functions, and (nearly) all distributions for both permanent and $MA(q)$ transitory shocks. We then discuss identification and estimation using a moment-based approach, restricting $\mu_t(\cdot)$ to be polynomial functions. Using data on log earnings for men ages 30-59 in the PSID, we estimate the evolution of unobserved skill pricing functions and the distributions of unobserved skills, transitory, and permanent shocks from 1970 to 2008. We highlight five main findings: (i) The returns to unobserved skill 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. Since the mid-1990s, we observe some evidence of polarization: the returns to unobserved skill declined at the bottom of the distribution while they remained relatively constant over the top half. (ii) The variance of unobserved skill changed very little across most cohorts in our sample (those born between 1925 and 1955). (iii) 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. (iv) The variance of permanent shocks declined very slightly over the 1970s, then rose systematically through the end of our sample by 15 to 20 log points. The increase in this variance over the 1980s and 1990s was strongest for workers with low unobserved ability. (v) In most years, the distribution of $\mu_t(\theta)$ is positively skewed, while the distributions of permanent and (especially) transitory shocks are negatively skewed.

*We thank Youngmin Park for excellent research assistance, Peter Gottschalk and Robert Moffitt for sharing their PSID data creation programs, and Maria Ponomareva and Steve Haider for their useful suggestions. We also thank seminar and conference participants from Carnegie Mellon University, Florida State University, FGV - Sao Paulo School of Economics, Indiana University, Montana State University, Ohio State University, University College of London, University of Illinois–Chicago, University of Iowa, University of Wisconsin, Uppsala University, and the MSU/UM/UWO Labor Day Conference. Lochner acknowledges financial research support from the Social Sciences and Humanities Research Council of Canada.

1 Introduction

Sustained growth in economic 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 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 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.

Two large and influential empirical literatures study the rise in residual inequality through very different lenses. Beginning with Katz and Murphy (1992) and Juhn, Murphy, and Pierce (1993), a literature based primarily on data from the Current Population Survey (CPS) has equated changes in residual inequality with 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 value of unobserved skill in the labor market.[5] This interpretation has fostered the development

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

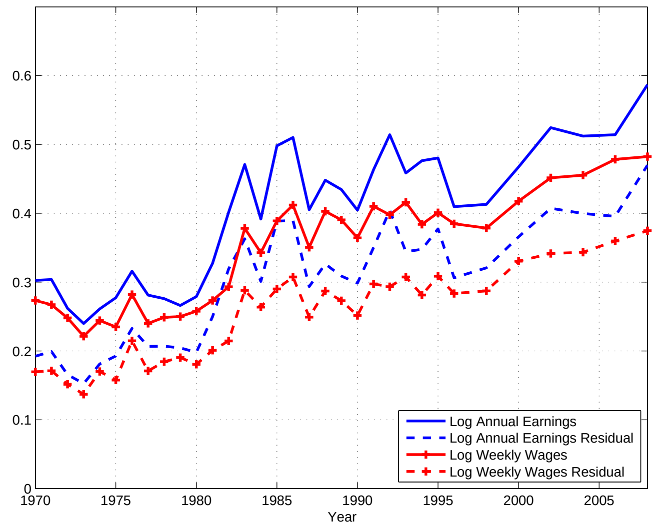
[2] Taber (2001) argues that increasing returns to unobserved skill in recent decades may be the main driver for the increase in measured returns to college, since individuals with higher unobserved skills are more likely to attend college.

[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). In a complementary literature, Heckman and Vytlacil (2001) and Murnane, Willett, and Levy (1995) directly estimate the changing effects of cognitive ability on wages over the 1980s and early 1990s using observable test scores.

[5] Lemieux (2006) argues that the composition of unobserved 'skills' has also changed over time as baby boomers have aged and the population has become more educated. Below, we explore differences in the distribution of unobserved ability across cohorts. Chay and Lee (2000) distinguish between changes in unobserved skill pricing and changes in the variance of transitory shocks over the 1980s by assuming that blacks and whites have different unobserved skill distributions but identical (time-varying) variances in transitory shocks.

Figure 1: Variance of Log Earnings and Weekly Wages and Residuals (Men Ages 30-59, PSID)



of new theories of skill-biased technical change as explanations for the increase in demand for skill (Acemoglu, 1999; Caselli, 1999; Galor and Moav, 2000; Gould, Moav, and Weinberg, 2001; Violante, 2002). Other studies have emphasized institutional changes like the declining minimum wage or de-unionization as causes for the declining wages paid to low-skilled workers (Card and DiNardo, 2002; Lemieux, 2006). Most recently, Autor, Levy, and Murnane (2003), 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.

A second, equally important literature studies the evolution of residual inequality with the aim of quantifying the relative importance of transitory and permanent shocks over time (Gottschalk and Moffitt, 1994; Blundell and Preston, 1998; Haider, 2001; Moffitt and Gottschalk, 2002; Meghir and Pistaferri, 2004; Gottschalk and Moffitt, 2009; Bonhomme and Robin, 2010; Heathcote, Storesletten, and Violante, 2010; Heathcote, Perri, and Violante, 2010; Moffitt and Gottschalk, 2012).[6] These studies examine the same type of wage or earnings residuals of the CPS-based literature, only

[6]A large related literature estimates earnings dynamics over the lifecycle without attempting to explain changes in residual inequality over time. See, e.g., Lillard and Weiss (1979); MaCurdy (1982); Abowd and Card (1989); Baker (1997); Guvenen (2007); Hryshko (2012).

they decompose the residuals into different idiosyncratic stochastic shocks (typically, permanent martingale shocks, autoregressive and/or 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 lifetime inequality as well as consumption and savings behavior at the individual and aggregate levels.[7] Studies using the Panel Study of Income Dynamics (PSID) typically find that growth in the variance of transitory shocks accounts 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, this growth in the variance of transitory shocks suggests that changes in the pricing of unobserved skills are unlikely to fully account for the observed rise in residual inequality. While this literature considers a rich structure for stochastic shocks, it typically neglects potential changes in the pricing of unobserved skills as emphasized by the CPS-based literature.[8]

In this paper, we consider a general framework for wage and earnings residuals that incorporates the features highlighted by both of these literatures: (i) unobserved skills with changing non-linear pricing functions and (ii) idiosyncratic shocks that follow a rich stochastic process including permanent and transitory components. Specifically, we consider log wage and earnings residuals for individual i in period t of the form:[9]

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

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

Economic shifts in the demand for or supply of unobserved skills 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 should co-move over time as the labor market rewards their

[7] The co-movement of consumption with permanent and transitory income shocks is also of interest to economists studying the structure of capital/insurance markets (Blundell, Pistaferri, and Preston, 2008; Krueger and Perri, 2006).

[8] A few exceptions are noteworthy. The model estimated by Moffitt and Gottschalk (2012) is most similar to ours, so we discuss it in some detail in Section 5.1.1. Haider (2001) estimates the evolution of lifetime earnings inequality allowing for changing observable and unobservable skill prices as well as an $ARMA(1,1)$ process for stochastic shocks. In their study of black – white wage inequality in the early 1980s, Card and Lemieux (1994) estimate the evolution of observed and unobserved skill prices accounting for $AR(1)$ stochastic shocks to unobserved skills.

[9] This specification is consistent with wage/earnings functions that are multiplicatively separable in observable factors (like education and experience), unobserved skills, and idiosyncratic wage/earnings shocks. More generally, observed and unobserved skills may be non-separable in wage or earnings functions. For example, in the case of fixed observable differences X_i (e.g. cohort of birth, race/ethnicity, or educational attainment prior to labor market entry), we could write $\mu_t(\theta_i, X_i)$ where the distribution of $\epsilon_{i,t}$ may depend on X_i. Our theoretical and empirical analyses below could all be conditioned on X_i in this case.

skill set more or less. The recent literature on 'polarization' in the U.S. labor market (Autor, Levy, and Murnane, 2003; 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 flexible non-linear $\mu_t(\cdot)$ pricing functions.

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 permanent 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) 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).[10] In our analysis, the sizeable increase in the variance of wages across firms (within industries and geographic regions) documented by Dunne et al. (2004) and Barth et al. (2011) is likely to manifest itself in a rise in the variance of permanent shocks, since most workers switch firms over the lifecycle but switches are infrequent. For similar reasons, any increase in the variance of wages due to increased occupational switching over time (Kambourov and Manovskii, 2009) is also likely to show up as an increase in the variance of permanent shocks.

Distinguishing between changes in the variance of shocks and in skill pricing can be useful for understanding household consumption and savings behavior, since household decisions are likely to respond quite differently to an increase in the price of skill than they would to an increase in the variance of permanent or transitory shocks. Transitory shocks have little effect on consumption behavior, while an increase in the variance of permanent shocks should lead to increases in consumption inequality over time within and across cohorts. 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 of and demand for skills or by major policy changes. To the extent that they are well-anticipated, changes in $\mu_t(\cdot)$ are likely to have little effect on consumption inequality over time for a given cohort; however, growth in the return to unobserved skill should raise consumption inequality across successive cohorts. Furthermore, changes in the variance of permanent shocks affect precautionary savings motives, while changes in skill prices should not (regardless of their predictability).[11]

[10] 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, Carrillo-Tudela, and Coles, 2011; Bagger, Fontaine, Postel-Vinay, and Robin, 2011). This literature incorporates unobserved worker human capital or productivity differences but does not account for changing market demands for human capital, implicitly assuming $\mu_t(\theta) = \mu(\theta)$.

[11] Changes in the underlying uncertainty of skill prices should affect precautionary savings motives in the same way

Workers of different ability/skill levels may face different levels of labor market risk, which may evolve differently over time. For example, the diffusion of new skill-biased technology in the presence of labor market frictions can produce different trends in the variance of permanent shocks by ability, since more able workers may adopt these technologies first (Violante, 2002). To account for this possibility, our analysis explicitly accounts for heteroskedasticity in permanent shocks, allowing the relationship between unobserved ability and the variance of permanent shocks to change freely over time.

An important question is whether all of these earnings components can be separately identified using standard panel data sets on wages/earnings without making strong distributional or functional form assumptions. We, therefore, begin with a formal analysis of nonparametric identification, drawing on insights from the measurement error literature (especially, Hu and Schennach (2008); Schennach and Hu (2013)) and the analysis of Cunha, Heckman, and Schennach (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.[12] 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 a $MA(q)$ process, our main theoretical 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 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. These results apply to any cohort, so we can also identify changes in the distribution of unobserved ability/skill across any cohorts we observe for enough time periods.

We next briefly discuss 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 and discuss minimal data and moment requirements for identification of $\mu_t(\cdot)$ and moments of the

as changes in the variance of permanent shocks.

[12] Assumptions on ε_{it} of this nature are often employed in the literature (Abowd and Card, 1989; Blundell and Preston, 1998; Haider, 2001; Gourinchas and Parker, 2002; Meghir and Pistaferri, 2004; Blundell, Pistaferri, and Preston, 2008; Heathcote, Perri, and Violante, 2010).

distributions for unobserved ability, transitory and permanent shocks. We use minimum distance estimation with second- and third-order residual moments to estimate our model with linear and cubic $\mu_t(\cdot)$ functions, considering 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 $\mu_t(\cdot)$ are fixed. Accounting for unobserved skills is important for understanding the evolution of log earnings residuals. Our results also demonstrate the importance of accounting for permanent and transitory shocks when estimating the pricing of unobserved skills. Time patterns for the distribution of log earnings residuals are quite different from the patterns estimated for the pricing of unobserved skills.[13]

We decompose the variance of residuals into components for (i) unobserved skill prices, (ii) permanent shocks, and (iii) transitory shocks. We also examine changes in $\mu_t(\theta)$ at the top and bottom of its distribution over time as well as potential changes in the distribution of θ across cohorts. We highlight five main findings. (i) 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. Since the mid-1990s, we observe some evidence of polarization: the returns to skill have declined at the bottom of the distribution while they have remained relatively constant over the top half. (ii) The variance of unobserved skill changed very little across most cohorts in our sample (those born between 1925 and 1955). (iii) 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. (iv) The variance of permanent shocks declined very slightly over the 1970s, then rose systematically through the end of our sample by 15 to 20 log points. The increase in this variance over the 1980s and 1990s was strongest for workers with low unobserved ability. (v) In most years, the distribution of $\mu_t(\theta)$ is positively skewed, while the distributions of permanent and (especially) transitory shocks are negatively skewed.

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 identification and estimation 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.

[13] Key findings also hold for log weekly wage residuals as shown in Appendix B.

2 Nonparametric Identification

In this section, we provide nonparametric identification results for our model. The baseline model is the following factor model:

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

where the distributions of unobserved factors θ_i and $\varepsilon_{i,t}$ are unspecified, the functional form of $\mu_t(\cdot)$ is also unspecified but strictly monotonic, and we only observe $\{W_{i,t}\}$ (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 drop the cross-sectional subscript i except where there may otherwise be some confusion. First, we discuss identification when ε_t is independent over t. Then, we generalize this result to account for 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.

Since all components are nonparametric, we need some normalization. Throughout this section, 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) W_1, W_2, and W_3 are mutually independent conditional on θ.

(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) For all $\bar{\theta}, \tilde{\theta} \in \Theta$, the set $\{w_3 : f_{W_3|\theta}(w_3|\bar{\theta}) \neq f_{W_3|\theta}(w_3|\tilde{\theta})\}$ has positive probability whenever $\bar{\theta} \neq \tilde{\theta}$.

(v) We normalize $\mu_1(\theta) = \theta$ and $E[\varepsilon_t|\theta] = 0$ for all t.

Assumption 1 is adopted from Hu and Schennach (2008) and is also used in Cunha, Heckman, and Schennach (2010). 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. Note that ε_t may be heteroskedastic, since this condition only requires conditional independence given θ. 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. Condition (iv) holds when $\mu_3(\cdot)$ is strictly monotonic, which is natural in our empirical context. Conditional heteroskedasticity can also ensure condition (iv) (see Hu and Schennach (2008)). 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.

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 data. As in Theorem 2 in Cunha, Heckman, and Schennach (2010), $f_\theta(\cdot)$ can be recovered from $f_{W_3,\theta}(\cdot,\cdot)$ by integrating out θ. 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 $f_{\varepsilon_t|\theta}(\cdot|\cdot)$ by 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 for our main identification result. Specifically, we decompose the idiosyncratic error $\varepsilon_t = \kappa_t + \nu_t$ into two components: a persistent shock, κ_t, and a transitory shock, ν_t. The persistent shock follows a martingale/unit root process with $\kappa_t = \kappa_{t-1} + \eta_t$, and the transitory shock has serial dependence only over q periods, i.e. $f_{\nu_t,\nu_{t+s}}(\nu_1, \nu_2) = f_{\nu_t}(\nu_1) \cdot f_{\nu_{t+s}}(\nu_2)$ for $s > q$. With this, the model can be written as

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

As discussed below, it is possible to identify both $f_\theta(\cdot)$ and $\mu_t(\cdot)$ under this general structure if a long enough time panel is available, i.e. $T \geq 6+3q$.

To further identify the processes for permanent and transitory shocks, we impose the following structure for κ_t and ν_t:

$$\kappa_t = \kappa_{t-1} + \eta_t = \kappa_{t-1} + \sigma_t(\theta)\zeta_t \tag{4}$$

$$\nu_t = \xi_t + \sum_{j=1}^{q} \beta_j \xi_{t-j} \tag{5}$$

where ζ_t and ξ_t are mean zero processes whose distributions can vary over time. The persistent shock η_t is conditional heteroskedastic through $\sigma_t(\theta)$, and the transitory shock follows a MA(q) process. 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. For $q=1$, equation (3) becomes

$$W_t = \mu_t(\theta) + \sum_{j=1}^{t} \sigma_j(\theta)\zeta_j + (\xi_t + \beta_t \xi_{t-1}). \tag{6}$$

In addition to $f_\theta(\cdot)$ and all $\mu_t(\cdot)$, we identify $\sigma_t(\cdot)$, $f_{\zeta_t}(\cdot)$, $f_{\xi_t}(\cdot)$, and β_t for all but the last two periods under the following assumption. (Define $\Delta A_t \equiv A_t - A_{t-1}$ and $\theta_t \equiv \mu_t(\theta)$.)

Assumption 2. *The following conditions hold in equation (6) 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.

(ii) Unobserved components ζ_t, ξ_t, and θ are mutually independent for all $t = 1, \ldots, 9$.

(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) For all $\bar{\theta}, \tilde{\theta} \in \Theta$, the set $\{w : f_{\Delta W_7|\theta}(w|\bar{\theta}) \neq f_{\Delta W_7|\theta}(w|\tilde{\theta})\}$ has positive probability whenever $\bar{\theta} \neq \tilde{\theta}$. The same condition holds for $f_{\Delta W_8|\theta}$ and $f_{\Delta W_9|\theta}$.

(v) We impose the following normalization: $\kappa_0 = \xi_0 = 0, \mu_1(\theta) = \theta, E[\zeta_t] = E[\xi_t] = 0$, $E[\zeta_t^2] = 1$, and $\sigma_t(\cdot) > 0$ for all t.

(vi) Let $m_{\zeta_t,n} = \int_{-\infty}^{\infty} \zeta^n dF_{\zeta_t}(\zeta)$ and $m_{\xi_t,n} = \int_{-\infty}^{\infty} \xi^n dF_{\xi_t}(\xi)$. For all t, the following Carleman's condition holds for ζ_t and ξ_t:

$$\sum_{n=1}^{\infty} (m_{\zeta_t,2n})^{-\frac{1}{2n}} = \infty \quad \text{and} \quad \sum_{n=1}^{\infty} (m_{\xi_t,2n})^{-\frac{1}{2n}} = \infty.$$

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. Condition (ii) implies similar mutual independence among $W_1, W_2, W_3, \Delta W_4, \ldots, \Delta W_9$ conditional on θ. Condition (vi) is a technical assumption that assures the equivalence between the sequence of moments and the distribution. We now state the main identification result.

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

The proof provided in Appendix A.2 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:

$$\begin{aligned} W_1 &= \theta + \varepsilon_1 = \theta + \eta_1 + \nu_1 = \theta + \sigma_1(\theta)\zeta_1 + \nu_1 \\ \Delta W_4 &= \Delta\mu_4(\theta) + \Delta\varepsilon_4 = \Delta\mu_4(\theta) + \eta_4 + \Delta\nu_4 = \Delta\mu_4(\theta) + \sigma_4(\theta)\zeta_4 + \Delta\nu_4 \\ \Delta W_7 &= \Delta\mu_7(\theta) + \Delta\varepsilon_7 = \Delta\mu_7(\theta) + \eta_7 + \Delta\nu_7 = \Delta\mu_7(\theta) + \sigma_7(\theta)\zeta_7 + \Delta\nu_7. \end{aligned}$$

These triplets are mutually independent conditional on θ. The differences ΔW_4 and ΔW_7 are analogous to W_2 and W_3 in the simpler model with serially independent ε_t. We can simply apply Lemma 1 to identify $f_\theta(\cdot)$ and both $\Delta\mu_4(\cdot)$ and $\Delta\mu_7(\cdot)$ functions. 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.[14] Notice that we actually identify $f_{W_t|\theta}(\cdot|\cdot)$ in this step, which will be used in the next steps.

In a second step, we establish identification of $E[\xi_t^2], \beta_t$, and $\sigma_t(\cdot)$ for $t = 1, \ldots, 7$ using various second moments between $W_1, \ldots, W_9$. For example, consider the triplet (W_1, W_2, W_3). From $E[W_1 W_3], E[W_1^2]$, and $E[W_1, W_2]$, we can identify $E[\sigma_1^2(\theta)], E[\xi_1^2]$, and β_1, respectively. With these, $\sigma_1(\cdot)$ is identified from $E[W_1^2|\theta]$, which can be calculated from $f_{W_1|\theta}(\cdot|\cdot)$. We can apply this same argument to triplets $\{(W_t, W_{t+1}, W_{t+2})\}_{t=2}^7$ to identify analogous elements up through $t = 7$.

[14]This step does not rely on the $MA(q)$ structure for ν_t. Only serial independence $\nu_t \perp\!\!\!\perp \nu_{t+s}$ for $s > q$ is needed to establish identification of $f_\theta(\cdot)$ and all $\mu_t(\cdot)$.

In a third step, we identify $f_{\eta_t}(\cdot)$ and $f_{\xi_t}(\cdot)$ for $t = 1, \ldots, 7$. In this step, we first identify the infinite sequence of moments $\{(E[\eta_t^k], E[\xi_t^k])\}_{k=1}^{\infty}$ by mathematical induction. Then, the Carleman's condition in Assumption 2 (vi) enables us to apply Hamburger's Theorem to determine $\{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot)\}_{t=1}^{7}$ uniquely.

2.3 Identification under Homoskedasticity

When $\eta_t \perp\!\!\!\perp \theta$ so ε_t is homoskedastic, condition (iv) of Assumption 2 requires strict monotonicity of $\Delta\mu_t(\cdot)$ for $t = 7, 8, 9$. This is quite strong, especially in light of recent discussions of polarization in the demand for skill. Fortunately, this assumption can be relaxed. In Appendix A.3, we show that the model is still fully identified by replacing condition (iv) with the following lower level assumption on $\mu_t(\cdot)$ (along with some minor additional technical conditions):

> For $t = 7, 8, 9$, the functions $\Delta\mu_t(\theta)$ are continuously differentiable with $\Delta\mu_t'(\theta^*) = 0$ for at most a finite number of θ^*.

This condition allows $\Delta\mu_t(\cdot)$ to be non-monotonic. However, it does not allow the marginal value of unobserved skill $\mu_t'(\cdot)$ to be time invariant over an interval of θ in later years. The complete set of assumptions and identification proof can be found in Appendix A.3.

2.4 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 $\{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot), \beta_t\}_{t=1}^{T-2}$. While not surprising, the fact that distributions of transitory and permanent shocks cannot be separately identified for the final two periods is useful to keep in mind during estimation.

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 an $AR(1)$ 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.[15]

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

$$\mu_t(\theta) = m_{0,t} + m_{1,t}\theta + ... + m_{p,t}\theta^p \tag{7}$$

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

$$\begin{aligned} W_{i,a,t} &= \mu_t(\theta_i) + \kappa_{i,a,t} + \nu_{i,a,t} & (8) \\ \kappa_{i,a,t} &= \kappa_{i,a-1,t-1} + \eta_{i,a,t} & (9) \\ \nu_{i,a,t} &= \xi_{i,a,t} + \beta_{1,t}\xi_{i,a-1,t-1} + \beta_{2,t}\xi_{i,a-2,t-2} + ... + \beta_{q,t}\xi_{i,a-q,t-q}. & (10) \end{aligned}$$

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 residual 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 $\kappa_{i,a,t}$, and (iii) transitory shocks $\nu_{i,a,t}$. 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: $\sigma^k_{\eta_t} \equiv E[\eta^k_{i,a,t}] = E[\sigma^k_t(\theta)]$

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

and $\sigma^k_{\xi_t} \equiv E[\xi^k_{i,a,t}]$ 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

$$\begin{aligned}
\sigma^k_{\kappa_{a,t}} &\equiv E[\kappa^k_{i,a,t}] = \sum_{j=0}^{a-1} \sigma^k_{\eta_{t-j}} \\
\sigma^k_{\nu_{a,t}} &\equiv E[\nu^k_{i,a,t}] = \sigma^k_{\xi_t} + \sum_{j=1}^{\min\{q,a-1\}} \beta^k_{j,t}\sigma^k_{\xi_{t-j}}.
\end{aligned}$$

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:

$$E[W^2_{i,a,t}|a,t] = \sum_{j=0}^{p}\sum_{j'=0}^{p} m_{j,t}m_{j',t}E[\theta^{j+j'}] + \sum_{j=0}^{a-1}\sigma^2_{\eta_{t-j}} + \sigma^2_{\xi_t} + \sum_{j=1}^{\min\{q,a-1\}} \beta^2_{j,t}\sigma^2_{\xi_{t-j}} \tag{11}$$

and covariances:

$$E[W_{i,a,t}W_{i,a+l,t+l}|a,t,l] = \sum_{j=0}^{p}\sum_{j'=0}^{p} m_{j,t}m_{j',t+l}E[\theta^{j+j'}] + \sum_{j=0}^{a-1}\sigma^2_{\eta_{t-j}} + E(\nu_{i,a,t}\nu_{i,a+l,t+l}) \text{ for } l \geq 1. \tag{12}$$

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 $q = 1$, 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 (11) 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 (12) but varying l, the second term due to permanent shocks remains constant, while the final term disappears altogether for $l > q$. As emphasised in 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 and homoskedasticity in η_t (i.e. $\sigma_t(\theta) = \sigma_t, \forall\theta$). In this case, we need to identify/estimate a total of $(4+p)T + p - 9$ parameters.[16] For T periods of data, we have a total of $T(T+1)/2 + T - 1$ moments, which includes

[16] The parameters include $2p-1$ 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 $\beta_{1,t}$, $t = 2, ..., T-2$.

$T(T+1)/2$ unique variance/covariance terms and $T-1$ moments coming from $E[\mu_t(\theta_i)] = 0$ for $t = 2, ..., T$. A necessary condition for identification is, therefore, $(4+p)T+p-9 \leq T(T+1)/2+T-1$. 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 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 second order moments to estimate the model.

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 $k = 2, ..., p$. Incorporating these moments adds $2(p-1)$ additional parameters $\left(\sigma^3_{\eta_1}, ..., \sigma^{p+1}_{\eta_1}, \sigma^3_{\xi_1}, ..., \sigma^{p+1}_{\xi_1}\right)$. While the number of additional parameters does not depend on T, 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. $\sigma^k_{\eta_t}$ and $\sigma^k_{\xi_t}$) 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.[17] In practice, it may be difficult to precisely estimate higher order residual moments given the sample sizes of typical 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 (11) and (12). Although the distributions for η_{it} and ξ_{it} are assumed to depend only on time and not age or

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 $q = 1$.

[17] Including all cross-product moments from order $2, ..., k$ yields a total of $\sum_{j=2}^{k} \begin{pmatrix} T+j-1 \\ j \end{pmatrix}$ moments plus $T-1$ moments for $E[\mu_t(\theta)] = 0$.

cohort, the distribution of $\kappa_{i,a,t}$ will vary with age since older cohorts have accumulated a longer history of permanent shocks.[18]

Our identification results above can be applied separately for each cohort; however, differences in the variance and covariance terms across age/cohort for the same time periods in equations (11) and (12) can be used to help identify the effects of permanent and transitory shocks. The fact that market-based skill pricing functions $\mu_t(\cdot)$ vary only with time and not age/cohort is particularly helpful. To see why, consider equation (12) for $l > q$. In this case, the final term due to transitory shocks disappears, while the first term is the same for all cohorts. The second term reflects the sum of all permanent shocks from the time of labor market entry to year t for each cohort. By comparing these covariances across cohorts for fixed t, we can recover the variances of permanent shocks from time zero through $t-1$. Of course, this very simple identification strategy for $\sigma^2_{\eta t}$ can no longer be used if the distributions of permanent shocks varies freely with age or cohort. Allowing for cohort differences in the distribution of θ also reduces the value of additional cohorts, since any terms related to unobserved skills would then become cohort-specific. Still, as long as skill pricing functions are independent of age and cohort, the inclusion of multiple cohorts provides additional variation that can be useful for identification and estimation even if the distributions of shocks are allowed to vary by age/cohort and the distribution of skills varies across cohorts.

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.[19] 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

[18] This variation adds new parameters to be identified/estimated for each additional cohort. In particular, we must identify/estimate separate $\sigma^k_{\kappa_{a(0),0}}$ 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.

[19] 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.

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, 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 log earnings residuals after controlling for differences in educational attainment, race, and age. (Log weekly wage residuals are considered in Appendix B.) Residuals are derived 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 results 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.

Figure 2: Selected Log Earnings Residual Quantiles, 1970–2008

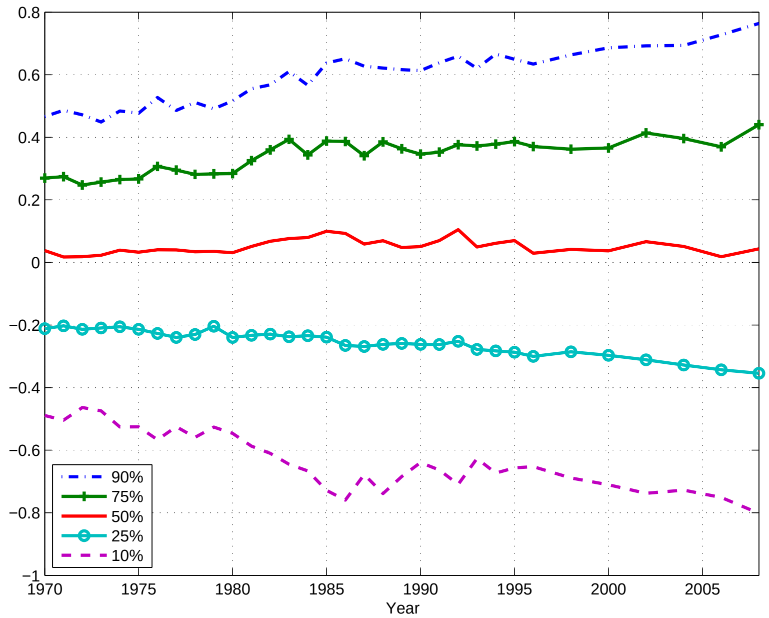
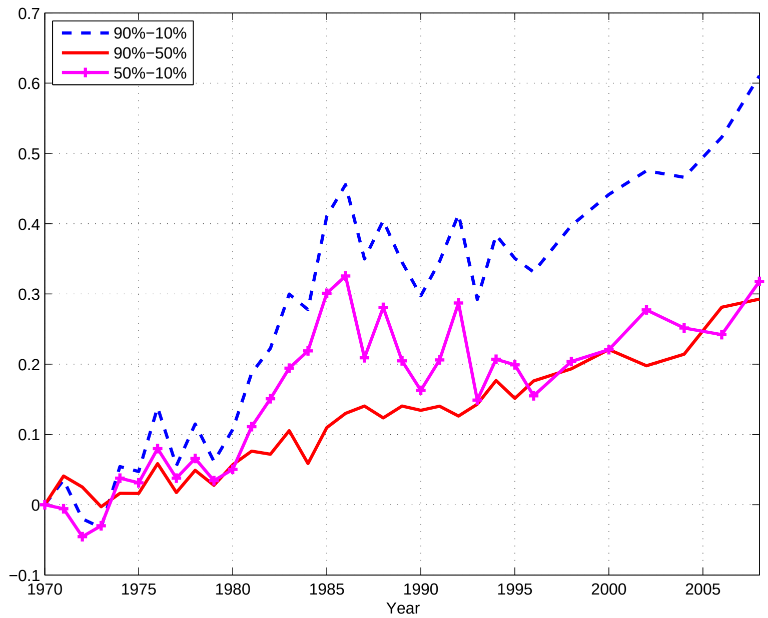


Figure 3: Changes in 90-10, 90-50, and 50-10 Ratios for Log Earnings Residuals, 1970–2008



5 Estimates of Unobserved Skill Pricing Functions and Earnings Dynamics over Time

We use minimum distance estimation and the residual moments described above to estimate the model for men using the PSID. We discuss results for log earnings residuals in the text; however, conclusions are quite similar for log weekly wages as reported in Appendix B. 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:

$$\begin{aligned}\frac{1}{n_{A,t,l}}\sum_{i:a\in A} W_{i,a,t}W_{i,a-l,t-l} &\xrightarrow{p} E[W_{i,a,t}W_{i,a-l,t-l}|a\in A,t,l] \\ &= \sum_{a\in A}\omega_{a,t,l}E[W_{i,a,t}W_{i,a-l,t-l}|a,t,l]\end{aligned}$$

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.[20]

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.[21]

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

1. <u>pricing of unobserved skills</u>: $Var[\mu_t(\theta)]$;
2. <u>permanent shocks</u>: $\sigma^2_{\kappa t}=\sum_a \varphi_{a,t}\left[\sum_{j=0}^{a-1}\sigma^2_{\eta t-j}\right]$ where $\varphi_{a,t}$ is the fraction of persons in period t that are age a;
3. <u>transitory shocks</u>: $\sigma^2_{\nu_t}=\sigma^2_{\xi_t}+\sum_{j=1}^{q}\beta^2_{jt}\sigma^2_{\xi_{t-j}}$.

[20]See the Online Appendix for additional estimation details.

[21]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.

We also discuss the evolution of different quantiles in the distributions 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 $\mu_t(\cdot)$, assuming homoskedasticity of all shocks. No distributional assumptions on θ or the permanent and transitory shocks are needed 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 in most of the PSID-based literature. The remaining columns allow $\mu_t(\cdot)$ to vary freely 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 for ν_t.[22] 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.[23] We also report results for an $MA(5)$ for comparison. Third, the estimated serial correlation in transitory shocks is weaker when $\mu_t(\cdot)$ is allowed to vary over time (e.g. β_1 estimates are more than 10% lower).

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 and an $ARMA(1,1)$ yield very similar conclusions (see Appendix B). Figure 4 reports the estimated variances (and standard errors) for $\mu_t(\theta)$, η_t, and ξ_t over time.[24] 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 more than 10 percentage points over the 1970s and early 1980s, then falls back to its original level by the late 1990s. It remains fairly constant thereafter. The

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

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

[24] 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.

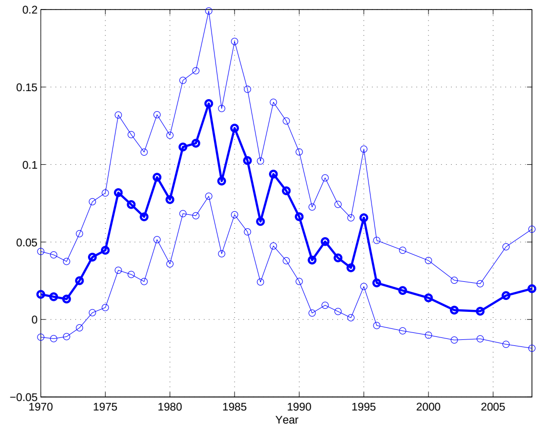
Table 1: Estimates Assuming $\nu_t \sim MA(q)$ using Variances/Covariances (Linear $\mu_t(\cdot)$)

	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
	(0.036)	(0.025)	(0.027)	(0.033)	(0.026)	(0.027)	(0.027)
β_2	·	0.257	0.222	·	0.186	0.172	0.194
	·	(0.030)	(0.025)	·	(0.025)	(0.021)	(0.020)
β_3	·	·	0.246	·	·	0.141	0.137
	·	·	(0.034)	·	·	(0.025)	(0.022)
β_4	·	·	·	·	·	·	0.126
	·	·	·	·	·	·	(0.020)
β_5	·	·	·	·	·	·	0.084
	·	·	·	·	·	·	(0.024)

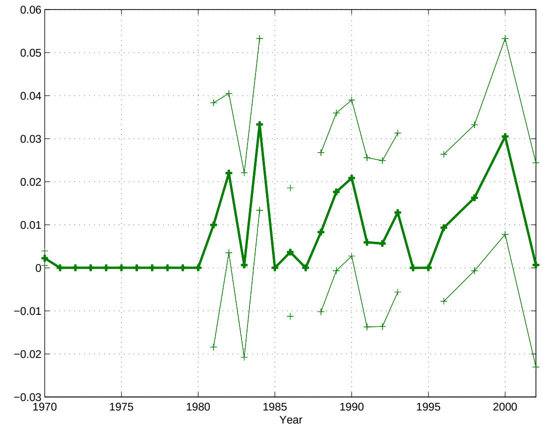
variance of the transitory component rises sharply (5-10 percentage points) in the early 1980s, then fluctuates up and down for about ten years before it stabilizes in the late 1990s. The variance of the permanent component declines slightly over the 1970s, then rises continuously and at roughly the same rate over the rest of the sample period. Between 1980 and 2002, the variance of permanent shocks increases by nearly 20 percentage points, more than the increase in total residual variance. 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.

The time patterns in Figure 5 differ from the conclusions reached in both the PSID- and CPS-based literatures on log wage/earnings dynamics and inequality. The CPS-based literature implicitly ignores the roles of permanent and transitory shocks, equating changes in the total variance of log earnings/weekly wage residuals with changes in the returns to unobserved skills. This need not be the case in our model. Figure 5 suggests very different time patterns for the returns to unobserved skills compared to the total variance of log earnings residuals. (Similar patterns are estimated for log weekly wage residuals as reported in Appendix B.) While the total variance of log earnings residuals mainly increased in the early 1980s and the late 1990s and 2000s, variation in the returns to unobserved skills increased smoothly from 1970 to the mid-1980s before declining over the next 15-20 years. Variation in the returns to unobserved skill is quite stable after

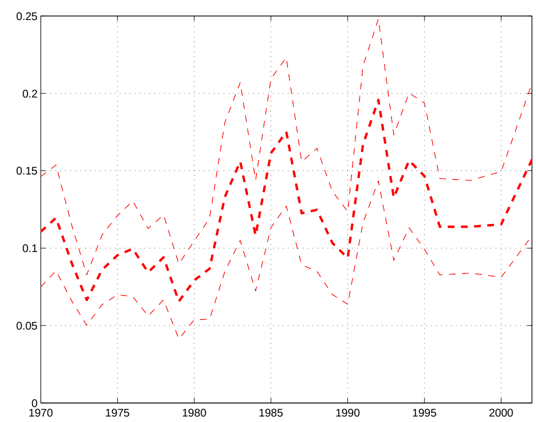
Figure 4: Variances of $\mu_t(\theta)$, η_t, and ξ_t



(a) Variance of $\mu_t(\theta)$

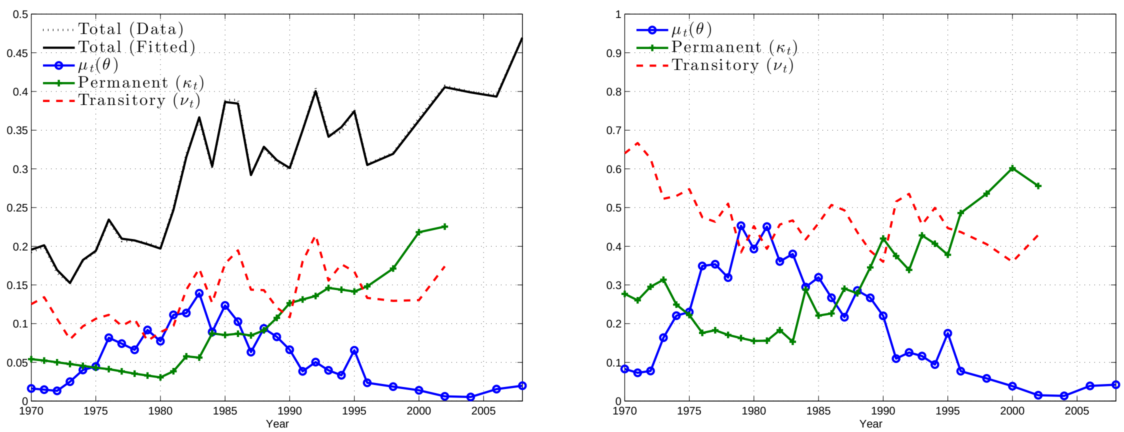


(b) Variance of Permanent Shocks η_t



(c) Variance of Transitory Shocks ξ_t

Figure 5: Variance Decomposition and Shares of Each Component



(a) Variance Decomposition

(b) Shares of Each Component

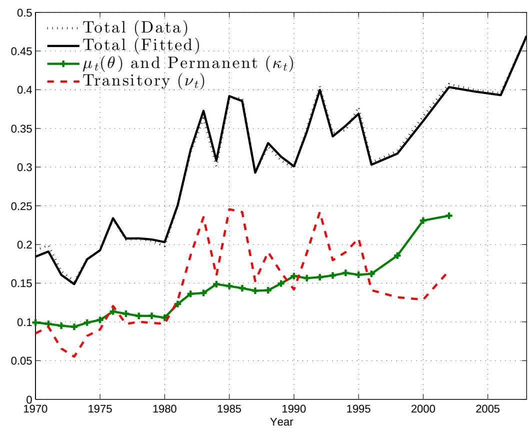
1995, in sharp contrast with the rapid increase in total residual variance.

The PSID-based literature explores the relative importance of permanent and transitory shocks; however, it typically ignores variation in the pricing of unobserved skills (i.e. assumes $\mu_t(\cdot) = \mu(\cdot)$).[25] Figure 6 shows that this is not innocuous, even if one is only interested in the relative importance of permanent and transitory components. This figure decomposes the total variance into 'permanent' and 'transitory' components based on a model that assumes $\mu_t(\cdot) = \mu(\cdot)$ is time invariant. Here, the 'permanent' component is given by the variance of $\mu(\theta) + \kappa_t$. These estimates suggest more modest increases in the permanent component and stronger increases in the transitory component over the early 1980s relative to estimates from our more general model that allows for variation in unobserved skill prices (Figure 5).

In Appendix B, we explore the robustness of our main variance decomposition results (Figure 5) to a few alternative specifications. First, we show that different assumptions about the transitory component yield very similar results. Specifically, the dynamics and relative importance of all three variance components are quite similar to those shown in Figure 5 if ν_t follows an $MA(1)$, $MA(5)$, or $ARMA(1,1)$ process. Second, we consider the possibility that the variance of transitory

[25] There are a few notable exceptions in the literature (e.g. Haider 2001, Moffitt and Gottschalk 2012); however, these studies abstract from other important features of the problem. Haider (2001) abstracts from permanent shocks, η_t. Moffitt and Gottschalk (2012) assume that the variance of permanent shocks remains constant over time, but they multiply both θ and κ_t by the same time-varying 'price'. We consider this specification below.

Figure 6: Variance Decomposition with Time-Invariant $\mu_t(\theta)$



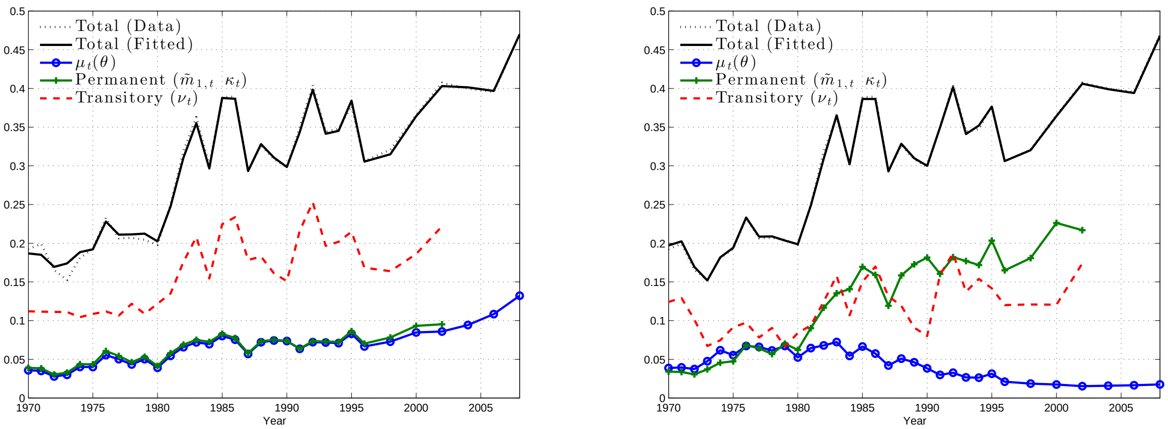
and permanent shocks may vary with age. Using rich administrative data from Norway, Blundell, Graber, and Mogstad (2013) estimate declines in the variance of both permanent and transitory shocks prior to age 35, with remarkable stability thereafter. We, therefore, estimate our model allowing the variances of both $\eta_{i,a,t}$ and $\xi_{i,a,t}$ to be linear functions of age over ages 20-35 and constant thereafter. Our estimates suggest that the variances of both permanent and transitory shocks decline by roughly 25% over these first fifteen years of workers' careers; however, the variance decompositions from this more general specification are nearly identical to that of Figure 5 (see Figure 16 in Appendix B).

5.1.1 Permanent Shocks to Skills

In our baseline model of earnings dynamics, shocks are distinct from ability or skills and, therefore, do not appear inside the pricing function $\mu_t(\cdot)$. Alternatively, one might treat permanent shocks η_t as shocks to unobserved skills, specifying log earnings residuals as $W_t = \tilde{\mu}_t(\theta + \kappa_t) + \nu_t$. Assuming $\kappa_0 = 0$, θ reflects initial skill levels. Moffitt and Gottschalk (2012) estimate this alternative model where $\tilde{\mu}_t(\cdot)$ is linear and the variance of permanent shocks, η_t, is constant over time.[26] The assumption of linearity implies that $\tilde{\mu}_t(\cdot) = \tilde{m}_{0,t} + \tilde{m}_{1,t}\theta + \tilde{m}_{1,t}\kappa_t$, so it is still possible to decompose the residual variance into three components: variation due to initial skill differences θ,

[26] Moffitt and Gottschalk (2012) also allow for heterogeneous growth rates in unobserved ability/skill; however, the estimated variance of these growth rates is insignificantly different from zero. Hryshko (2012) also finds no evidence of heterogenous growth rates when Martingale/unit root shocks are included.

Figure 7: Variance Decomposition with Permanent Shocks to Skills



(a) Fixed Variance of Permanent Shocks ($\sigma^2_{\eta_t} = \sigma^2_\eta$)

(b) Time-Varying $\sigma^2_{\eta_t}$

variation due to permanent innovations in unobserved skills κ_t, and variation due to transitory shocks ν_t. Assuming the variance of η_t remains constant over time effectively restricts the variance contribution due to initial skills to perfectly co-move over time with the contribution due to innovations in skill as seen in Figure 7(a).[27] This alternative model yields a slightly greater increase in the transitory component when compared with our baseline model (Figure 5), but its time pattern is quite similar. The time patterns for components related to θ and κ_t are quite different from their counterparts in Figure 5, with the value of skills increasing slowly but continually over the entire sample period in Figure 7(a).

Figure 7(b) reveals that much of the discrepancy between these results and those of our baseline model is due to the time-invariance assumption that $\sigma^2_{\eta_t} = \sigma^2_\eta$. Relaxing this assumption, Figure 7(b) shows that the return to unobserved skill rose and then fell over time (peaking in the early 1980s) as observed in Figure 5. While this rise and fall is noticeably more muted in Figure 7(b), the transitory and permanent components show fairly similar time patterns across the two figures.

Although, both our baseline model in equations (8)-(10) and this alternative specification (with time-varying $\sigma^2_{\eta_t}$) produce a very similar fit to the data, we focus on our baseline specification

[27]The co-movement of the first two components is not exactly the same due to changes in the age distribution over time, which affects the variance of κ_t across the entire population.

in the rest of the paper for two primary reasons. First, our nonparametric identification results of Section 2 directly apply to the baseline model, and we are interested in more general $\mu_t(\cdot)$ pricing functions. Analogous nonparametric identification results for the alternative model with permanent ability/skill shocks would require a different approach. Second, recent studies document large increases in the variance of wages across firms/plants for similar workers (Dunne et al., 2004; Barth et al., 2011). Barth et al. (2011) show that this increase is not caused by changes in the sorting of workers across establishments by skill, arguing instead that it is the result of a widening in the dispersion of productivity across plants. This is consistent with an increase in the variance of permanent shocks in our baseline model, where those shocks (at least partially) incorporate differences in wage payments across firms.[28]

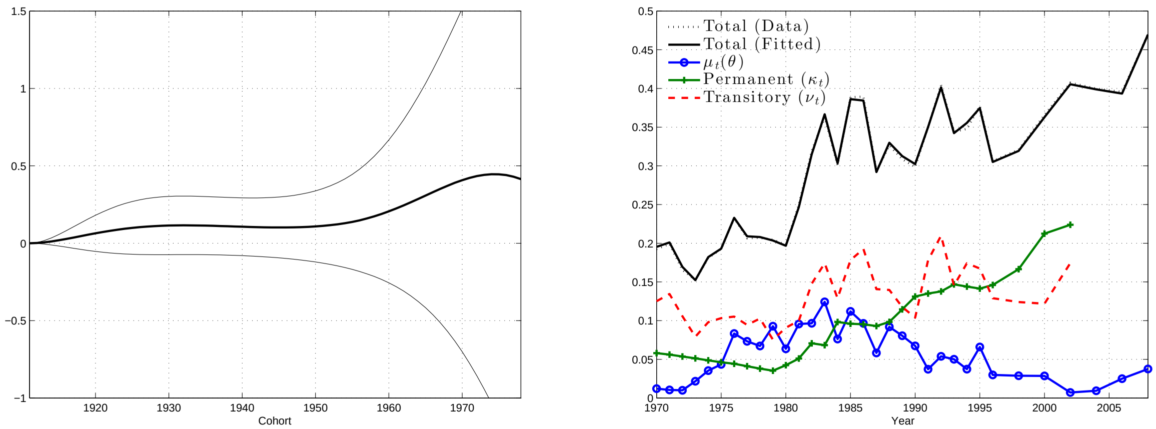
5.1.2 Cohort Differences in the Distribution of Unobserved Ability

Thus far, we have assumed that all cohorts have the same distribution of unobserved ability. Our identification results can be applied separately for each cohort (assuming they appear in the data for at least 9 years), so it is possible, in theory, to estimate everything separately for each cohort. Given sample sizes in the PSID, this is impractical. Instead, we allow the variance of θ to vary across cohorts, assuming the skill pricing functions and stochastic processes for all shocks depend only on calendar time.[29] We re-estimate our baseline model assuming the path for $\sigma_{\theta,c}$ across cohorts can be represented by a cubic spline (with two interior knots). Figure 8(a) reports the estimated variance of θ across cohorts along with 95% confidence bands. These results suggest very similar variation in unobserved ability across most of the cohorts in our sample. The variance of θ for cohorts born between 1925 and 1955 ranges between 0.1 and 0.15. The point estimates suggest a sizeable increase in the variance for later cohorts; however, these estimates are very imprecise as evident from the growing standard errors. These later cohorts do not appear in our data for many years, since we first start following them when they turn age 30 (and the PSID moves to a biennial survey in 1997). Importantly, Figure 8(b) shows that allowing for cohort variation in the variance of θ has little effect on our baseline variance decomposition results.

[28] It would be interesting to exploit worker-firm matched panel data to further decompose the variances of shocks into worker- and firm-specific components.

[29] Given linear skill pricing functions $\mu_t(\cdot)$, our first stage log earnings regressions absorb any mean differences in the distribution of θ across cohorts. Controlling for unrestricted year-specific age effects makes it impossible to separately identify mean differences across cohorts in this case.

Figure 8: Accounting for Cohort Differences in the Variance of Unobserved Ability



(a) Estimated $\sigma^2_{\theta,c}$ (with 95% CI) Across Cohorts

(b) Variance Decomposition

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

We next estimate our baseline 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.[30] This aids in identification of $\mu_t(\cdot)$ pricing functions and allows for estimation of third-order moments for permanent and transitory shocks. We continue to assume that shocks are homoskedastic for now, relaxing this assumption below.

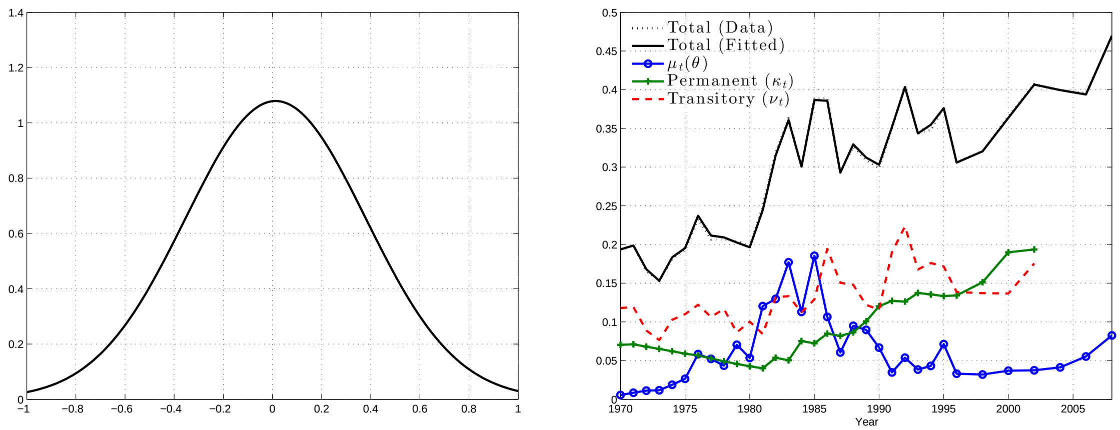
Our third-order moment conditions contain moments of θ up to $E(\theta^9)$. While we could estimate these higher moments directly along with all other parameters of the model, we instead assume that f_θ is a mixture of two normal distributions. Figure 9(a) shows the estimated distribution for θ.[31] Figure 9(b) performs the same type of variance decomposition as above. The results are quite similar to those assuming linear $\mu_t(\cdot)$ functions.

Figure 10 shows the evolution of estimated $\mu_t(\cdot)$ pricing functions for each decade. These

[30] Specifically, we include all $E[W_{a,t}, W_{a-l,t-l}, W_{a-l-k,t-l-k}]$ moments along with all variances/covariances. In aggregating across cohorts, we calculate these third-order moments in the same way we calculate variance/covariance terms. See the Online Appendix for additional details.

[31] 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 9: Estimates Assuming Cubic $\mu_t(\cdot)$ Functions



(a) Distribution of θ

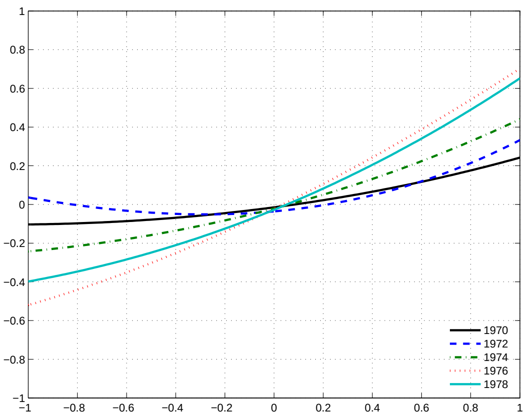
(b) Variance Decomposition

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.[32] At the same time, the $\mu_t(\theta)$ functions are quite stable or even steepening slightly at the top of the distribution.

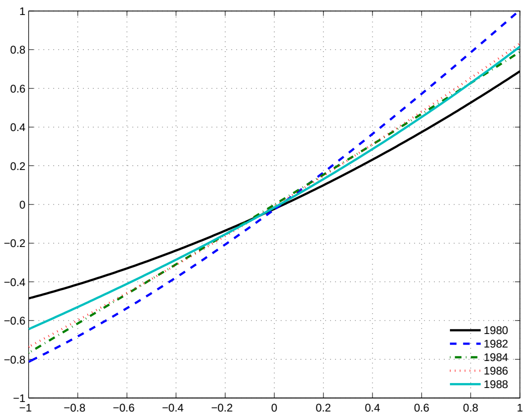
These patterns are more simply summarized in Figure 11, which shows the evolution of selected quantiles and the 90-10, 90-50, and 50-10 ratios for the distribution for $\mu_t(\theta)$ (the latter are relative to their 1970 values). The 90-10 ratio follows a similar pattern to that observed for the variance 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

[32]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.

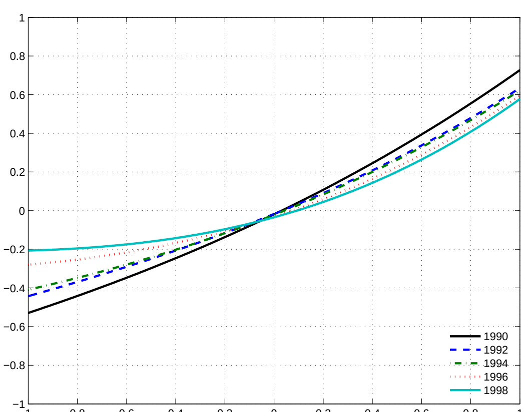
Figure 10: Estimated Cubic $\mu_t(\theta)$ functions



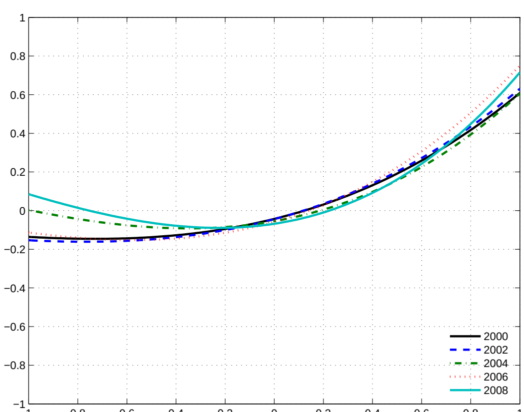
(a) 1970s



(b) 1980s

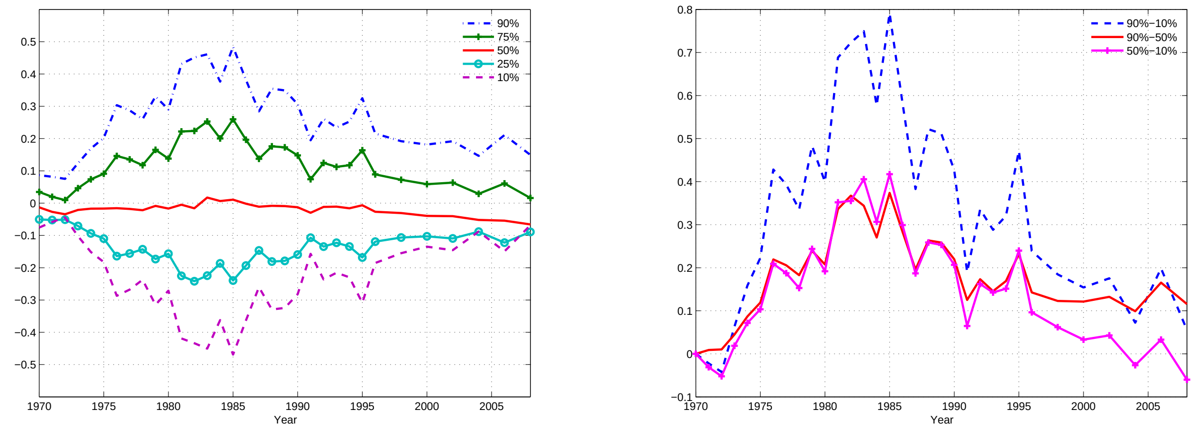


(c) 1990s



(d) 2000s

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



(a) Quantiles of $\mu_t(\theta)$

(b) Changes in 90-10, 90-50, and 50-10 Gaps for $\mu_t(\theta)$

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.

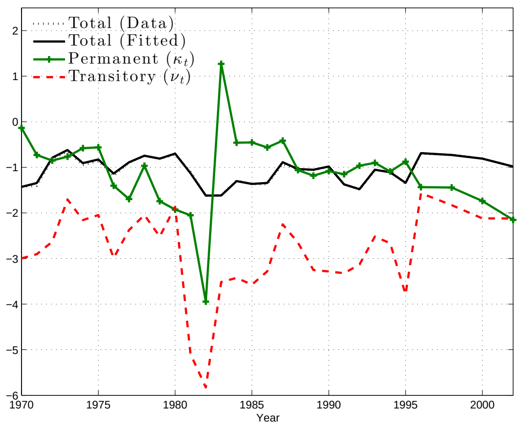
The patterns for $\mu_t(\cdot)$ imply an increasing (more positive) skewness over the 1990s. Figure 12 shows the skewness of total log earnings residuals along with the skewness for the permanent and transitory components of earnings over time. All are typically negatively skewed, in contrast to the 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 -4 in most years.

5.2.1 Heteroskedasticity in Permanent Shocks

Thus far, we have assumed that the distributions of all shocks are independent of unobserved ability. We next examine whether the variance of permanent shocks depends on θ as in equation (4), assuming $\sigma_t(\theta)$ is linear in θ for all t.

Figure 13(a) shows the variance of κ_t for three different quantiles of the θ distribution. Interestingly, our estimates imply notable differences in the variability of permanent shocks (and their time patterns) for workers of different ability levels. Among low-ability workers at the 10th percentile, the the variance of permanent shocks increased rapidly in the 1980s and early 1990s

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



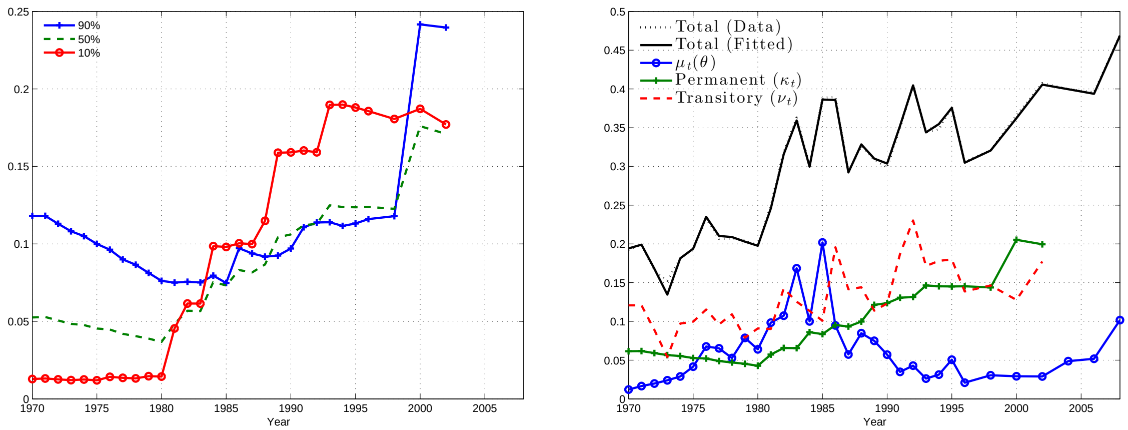
(from less than 0.03 in 1980 to nearly 0.20 by the mid-1990s). The increase in variability over the 1980s and early 1990s was much more modest for high ability types (at the 90th percentile) and followed a decade of declining variability. Near the end of our sample period in the late 1990s, the variability of permanent shocks jumps up, such that the total increase in the variance from 1970 to 2000 was quite similar across the entire distribution of unobserved ability. Figure 13(b) shows that the overall variance decomposition when accounting for heteroskedasticity is quite similar to our results in Figure 9(b) that assume all workers face the same distribution of shocks.

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 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

Figure 13: Estimates Allowing for Heteroskedasticity in Permanent Shocks (Cubic $\mu_t(\cdot)$ Functions)



(a) $\sigma^2_{\kappa,t}(\theta)$ at 10th, 50th, and 90th Quantiles

(b) Variance Decomposition

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. Furthermore, accounting for variation in the distributions of transitory and permanent shocks is important for identifying the evolution of skill pricing functions. Using our estimates, we decompose the variance of log earnings residuals over time into three components: the pricing of unobserved skills, permanent and transitory shocks.

Our results suggest little change in the variance of unobserved skills across cohorts born between 1925 and 1955; however, the pricing of unobserved skills changed substantially over time. There was a sizeable increase in the returns to unobserved skill over the 1970s and early 1980s, but this trend reversed itself in the late 1980s and 1990s, with the pricing of unobserved skills falling back to what it was in 1970. From 1995 onward, unobserved skill prices were fairly 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 (especially for workers with low unobserved skill). 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 pricing functions over the top half of the distribution; however, the value of unobserved skill declined over the bottom half of the distribution as earnings differences between middle- and low-skilled workers narrowed. By the mid-2000s, skill pricing functions were essentially flat over the entire bottom half of the distribution. These changes at the bottom are 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 necessarily informative about the evolution of unobserved skill pricing functions, especially in recent decades.

It is difficult to reconcile current theories of skill-biased technical change with the broad trends we estimate for the pricing of unobserved skills and the variance of permanent shocks. Many theories, motivated largely by CPS-based evidence, have sought to explain a long-run rise in unobserved skill prices, but they offer little insight into the subsequent decline we observe. Theories based on the slow diffusion of skill-biased technology in frictional labor markets are helpful for understanding the simultaneous increase in unobserved skill prices and in the variance of permanent shocks (Violante, 2002), since skilled workers are more likely to adopt new technologies when they are lucky enough to find employment at a firm that has upgraded. Yet, this suggests that an increase in the variance of permanent shocks (i.e. the luck of matching with more cutting edge firms) should be seen among higher skilled workers first, in contrast with our findings. Our results are likely to be more consistent with a theory based on the introduction of new technologies that are adopted broadly by skilled workers early on (i.e. the 1970s and early 1980s) but which then diffuse more slowly (and randomly) to low-skilled workers (i.e. the late 1980s and 1990s). The faster rise in permanent shocks for less-skilled workers from the early 1980s on suggests that it may have taken time for many of these workers to match with firms adopting newer technologies (due to labor market frictions). This theory would also be consistent with the rising variance of wages paid across firms as observed in Dunne et al. (2004) and Barth et al. (2011).

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). The completeness assumption in condition (iii) is sufficient for injectivity. 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

$$f_{W_3,W_1,W_2}(w_3, w_1, w_2) = \int_\Theta f_{W_1|\theta}(w_1|\theta) f_{W_3,\theta}(w_3, \theta) f_{W_2|\theta}(w_2|\theta)\, d\theta \quad \text{for all} \quad w_t \in \mathcal{W}_t \tag{13}$$

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 $f_{\varepsilon_t|\theta}(\varepsilon|\theta) = f_{W_t|\theta}(\mu_t(\theta) + \varepsilon)$ since both $f_{W_t|\theta}$ and $\mu_t(\cdot)$ are already known. $\square$

A.2 Proof of Theorem 1

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 $t = 1, 2, 3$. Begin with the following subset of equations:

$$\begin{aligned}
W_1 &= \theta + \varepsilon_1 = \theta + \eta_1 + \nu_1 = \theta + \sigma_1(\theta)\zeta_1 + \nu_1 \\
\Delta W_4 &= \Delta\mu_4(\theta) + \Delta\varepsilon_4 = \Delta\mu_4(\theta) + \eta_4 + \Delta\nu_4 = \Delta\mu_4(\theta) + \sigma_4(\theta)\zeta_4 + \Delta\nu_4 \\
\Delta W_7 &= \Delta\mu_7(\theta) + \Delta\varepsilon_7 = \Delta\mu_7(\theta) + \eta_7 + \Delta\nu_7 = \Delta\mu_7(\theta) + \sigma_7(\theta)\zeta_7 + \Delta\nu_7.
\end{aligned}$$

Assumption 2 (ii) implies that these triplets are mutually independent conditional on θ. Thus, the problem simplifies to that of the serially independent case of Lemma 1. Therefore, we can identify $f_\theta(\cdot)$, $\Delta\mu_4(\cdot)$, and $\Delta\mu_7(\cdot)$ from Lemma 1. As a by-product, we can also identify conditional density functions $f_{W_1|\theta}(\cdot|\cdot)$, $f_{\Delta W_4|\theta}(\cdot|\cdot)$, and $f_{\Delta W_7|\theta}(\cdot|\cdot)$, which will be used in Step 2.

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 $\theta_t \equiv \mu_t(\theta)$ and $g_t(\theta_2)$ is implicitly defined as $\Delta\mu_t(\theta) = g_t(\mu_2(\theta))$. We apply Lemma 1 again to identify $f_{\theta_2}(\cdot)$, $g_5(\cdot)$, and $g_8(\cdot)$. Given monotonicity of all $\mu_t(\cdot)$, we can recover the function $\mu_2(\cdot)$ by $\mu_2(\theta) = F_{\theta_2}^{-1}\left(F_\theta\left(\theta\right)\right)$. 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 $E[\xi_t^2]$, β_t, and $\sigma_t(\cdot)$ for $t = 1, \ldots, 7$.

Consider the following three equations:

$$W_1 = \theta + \kappa_1 + \nu_1 = \theta + \sigma_1(\theta)\zeta_1 + \nu_1$$
$$W_2 = \mu_2(\theta) + \kappa_2 + \nu_2 = \mu_2(\theta) + \sum_{j=1}^{2}\sigma_j(\theta)\zeta_j + \nu_2$$
$$W_3 = \mu_3(\theta) + \kappa_3 + \nu_3 = \mu_3(\theta) + \sum_{j=1}^{3}\sigma_j(\theta)\zeta_j + \nu_3.$$

Recall that we already know $f_\theta(\cdot), \mu_t(\cdot)$, and $f_{W_t|\theta}(\cdot|\cdot)$ for all $t = 1, \ldots, 7$. First, by looking at

$$E\left[W_1 W_3\right] = E[\theta\mu_3(\theta)] + E[\sigma_1^2(\theta)],$$

we can recover $E[\sigma_1^2(\theta)]$. Second, we look at

$$E\left[W_1^2\right] = E[\theta^2] + E[\sigma_1^2(\theta)] + E[\xi_1^2],$$

to identify $E[\xi_1^2]$. Third, we focus on

$$E\left[W_1 W_2\right] = E[\theta\mu_2(\theta)] + E[\sigma_1^2(\theta)] + \beta_1 E[\xi_1^2],$$

which identifies β_1. Finally, $\sigma_1(\cdot)$ is identified from

$$E\left[W_1^2|\theta\right] = \theta^2 + \sigma_1^2(\theta) + E[\xi_1^2].$$

Under the sign normalization, it is given by $\sigma_1(\theta) = \left(E\left[W_1^2|\theta\right] - \theta^2 - E[\xi_1^2]\right)^{1/2}$. Applying the same logic to triplets $\{W_t, W_{t+1}, W_{t+2}\}_{t=2}^{7}$, we can identify $E[\xi_t^2], \beta_t$, and $\sigma_t(\cdot)$ for all $t = 2, \ldots, 7$.

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

Under the Carleman's condition in Assumption 2 (vi), the infinite sequence of moments $\left\{\left(E[\eta_t^k], E[\xi_t^k]\right)\right\}_{k=1}^{\infty}$ for $t = 1, \ldots, 7$ determines $\{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot)\}_{t=1}^{7}$ uniquely.[33] We show that the infinite sequence of moments are identified by mathematical induction. Since we already know $\left\{\left(E[\zeta_t^2], E[\xi_t^2]\right)\right\}_{t=1}^{7}$ by the normalization and Step 2, we now suppose that $E[\zeta_t^k]$ and $E[\xi_t^k]$ are known for some $k \geq 2$. It remains to show that $E[\zeta_t^{k+1}]$ and $E[\xi_t^{k+1}]$ are identified.

Consider the case of $t = 1$. We look at the following two moment conditions: $E[W_1^{k+1}]$ and $E\left[W_1^k W_3\right]$:

$$E[W_1^{k+1}] = E[\sigma_1^{k+1}(\theta)]E[\zeta_1^{k+1}] + E[\xi_1^{k+1}] + C_1$$
$$E[W_1^k W_3] = E[\sigma_1^{k+1}(\theta)]E[\zeta_1^{k+1}] + C_2,$$

where two constants C_1 and C_2 can be calculated from known moments up to k. Also, note that $E[\sigma_1^{k+1}(\theta)]$ is known from $\sigma_1(\cdot)$ and $f_\theta(\cdot)$ and that $E[\sigma_1^{k+1}(\theta)]$ is bounded above zero. Solving these linear equation, we have

$$E[\zeta_1^{k+1}] = \frac{E[W_1^k W_3] - C_2}{E[\sigma_1^{k+1}(\theta)]}$$
$$E[\xi_1^{k+1}] = E[W_1^{k+1}] - E[W_1^k W_3] - (C_1 - C_2).$$

By applying the same arguments over $\left\{E[W_t^{k+1}], E[W_t^k W_{t+2}]\right\}$ for $t = 2, \ldots, 7$, we can identify $E[\zeta_t^{k+1}]$ and $E[\xi_t^{k+1}]$ for $t = 2, \ldots, 7$.

The infinite sequence of moments $\left\{\left(E[\eta_t^k], E[\xi_t^k]\right)\right\}_{k=1}^{\infty}$ is recovered by mathematical induction. □

A.3 Identification under Homoskedasticity

We next provide our identification result under homoskedasticity without requiring strict monotonicity of $\Delta\mu_t(\cdot)$. The model is given by equation (3) with $\kappa_t = \kappa_{t-1} + \eta_t$, $\nu_t = \xi_t + \beta_1 \xi_{t-1}$, and η_t and ξ_t independent of θ for all t. We assume the following conditions similar to Assumption 2.

Assumption 3. *The following conditions hold 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.

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

[33] This is known as the Hamburger moment problem, and the Carleman's condition is a sufficient condition for identification of the distributions, e.g. see Shiryaev (1995), pp. 295–296.

(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) For $t = 7, 8, 9$*, the functions* $\Delta\mu_t(\theta)$ *are continuously differentiable with* $\Delta\mu_t'(\theta^*) = 0$ *for at most a finite number of* θ^**. The density of* θ*,* $f_\theta(\cdot)$*, does not vanish in the neighborhood of* θ^**.*

(v) We impose the following normalization: $\kappa_0 = \xi_0 = 0$*,* $\mu_1(\theta) = \theta$*, and* $E[\eta_t] = E[\xi_t] = 0$*, for all* t*.*

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

Conditions (i) and (iii) are the same as in Assumption 2. Condition (ii) assumes full independence of permanent and transitory shocks with θ, ruling out heteroskedasticity. Condition (iv) allows for non-monotonic changes in unobserved skill pricing functions $\Delta\mu_t(\cdot)$, but it requires that $\mu_t(\cdot)$ is changing in later periods for all but a finite number of θ values. Condition (vi) is a standard technical assumption. This alternative set of assumptions ensures identification under homoskedasticity of the shocks.

Theorem 2. *Under Assumption 3,* $f_\theta(\cdot), \{\mu_t(\cdot)\}_{t=1}^{9}$*,* $\{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot), \beta_t\}_{t=1}^{7}$ *are identified.*

Proof of Theorem 2

We again 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 3 (ii). Begin with the following subset of equations:

$$\begin{aligned} W_1 &= \theta + \varepsilon_1 = \theta + \eta_1 + \nu_1 \\ \Delta W_4 &= \Delta\mu_4(\theta) + \Delta\varepsilon_4 = \Delta\mu_4(\theta) + \eta_4 + \Delta\nu_4 \\ \Delta W_7 &= \Delta\mu_7(\theta) + \Delta\varepsilon_7 = \Delta\mu_7(\theta) + \eta_7 + \Delta\nu_7. \end{aligned}$$

Note that, different from the baseline model (Lemma 1), Assumption 4 in Hu and Schennach (2008) described above may not hold since the difference of two monotone functions, $\Delta\mu_7(\cdot)$, is not always monotone. 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 $\bar{\theta} \neq \tilde{\theta}$, we have $\Delta\mu_7(\bar{\theta}) \neq \Delta\mu_7(\tilde{\theta})$ because both the logarithmic and exponential functions are strictly monotone. Therefore, $f_{\Delta W_7|\theta}(w|\bar{\theta}) \neq f_{\Delta W_7|\theta}(w|\tilde{\theta})$ for some w with positive probability, and we can 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:

$$W_1 = \theta + \varepsilon_1$$
$$\Delta W_7 = \Delta\mu_7(\theta) + \Delta\varepsilon_7.$$

Notice that Assumptions 1–6 in their paper are implied by Assumption 3 (i), (ii), (iv), and (v). Therefore, we can identify the function $\Delta\mu_7(\cdot)$ and densities $f_\theta(\cdot), f_{\varepsilon_1}(\cdot), f_{\Delta\varepsilon_7}(\cdot)$. To identify $\Delta\mu_4(\cdot)$, we need some additional notation. Let $L_{A|B}$ be a linear operator defined as

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

where $\mathcal{A}$ is the 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

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

Using Assumption 3 (ii), we can rewrite the conditional density $f_{\Delta W_7,W_1|\Delta W_4}$ as follows

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

which is equivalent to

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

By integrating over w_7, we have

$$L_{W_1|\Delta W_4} = L_{W_1|\theta}L_{\theta|\Delta W_4} \tag{14}$$
$$L_{\theta|\Delta W_4} = L^{-1}_{W_1|\theta}L_{W_1|\Delta W_4}, \tag{15}$$

where equation (15) is made possible from Assumption 3 (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 (15) and identify the density $f_{\theta|\Delta W_4}$. Applying Bayes' rule 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 $\mu_2(\theta) = F_{\theta_2}^{-1}\left(F_\theta\left(\theta\right)\right)$. 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 $\nu_3' = \eta_2 + \eta_3 + \nu_3$. 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 $t = 1, \ldots, 7$.

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

$$Cov(W_1, W_2) = Cov\left(\theta, \mu_2(\theta)\right) + Var(\eta_1) + \beta_1 Var(\xi_1),$$

since we know all other terms except β_1. Therefore, unless $Var(\xi_1) = 0$, we can identify β_1. In the above equation, note that all cross moments between unobservables are zero because of the conditional mean zero assumption. For example,

$$Cov(\mu_2(\theta), \eta_1) = E[\mu_2(\theta)\eta_1] = E_\theta\left[E[\mu_2(\theta)\eta_1|\theta]\right] = E_\theta\left[\mu_2(\theta)E[\eta_1|\theta]\right] = 0$$

We next identify the distribution of ξ_2 using the standard deconvolution method as

$$\phi_{\xi_2}(\tau) = \frac{\phi_{\nu_2}(\tau)}{\phi_{\beta_1\xi_1}(\tau)},$$

where $\phi_{\nu_2}(\cdot)$ and $\phi_{\beta_1\xi_1}(\cdot)$ are identified. In the same way, we expand $Cov(W_t, W_{t+1})$ and identify β_t and $f_{\xi_t}(\cdot)$ 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 Steps 1–3, we establish the identification of $f_\theta(\cdot), \{f_{\eta_t}(\cdot), f_{\xi_t}(\cdot), \beta_t\}_{t=1}^{7}$, and $\{\mu_t(\cdot)\}_{t=1}^{9}$. $\square$

B Additional Estimates

In this appendix, we report results from a number of alternative specifications, all assuming linear $\mu_t(\cdot)$. Unless otherwise noted, model assumptions are the same as those of subsection 5.1. Consistent with the results of that subsection, we also use only variance and covariance moments in estimation. We begin with an examination of the robustness of our results for log earnings residuals, then consider results from our baseline specification for log weekly wage residuals.

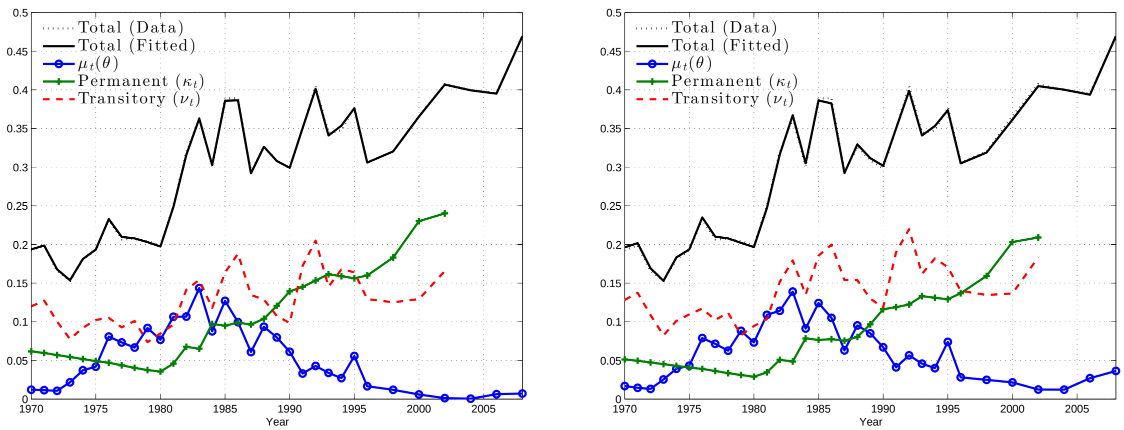
B.1 Additional Specifications for Log Earnings Residuals

B.1.1 Different Processes for ν_t

We first consider estimates when $\nu_{a,t}$ follows a $MA(1)$ or $MA(5)$ process. See Figure 14. The $MA(1)$ model slightly over-predicts the relative importance of permanent shocks, since some of

Figure 14: Variance Decomposition Assuming ν_t is $MA(1)$ or $MA(5)$ (Linear $\mu_t(\cdot)$)

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



the persistence attributed to 'transitory' shocks in the more general $MA(3)$ and $MA(5)$ models 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 the process for transitory shocks.

We next consider the case when $\nu_{a,t}$ follows an $ARMA(1,1)$ stochastic process: $\nu_{a,t} = \rho\nu_{a-1,t-1} + \xi_{a,t} + \beta_t \xi_{a-1,t-1}$. With this assumption,

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

due to the mutual independence of $\xi_{i,a,t}$ across time. For $k = 2$, this expression defines the variance of the 'transitory' component in our variance decompositions. Other variance components are unchanged, so

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

and

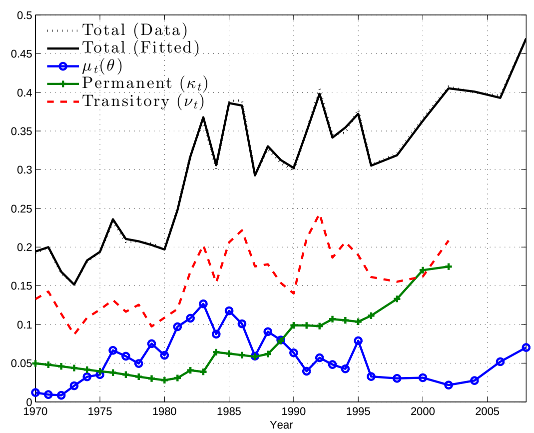
$$E[W_{a,t}W_{a+l,t+l}|a,t,l] = \sum_{j=0}^{p}\sum_{j'=0}^{p} m_{j,t}m_{j',t+l}E[\theta^{j+j'}]+\sum_{j=0}^{a-1}\sigma_{\eta_{t-j}}^2+\rho^{l-1}(\rho+\beta_{t+1})\sigma_{\xi_t}^2+\rho^l\sum_{j=0}^{a-2}\rho^{2j}(\rho+\beta_{t-j})^2\sigma_{\xi_{t-j-1}}^2$$

for $l \geq 1$.

Table 2: Estimates Assuming $\nu_t \sim ARMA(1,1)$ 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
	(0.031)	(0.056)
β	-0.529	-0.496
	(0.037)	(0.058)

Figure 15: Variance Decomposition Assuming $\nu_t \sim ARMA(1,1)$ (Linear $\mu_t(\cdot)$)



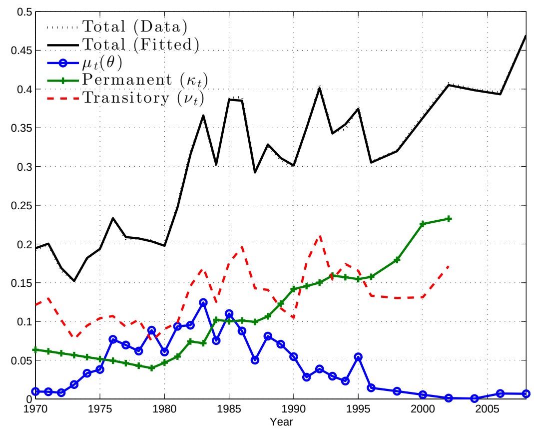
Using variance/covariance moments and assuming linear $\mu_t(\cdot)$, we estimate this model assuming $\beta_t = \beta$ for all t. Results from this model are reported in Table 2 along with analogous results assuming time invariant pricing functions (i.e. $\mu_t(\theta) = \theta$ for all t).[34] Figure 15 shows the variance decomposition associated with these estimates (for time-varying $\mu_t(\cdot)$). As with the $MA(1)$ and $MA(5)$ cases, the decomposition results are quite similar to those in Figure 5 of the paper.

B.1.2 Age-Dependent Variances of Permanent and Transitory Shocks

We next consider the possibility that the variance of transitory and permanent shocks may vary with age. Specifically, we allow the variances of both $\eta_{a,t}$ and $\xi_{a,t}$ to be linear functions of age over

[34] Notice that the $ARMA(1,1)$ specification has a slightly lower minimized objective function than does our baseline $MA(3)$ model. However, we cannot reject that adding an autoregressive component to the $MA(5)$ model improves the fit (at the 5% significance level). The minimized objective function for an $ARMA(1,5)$ is 114.1.

Figure 16: Variance Decomposition with Age-Dependent Variances of $\eta_{a,t}$ and $\xi_{a,t}$ (Linear $\mu_t(\cdot)$)



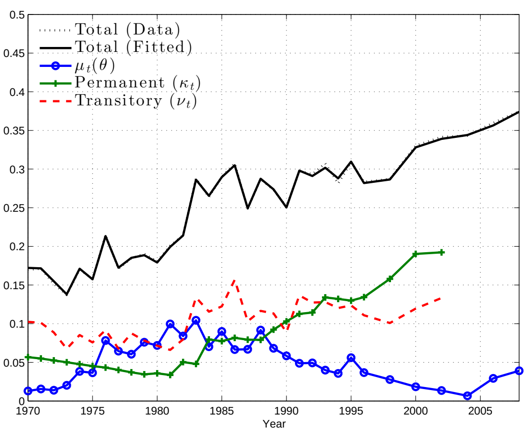
ages 20-35 and constant thereafter. Estimates imply that the variance of permanent shocks declines by a total of 26% over these fifteen years, while the variance of transitory shocks declines by 24%. The minimized objective function improves (insignificantly) to 118.5. The variance decomposition shown in Figure 16 is quite similar to that of Figure 5 in the text.

B.2 Log Weekly Wage Residuals

Figure 17 shows the estimated variance decomposition for log weekly wage residuals using our baseline specification with linear time-varying $\mu_t(\cdot)$, an $MA(3)$ process for $\nu_{a,t}$, and homoskedastic permanent shocks. This figure is quite similar to the analogous figure for log earnings residuals (Figure 5). Estimated persistence of transitory shocks are also quite similar to those in column 6 of Table 1 with $\hat{\beta}_1 = 0.282$ (0.028), $\hat{\beta}_2 = 0.162$ (0.021) and $\hat{\beta}_3 = 0.124$ (0.028). As with log earnings residuals, accounting for time-varying $\mu_t(\cdot)$ functions substantially improves the fit to the data.[35]

[35]The minimized objective function is 85.65 for time varying $\mu_t(\cdot)$ compared to 110.12 for time invariant $\mu_t(\cdot)$.

Figure 17: Variance Decomposition for Log Weekly Wage Residuals (Baseline Specification)



References

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

Acemoglu, D. (1999). Changes in unemployment and wage inequality: An alternative theory and some evidence. *American Economic Review 89*(5), 1259–1278.

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., F. Levy, and R. Murnane (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics 118*(4), 1279–1333.

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.

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2011). The contribution of dispersion across plants to the increase in us earnings dispersion. Working Paper.

Blundell, R., M. Graber, and M. Mogstad (2013). Labor income dynamics and the insurance from taxes, transfers, and the family. Working Paper.

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

Card, D. and T. Lemieux (1994). Changing wage structure and black–white wage differentials. *American Economic Review 84*(2), 29–33.

Caselli, F. (1999). Technological revolutions. *American Economic Review 89*(1), 78–102.

Chay, K. Y. and D. S. Lee (2000). Changes in relative wages in the 1980s: Returns to observed and unobserved skills and black-white wage differentials. *Journal of Econometrics 99*(1), 1–38.

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

Dunne, T., L. Foster, J. Haltiwanger, and K. R. Troske (2004). Wage and productivity dispersion in united states manufacturing: The role of computer investment. *Journal of Labor Economics 22*(2), 397–429.

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

Galor, O. and O. Moav (2000). Ability-biased technological transition, wage inequality, and economic growth. *Quarterly Journal of Economics 115*(2), 469–497.

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.

Gould, E., O. Moav, and B. Weinberg (2001). Precautionary demand for education, inequality, and technological progress. *Journal of Economic Growth 6*(4), 285–315.

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. and E. Vytlacil (2001). Identifying the role of cognitive ability in explaining the level of and change in the return to schooling. *Review of Economics and Statistics 83*(1), 1–12.

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.

Hryshko, D. (2012). Labor income profiles are not heterogeneous: Evidence from income growth rates. *Quantitative Economics 3*(2), 177–209.

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.

Kambourov, G. and I. Manovskii (2009). Occupational mobility and wage inequality. *Review of Economic Studies 76*(2), 731–759.

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.

Krueger, D. and F. Perri (2006). Does income inequality lead to consumption inequality? evidence and theory. *Review of Economic Studies 73*(1), 163–193.

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.

Murnane, R., J. Willett, and F. Levy (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics 77*(2), 251–266.

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.

Shiryaev, A. N. (1995). *Probability* (Second ed.). Secaucus, NJ, USA: Springer-Verlag.

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.

Violante, G. (2002). Technological acceleration, skill transferability, and the rise in residual inequality. *Quarterly Journal of Economics 117*(1), 297–338.

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