Wage Dynamics and Returns to Unobserved Skill*

Lance Lochner University of Western Ontario

Youngmin Park Bank of Canada

Youngki Shin McMaster University

October 3, 2017

Abstract

Economists disagree about the factors driving the substantial increase in residual wage inequality in the U.S. over the past few decades. We identify and estimate a general model of log wage residuals that incorporates: (i) changing returns to unobserved skills, (ii) a changing distribution of unobserved skills, and (iii) changing volatility in wages due to factors unrelated to skills. Using data from the PSID, we estimate that the returns to unobserved skills have declined by as much as 50% since the mid-1980s despite a sizeable increase in residual inequality. Instead, the variance of skills rose over this period due to increasing variability in lifecycle skill growth. Finally, we develop an assignment model of the labor market and show that both demand and supply factors contributed to the downward trend in the returns to skills over time, with demand factors dominating for non-college men.

1 Introduction

The U.S. has experienced substantial and sustained growth in wage inequality since the 1960s. In addition to the long-run increase in wage differentials across workers with different levels of education and experience, inequality within narrowly defined groups (e.g. by race, education, and age/experience) also rose dramatically (see, e.g., Katz and Autor, 1999; Autor, Katz, and Kearney, 2008). While the interpretation of the latter trend is not fully understood, its underlying cause is economically important. Whether it reflects an increase in the returns to unobserved skills, variance of unobserved skills, or variance of short-term volatility unrelated to skills (or measurement error) is critical to our understanding of both the causes and welfare consequences of rising inequality.

*For their valuable comments, we thank seminar/conference participants at the University of Rochester, University of Maryland, and 2017 Annual Meeting of the Society of Economic Dynamics. Lochner acknowledges generous support from SSHRC.

Since the seminal work of Juhn, Murphy, and Pierce (1993), many economists (implicitly or explicitly) equate the rising within-group, or residual, inequality with an increase in the returns to unobserved ability or skill (see, e.g., Card and Lemieux, 1996; Katz and Autor, 1999; Acemoglu, 2002; Autor, Katz, and Kearney, 2008). Indeed, this interpretation, along with the rising returns to observable skill, motivated an enormous and still influential literature on skill-biased technical change (SBTC).[1]

Challenging the conventional view, Lemieux (2006) demonstrates that the rise in residual inequality is at least partially explained by an increase in the variance of unmeasured skills resulting from composition changes in the labor market, especially in the late 1980s and 1990s, as the workforce shifted increasingly to older and more educated workers who exhibit greater within-group inequality. Lemieux (2006) and Gottschalk and Moffitt (2009) further argue that increasing measurement error and short-term volatility in wages may have also contributed to rising residual inequality.

A few studies have turned to richer data to incorporate additional measures of skill or occupational tasks, directly estimating their effects on wages at different points in time. Using the 1979 and 1997 Cohorts of the National Longitudinal Surveys of Youth (NLSY), Castex and Dechter (2014) estimate that the wage returns to cognitive achievement, as measured by the Armed Forces Qualifying Test (AFQT), declined substantially between the 1980s and early 2000s. By contrast, Deming (forthcoming) estimates that the returns to social skills have risen across these two cohorts. Among others, Autor, Levy, and Murnane (2003) and Autor and Dorn (2013) document a decline in demand for middle-skill workers caused by the automation of routine tasks, which has led to a fall in the wages for workers in many middle-skill relative to low- and high-skill occupations, dubbed 'polarization'. Caines, Hoffmann, and Kambourov (2017) instead argue that occupational task complexity has become a stronger determinant of wages in recent years, more so than routineness.

While efforts to better measure skills and job tasks have greatly enriched our understanding of wage inequality, much of the cross-sectional variation in wages remains unexplained in these studies. More importantly, difficult measurement challenges have led to strong (often implicit) assumptions on the evolution of skills over the lifecycle and across time. For example, Castex and Dechter (2014) and Deming (forthcoming) examine the effects of pre-market skills on wages ignoring subsequent lifecycle skill accumulation that may vary across workers and over time. Because the vast majority of studies taking a task-based approach do not use individual-level data on skills or job tasks, they implicitly assume that worker skills and tasks within each occupation are time invariant and attribute all time variation in wages across occupations to changes in the returns to skills/tasks.[2]

[1]Many of these studies aimed specifically to explain rising residual inequality and returns to unobserved ability/skill (e.g., Galor and Tsiddon, 1997; Acemoglu, 1999; Caselli, 1999; Galor and Moav, 2000; Violante, 2002). Card and DiNardo (2002) question the influence of SBTC on the overall wage structure based largely on the failure of SBTC to explain the evolution of important between-group differences (largely across race and gender) in wages.

[2]Using data with individual-level measures of job tasks, Spitz-Oener (2006) shows that most task changes in Germany

Most of the literature studying the evolution of residual wage inequality relies on repeated samples of cross-sectional data, which makes it difficult to sort out changes in skill returns vs. the distribution of skills over time. In this paper, we show how panel data on wages can be used to separately identify the evolution of (i) returns to unobserved skills, (ii) distributions of unobserved skills and skill growth rates, and (iii) the volatility of transitory shocks unrelated to skills. Building on the literature on earnings dynamics, our key source of identification (long-run covariances in wage residuals) motivates a simple instrumental variable strategy for estimating the returns to skill over time, even when lifecycle skill growth varies systematically across individuals and is subject to time-varying idiosyncratic shocks.[3] This approach also allows for a very general structure for transitory non-skill shocks. Importantly, there is no need to observe anything about what workers do on their jobs, enabling our approach in most widely available panel data sets.

We estimate the evolution of returns to unobserved skills, variances of unobserved skills, and variances of the non-skill component of wages from 1970-2012 using data on log hourly wages for men ages 16-64 in the Panel Study of Income Dynamics (PSID), performing separate analyses by college attendance. Our main finding is that the returns to unobserved skills were relatively stable from 1970 to the mid-1980s, then *fell* considerably through the 1990s, stabilizing thereafter. The decline in returns was more dramatic for non-college workers, consistent with the recent literature on polarization (Autor, Levy, and Murnane, 2003; Autor, Katz, and Kearney, 2008; Acemoglu and Autor, 2011; Autor and Dorn, 2013). The estimated declines in returns to unobserved skill are robust to different estimation strategies and to very general specifications regarding heterogeneity across cohorts and experience levels. The flip side of declining returns is that the variance of unobserved skills rose substantially, driving the increase in the residual variance of log wages. The widening skill distribution is largely explained by an increase in the variance of lifecycle skill growth over time, not in the variance of initial skill levels across cohorts.[4] We further show that the increasing variance of skill growth reflects increases in the variances of both idiosyncratic skill growth shocks and heterogeneous systematic lifecycle skill growth as in the heterogeneous income profile (HIP) framework of Baker and Solon (2003) and Guvenen (2009).

over the 1980s and 1990s actually occurred within occupations. Autor and Handel (2013) document substantial heterogeneity in worker tasks within occupations that are powerful predictors of wage variation (within occupational, demographic, and education groups). Their finding that person-level job tasks vary systematically across demographic groups within narrowly defined occupations suggests that worker skills and job tasks are unlikely to be time invariant within occupations as the composition of the workforce changes.

[3]The voluminous literature on earnings dynamics focuses on a different question from ours: identifying the changing importance of permanent vs. transitory shocks in earnings and the resulting implications for consumption and wealth inequality (e.g., Gottschalk and Moffitt, 1994; Blundell and Preston, 1998; Haider, 2001; Moffitt and Gottschalk, 2002; Meghir and Pistaferri, 2004; Bonhomme and Robin, 2010; Heathcote, Storesletten, and Violante, 2010; Heathcote, Perri, and Violante, 2010; Moffitt and Gottschalk, 2012; Blundell, Graber, and Mogstad, 2015). An important conclusion from this literature is that a considerable share of the rise in residual earnings inequality over the 1980s is attributable to a growing variance in short-term fluctuations that is unlikely to have much to do with the returns to unobserved skills (Gottschalk and Moffitt, 2009).

[4]Since we estimate widening skill growth distributions within education and experience groups, these trends are not accounted for in the composition adjustments of Lemieux (2006).

These results highlight the importance of accounting for changes in the distribution of unobserved skills over time. Our estimated time patterns for returns to unobserved skill are fundamentally different from those estimated in previous studies assuming time-invariant unobserved skill distributions (e.g., Juhn, Murphy, and Pierce, 1993; Moffitt and Gottschalk, 2012). They are more consistent with the falling returns to AFQT between the 1980s and early 2000s as estimated by Castex and Dechter (2014).

To interpret our empirical findings, we develop a simple demand and supply framework based on the assignment model of Sattinger (1979). In our model, the returns to skills are determined by the assignment of workers with heterogeneous skills to jobs with heterogeneous productivity (e.g. different capital quality, tasks), as well as the technology of production itself. More skilled workers earn more partly because they work at more productive jobs. As a result, the skill distribution affects the return to skill by changing the equilibrium assignment. An increase in the variance of skills reduces the returns to skills by shrinking the productivity differential across jobs among workers of different skill levels. We recover the demand and supply factors by combining our estimates of the returns to skills and variance of skills with key restrictions implied by the model. Our estimates suggest that both demand and supply factors played important roles in the declining returns to unobserved skills since the mid-1980s for both college and non-college workers. The decline in skill demand was a more important factor than supply shifts for non-college workers, consistent with the automation of routine tasks in middle-skill jobs as emphasized by Autor, Levy, and Murnane (2003), Autor and Dorn (2013), and many others.

This paper proceeds as follows. In Section 2, we provide identification results for our model where the returns to unobserved skills, the variance of unobserved skills, and the variance of non-skill component of earnings change over time. Section 3 describes the PSID data used to estimate earnings dynamics for American men. Sections 4 and 5 report our empirical findings. Section 6 develops our assignment model of the labor market and uses it to interpret the evolution of estimated returns to unobserved skills in terms of changes in demand and supply. Section 7 concludes.

2 Identifying the Returns to and Distributions of Unobserved Skills

In this section, we describe a general specification for wages that is consistent with much of the empirical literature on residual wage inequality and our theoretical framework in Section 6. We then establish conditions under which the time series for the returns to unobserved skills, the variances of unobserved skills, and the variances of transitory non-skill shocks (or measurement error) are identified.

2.1 Log Wage Functions

We consider the following specification for log wages motivated by the literature on unobserved skills (e.g., Juhn, Murphy, and Pierce, 1993; Lemieux, 2006):

$$\ln W_{i,t} = f_t(x_{i,t}) + \underbrace{\mu_t \theta_{i,t} + \varepsilon_{i,t}}_{\equiv w_{i,t}}, \tag{1}$$

where $W_{i,t}$ reflects wages, $x_{i,t}$ observed characteristics (e.g. education, race, experience), and $\theta_{i,t}$ (log) unobserved skill for individual i in period t. The time-varying function $f_t(\cdot)$ incorporates effects of observable characteristics on period t log wages, while the 'residual' $w_{i,t} \equiv \mu_t \theta_{i,t} + \varepsilon_{i,t}$ reflects the contributions of unobserved skills and idiosyncratic non-skill shocks $\varepsilon_{i,t}$ (including measurement error).[5] In Section 6, we develop and study an assignment model of the labor market that produces log wage functions of this form.[6]

The log wage residual $w_{i,t}$ is the primary focus of our efforts to identify and estimate the evolution of 'returns' to unobserved skill, distributions of unobserved skill, and volatility of non-skill shocks.

2.2 Identification

Suppose that we observe log wage residuals from equation (1) for a large number of individuals $i = 1, \ldots, N$ for periods $t = \underline{t}, \ldots, \bar{t}$:

$$w_{i,t} = \mu_t \theta_{i,t} + \varepsilon_{i,t}, \tag{2}$$

where $\theta_{i,t}$ represents unobserved (log) skill, μ_t the period t return to unobserved skill, and $\varepsilon_{i,t}$ idiosyncratic shocks. Since $w_{i,t}$ is a mean zero residual, we normalize $\theta_{i,t}$ and $\varepsilon_{i,t}$ so that both are mean zero for all t. This implies that unobserved skill growth innovations,

$$\nu_{i,t} \equiv \theta_{i,t} - \theta_{i,t-1}$$

[5]The separability between $x_{i,t}$ and $\theta_{i,t}$ is both common and convenient, though not necessary. One could simply condition the analysis that follows on $x_{i,t}$. Indeed, much of our empirical analysis separately studies non-college and college educated workers.

[6]Notice that wage levels are non-linear in unobserved skill. As such, variation in μ_t over time is inconsistent with perfect substitutability across workers of different skill levels, since perfect substitutability would imply log wages functions that are additively separable in 'prices' and skills.

are also mean zero for all t.[7] We further assume that $\mathrm{Cov}(\theta_t, \varepsilon_{t'}) = \mathrm{Cov}(\nu_t, \varepsilon_{t'}) = 0$ for all t, t'.[8] Thus, any shocks related to skills are embedded in $\theta_{i,t}$, while $\varepsilon_{i,t}$ reflects factors unrelated to skills. We note that individuals may come from different cohorts (i.e. different years of labor market entry), which we discuss further below.

Central to our approach is the classical idea of Friedman and Kuznets (1954) that earnings consist of a permanent component related to skills and a transitory component that reflects short-run variation unrelated to skills. Although the transitory component, which may include measurement error, can be serially correlated, the correlation between transitory components far apart in time is likely to be quite small. Taking this to the limit, we assume that there exists some $k > 0$ such that $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$ for $|t' - t| \geq k$, so the "long" autocovariance of wage residuals reflects only the skill-related component (Carroll, 1992; Moffitt and Gottschalk, 2012):

$$\mathrm{Cov}(w_t, w_{t'}) = \mu_t \mu_{t'} \mathrm{Cov}(\theta_t, \theta_{t'}), \qquad \text{for } |t' - t| \geq k. \tag{3}$$

In short, we assume that persistent long-term differences in wages across workers are driven by lasting differences in skills. Appendix A.1 shows that our main identification results continue to hold when the 'transitory' component $\varepsilon_{i,t}$ contains an autoregressive component, such that the serial correlation in non-skill shocks depreciates exponentially over time but never fully disappears. Empirical results are also quite similar in this case.

The properties of skill growth innovations, $\nu_{i,t}$, are important for identification. We begin by assuming that these innovations are serially uncorrelated, then show how systematic heterogeneity in skill growth rates can be incorporated.

2.2.1 Serially Uncorrelated Skill Growth Shocks

First, consider the case in which unobserved skill growth innovations $\nu_{i,t}$ are uncorrelated with past skill levels and growth shocks, i.e. $\mathrm{Cov}(\theta_{t'}, \nu_t)$ for all $t' < t$ and $\mathrm{Cov}(\nu_t, \nu_{t'}) = 0$ for all $t \neq t'$. This implies that skill levels are persistent in the sense that, for all $t' \leq t$, $E(\theta_{i,t}|\theta_{i,t'}) = \theta_{i,t'}$ and $\mathrm{Cov}(\theta_t, \theta_{t'}) = \mathrm{Var}(\theta_{t'})$. The latter is particularly useful, since it implies that the following ratio of residual covariances identifies the ratio of skill returns:

$$\frac{\mathrm{Cov}(w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\mu_t \mu_{t'} \mathrm{Var}(\theta_{t'})}{\mu_{t-1} \mu_{t'} \mathrm{Var}(\theta_{t'})} = \frac{\mu_t}{\mu_{t-1}}, \qquad \text{for } t' < t - k. \tag{4}$$

[7] Average skill growth rates, which may vary by observable characteristics, are reflected in changes in $f_t(x_{i,t})$.

[8] Let x_t be a random variable and its realization for individual i be $x_{i,t}$. Denote its cross-sectional second moments by $\mathrm{Var}(x_t)$ and $\mathrm{Cov}(x_t, x_{t'})$.

This suggests that μ_t/μ_{t-1} can be easily estimated by regressing $w_{i,t}$ on $w_{i,t-1}$ using sufficiently lagged $w_{i,t'}$ as an instrumental variable (IV).

An IV estimation approach can also be motivated using the framework of Holtz-Eakin, Newey, and Rosen (1988). We can substitute in for $\theta_{i,t}$ in equation (2) to obtain an expression for $w_{i,t}$ in terms of $w_{i,t-1}$:

$$w_{i,t} = \mu_t \left(\frac{w_{i,t-1} - \varepsilon_{i,t-1}}{\mu_{t-1}} + \nu_{i,t} \right) + \varepsilon_{i,t} = \frac{\mu_t}{\mu_{t-1}} w_{i,t-1} + \left(\varepsilon_{i,t} - \frac{\mu_t}{\mu_{t-1}} \varepsilon_{i,t-1} + \mu_t \nu_{i,t} \right), \tag{5}$$

suggesting that lagged residuals $w_{i,t-1}$ might serve as a proxy for unobserved skills. However, because $w_{i,t-1} = \mu_{t-1}\theta_{i,t-1} + \varepsilon_{i,t-1}$ is a 'noisy' measure of unobserved skill $\theta_{i,t}$, it is correlated with the measurement error $\varepsilon_{i,t-1}$ (and $\varepsilon_{i,t}$ if $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1}) \neq 0$). Simply regressing $w_{i,t}$ on $w_{i,t-1}$ would, therefore, produce a biased estimate of μ_t/μ_{t-1}. To address this problem, wage residuals from the distant past (i.e. any $w_{i,t'}$ for $t' < t - k$) can be used as instrumental variables, since they are correlated with $w_{i,t-1}$ (through unobserved skills) but uncorrelated with $\varepsilon_{i,t-1}$, $\varepsilon_{i,t}$, and $\nu_{i,t}$. Therefore, μ_t/μ_{t-1} can be estimated using IV methods for all but the first few sample years, i.e. $t > \underline{t} + k$.

Future wage residuals are not valid instruments in equation (5), because skill growth has permanent effects on future skills. Indeed, IV regression using future wage residuals as instruments produces an upward biased estimate of μ_t/μ_{t-1} with the bias proportional to the variance of skill growth:

$$\frac{\mathrm{Cov}(w_t, w_{t''})}{\mathrm{Cov}(w_{t-1}, w_{t''})} = \frac{\mu_t \mu_{t''} \mathrm{Var}(\theta_t)}{\mu_{t-1}\mu_{t''} \mathrm{Var}(\theta_{t-1})} = \frac{\mu_t}{\mu_{t-1}} \left[1 + \frac{\mathrm{Var}(\nu_t)}{\mathrm{Var}(\theta_{t-1})} \right] \quad \text{for } t'' \geq t + k. \tag{6}$$

Since IV estimates using past residuals consistently estimate μ_t/μ_{t-1}, the ratio of IV estimates using future vs. past residuals as instruments can be used to identify the importance of skill growth shocks (relative to variation in lagged skill levels):

$$\frac{\mathrm{Var}(\nu_t)}{\mathrm{Var}(\theta_{t-1})} = \frac{\mathrm{Cov}(w_t, w_{t''})/\mathrm{Cov}(w_{t-1}, w_{t''})}{\mathrm{Cov}(w_t, w_{t'})/\mathrm{Cov}(w_{t-1}, w_{t'})} - 1, \quad \text{for } t' + k + 1 \leq t \leq t'' - k.$$

The bias for μ_t/μ_{t-1} when using future residuals as instruments poses an identification challenge for μ_t in early sample periods when past observations are not available to serve as instruments. Fortunately, two additional conditions enable estimation of μ_t/μ_{t-1} for early years by differencing out this bias across cohorts. To see this, let c reflect the period of labor market entry ('cohort'), and suppose that two cohorts exist, c and $\tilde{c}$, such that $\mathrm{Var}(\theta_{t-1}|c) \neq \mathrm{Var}(\theta_{t-1}|\tilde{c})$ and $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|\tilde{c})$. The first condition is likely to hold quite generally. For example, differences in the variance of initial skill levels would contribute to different variances later in life. Even if initial skill distributions were identical across cohorts, the older cohort is likely to have accumulated more skill growth innovations over its longer career. The second condition holds when the skill growth variance depends only on time (and not experience) or when there is a non-monotonic experience trend in the variance of skill changes. For example, young workers may experience greater variation in skill growth than middle age workers due to differences in training or learning opportunities, while older workers may have a greater variance in skill changes due to differences in health shocks or skill obsolescence. Indeed, Baker and Solon (2003) and Blundell, Graber, and Mogstad (2015) estimate a U-shaped age profile for the variance of earnings shocks. In this case, the ratio of the difference of the long autocovariances between cohorts identifies the ratio of skill returns for early periods:

$$\frac{\mathrm{Cov}(w_t, w_{t'}|c) - \mathrm{Cov}(w_t, w_{t'}|\tilde{c})}{\mathrm{Cov}(w_{t-1}, w_{t'}|c) - \mathrm{Cov}(w_{t-1}, w_{t'}|\tilde{c})} = \frac{\mu_t \mu_{t'}\left[\mathrm{Var}(\theta_{t-1}|c) - \mathrm{Var}(\theta_{t-1}|\tilde{c})\right]}{\mu_{t-1}\mu_{t'}\left[\mathrm{Var}(\theta_{t-1}|c) - \mathrm{Var}(\theta_{t-1}|\tilde{c})\right]} = \frac{\mu_t}{\mu_{t-1}}, \quad \text{for } t' \geq t+k. \tag{7}$$

With this, we can identify the sequence of all μ_t over the sample period, where we must normalize one skill return in a single time period, say t^*. Normalizing $\mu_{t^*} = 1$ effectively sets the units for unobserved skill in terms of dollars per hour of work in year t^*.

Once the sequence of all μ_t have been identified, the variances of unobserved skills for all $t \leq \bar{t} - k$ are identified from the covariance between current and future wage residuals $\mathrm{Var}(\theta_t) = \mathrm{Cov}(w_t, w_{t'})/(\mu_t \mu_{t'})$ for $t' \geq t+k$. Although the return to unobserved skill is identified for all periods, the variance of unobserved skills is not identified for later periods $t > \bar{t} - k$ (without further assumptions), because it is impossible distinguish between unobserved skills and wage shocks without observing future wage residuals. Having identified the variance of unobserved skills over time, it is straightforward to then identify $\mathrm{Var}(\nu_t) = \mathrm{Var}(\theta_t) - \mathrm{Var}(\theta_{t-1})$ and $\mathrm{Var}(\varepsilon_t) = \mathrm{Var}(w_t) - \mu_t^2 \mathrm{Var}(\theta_t)$.

We summarize the above discussion in the following Proposition, where we also note that the variances of unobserved skills, skill growth shocks, and non-skill transitory shocks can be identified separately by cohort.

Proposition 1. *Assume: (i) there exists* $k > 0$ *such that* $\bar{t} - \underline{t} \geq 2k$ *and* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = 0$ *for all* (c, t, t') *such that* $|t' - t| \geq k$*, (ii)* $\mathrm{Cov}(\varepsilon_t, \theta_c|c) = \mathrm{Cov}(\varepsilon_t, \nu_{t'}|c) = 0$ *for all* (c, t, t')*, (iii)* $\mathrm{Cov}(\nu_t, \theta_{t-j}|c) = 0$ *for all* (c, t)*, and* $j > 0$*, and (iv)* $\mathrm{Var}(\theta_{t-1}|c) \neq \mathrm{Var}(\theta_{t-1}|\tilde{c})$ *and* $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|\tilde{c})$ *for some* $c \neq \tilde{c}$*. Then, (i)* μ_t *is identified for all* t *up to a normalization* $\mu_{t^*} = 1$ *for some period* t^**, (ii)* $\mathrm{Var}(\theta_t|c)$ *and* $\mathrm{Var}(\varepsilon_t|c)$ *are identified for all* (c, t) *such that the cohort* c *is observed both in period* t *and some other period* $t' \geq t+k$*, and (iii)* $\mathrm{Var}(\nu_t|c)$ *is identified for all* (c, t) *such that* $\mathrm{Var}(\theta_t|c)$ *and* $\mathrm{Var}(\theta_{t-1}|c)$ *are identified.*

Our identification strategy has relied on the assumption that non-skill shocks, $\varepsilon_{i,t}$, become serially uncorrelated when observations are far enough apart. This is not critical; although, identification is most transparent in this case. Appendix A.1 provides an analogous identification analysis when $\varepsilon_{i,t}$ follows an $ARMA(1, q)$ process, such that the serial correlation in non-skill shocks never fully disappears.

2.2.2 Heterogeneity in Lifecycle Skill Growth

We now consider the possibility that unobserved skill growth innovations $\nu_{i,t}$ may be correlated over time as in the HIP models estimated in, for example, Haider (2001), Guvenen (2009), and Moffitt and Gottschalk (2012). We consider a flexible process governing this skill growth heterogeneity, assuming

$$\nu_{i,t} = \tau_t(c_i)\delta_i + \tilde{\nu}_{i,t}, \tag{8}$$

where δ_i is a mean zero individual-specific lifecycle growth rate factor and the $\tau_t(c) \geq 0$ terms allow for variation in systematic skill growth across time and cohorts/experience. If we let ψ_i reflect the initial skill level for an individual just entering the labor market, then the level of unobserved skill for individual i from cohort c_i in year t can be written as

$$\theta_{i,t} = \psi_i + \sum_{j=0}^{t-c_i-1} \tau_{t-j}(c_i)\delta_i + \sum_{j=0}^{t-c_i-1} \tilde{\nu}_{i,t-j}.$$

We assume that idiosyncratic skill growth shocks $\tilde{\nu}_{i,t}$ are serially uncorrelated and uncorrelated with initial skills ψ_i and systematic skill growth δ_i; however, we make no assumptions about the correlation between heterogeneous skill growth rates δ_i and initial skill levels ψ_i. We continue to assume that non-skill shocks $\varepsilon_{i,t}$ are uncorrelated with all skill-related components ψ_i, δ_i, and $\tilde{\nu}_{i,t'}$ for all t, t'. Altogether, these assumptions imply the following variance of skills:

$$\mathrm{Var}(\theta_t|c) = \mathrm{Var}(\psi|c) + \left(\sum_{j=0}^{t-c-1} \tau_{t-j}(c)\right)^2 \mathrm{Var}(\delta|c) + 2\sum_{j=0}^{t-c-1} \tau_{t-j}(c)\,\mathrm{Cov}(\psi,\delta|c) + \sum_{j=0}^{t-c-1} \mathrm{Var}(\tilde{\nu}_{t-j}|c). \tag{9}$$

This includes two new terms reflecting (i) the variance of accumulated (systematic) skill growth and (ii) the covariance between this accumulated skill growth and initial skills. The covariance between skills in periods t and $t' < t$ can be written as

$$\mathrm{Cov}(\theta_t, \theta_{t'}|c) = \mathrm{Var}(\theta_{t'}|c) + \sum_{j=t'+1}^{t} \tau_j(c)\,\mathrm{Cov}(\theta_{t'}, \delta|c) \tag{10}$$

Unless $\sum_{j=t'+1}^{t} \tau_j(c) = 0$, it is clear that $\mathrm{Cov}(\theta_t, \theta_{t'}|c)$ will not generally equal $\mathrm{Var}(\theta_{t'}|c)$, and the IV approach described in the previous subsection (see equation (4)) cannot be used to identify/estimate μ_t/μ_{t-1}.[9] An additional assumption is needed.

[9]Notice that $\mathrm{Cov}(\theta_{t'}, \delta|c) = \left(\sum_{j=0}^{t-c-1} \tau_{t-j}(c)\right) \mathrm{Var}(\delta|c) + \mathrm{Cov}(\psi, \delta|c)$, which can be negative very early in worker's careers

Human capital theory (Becker, 1964; Ben-Porath, 1967) predicts that skill investment and accumulation should be negligible as workers approach the end of their careers, a prediction confirmed by the lack of wage growth among most older workers. While assuming zero skill growth among older workers would enable identification of μ_t/μ_{t-1}, such an assumption is stronger than needed. Instead, it is sufficient to assume that there is no unobserved heterogeneity in skill growth among the most experienced workers, i.e. $\tau_t(c) = 0$ for all (t, c) satisfying $e = t - c \geq \bar{e}$.[10] This assumption implies that skill innovations are serially uncorrelated and that $\text{Cov}(\theta_t, \theta_{t'}|c) = \text{Var}(\theta_{t'}|c)$ for workers with $e \geq \bar{e}$. Based on the analysis of Section 2.2.1, the returns to skill can be identified and estimated by following these workers over time.[11] No other assumptions are needed regarding the structure of skill growth heterogeneity across time or cohort/experience (i.e. $\tau_t(c)$) over the rest of the lifecycle. In practice, many cohorts may be needed to recover a long time series of skill returns μ_t by using overlapping subsamples of sufficiently experienced workers. See Appendix A.2 for details on identification of the full model.

3 PSID Data

To estimate the evolution of returns to unobserved skills, variances of unobserved skills and skill growth innovations, and variances of transitory non-skill shocks (including measurement error), we utilize data for American men from the PSID. 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 2013. Since earnings were collected for the year prior to each survey, our analysis studies hourly wages from 1970 to 2012. (Key findings also hold for annual earnings.)[12]

Our sample is restricted to male heads of households from the core (SRC) sample.[13] We use earnings (total wage and salary earnings, excluding farm and business income) from any year these men were ages 16-64, had potential experience of 1-40 years, had positive wage and salary income, had positive

if initial skill levels and skill growth rates are negatively correlated.

[10]This assumption is weaker than the assumption of zero skill growth for older workers used to identify skill price patterns in Heckman, Lochner, and Taber (1998) and Bowlus and Robinson (2012) and the distribution of human capital shocks in Huggett, Ventura, and Yaron (2011). Our assumption accommodates any average level of skill growth, as well as systematic differences in skill growth based on observable characteristics like education, experience, and time.

[11]As discussed in Appendix A.2, the experience level $\bar{e}$ above which $\tau_t(c) = 0$ must leave enough years prior to retirement to identify skill returns based on workers with $e \geq \bar{e}$. For example, μ_t/μ_{t-1} for $t > \underline{t} + k$ is identified if $\bar{e} + k$ is less than the maximum experience level used in the analysis.

[12]Results available upon request. Also, see Lochner and Shin (2014) for similar results using a slightly different specification.

[13]We exclude those from any PSID oversamples (SEO, Latino) as well as those with 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 Moffitt and Gottschalk (2012).

hours worked, and were not enrolled as a student. We calculate the wage measure we use in our analysis by dividing annual earnings by annual hours worked, trimming the top and bottom 1% of all wages within year and sector by ten-year experience cells. The resulting sample contains 3,766 men and 44,547 person-year observations – roughly 12 observations for each individual.

Our sample is composed of 92% whites, 6% blacks, and 1% hispanics with an average age of 39 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. College workers are defined as those with more than 12 years of schooling. In our sample, 13% of respondents finished less than 12 years of schooling, 35% had exactly 12 years of completed schooling, 21% 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 wage residuals $w_{i,t}$ from equation (1) after controlling for differences in educational attainment, race, and experience. Specifically, we estimate $f_t(\boldsymbol{x}_{i,t})$ by year and college vs. non-college status from separate linear regressions of log hourly wages on indicators for each year of potential experience, race, and our educational attainment categories, along with interactions between race and education indicators and a third-order polynomial in experience.

Figure 1 shows the total variance, between-group variance, and within-group variance (variance of residuals) of log wages over time. The variance of log wages increases sharply in the early 1980s and after the late 1990s. The evolution of the within-group variance closely mirrors this pattern. The within-group variance explains a larger share of the total variance than the between-group variance, and it also explains an increasing share of the total variance since the early 1990s.

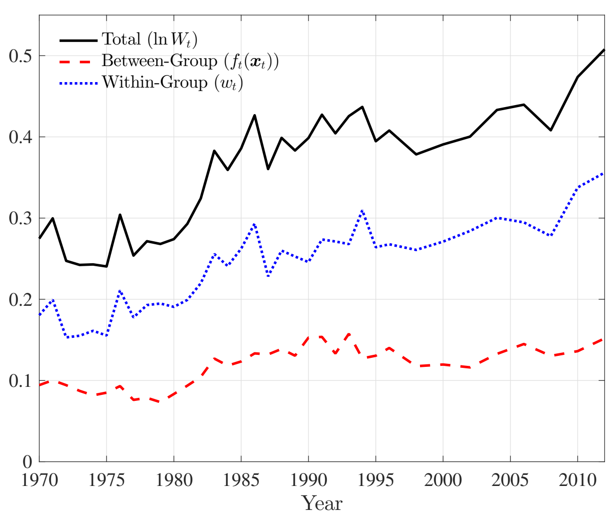
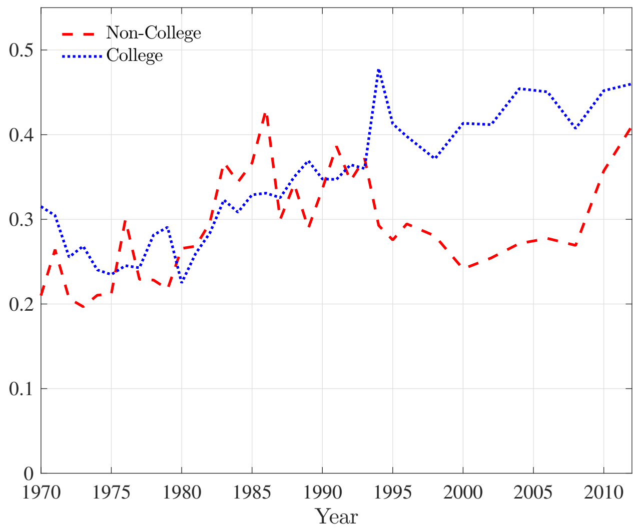


Figure 1: Between- and Within-Group Variances of Log Wages

The total and within-group variance of log wages evolved quite differently for non-college and college men, as shown in Figure 2. The rise in inequality in the early 1980s is stronger among non-college workers, while the increasing inequality among college workers is the main driver of the surge in inequality in the 1990s and early 2000s. Between the mid-1980s and early-2000s, wage inequality fell among non-college workers and increased among college workers, consistent with the literature on 'polarization' in the U.S. labor market (Autor, Levy, and Murnane, 2003; Autor, Katz, and Kearney, 2008; Acemoglu and Autor, 2011; Autor and Dorn, 2013).



(a) Total Variance

(b) Within-Group Variance

Figure 2: Variance of Log Wages by Educational Attainment

As discussed in Section 2, long autocovariances in wage residuals are central to identifying changes in the returns to unobserved skill. Figure 3 reports $\text{Cov}(w_b, w_t)$ for $6 \leq t - b \leq 20$ with each line reporting autocovariances for a different 'base' year b and 15 subsequent years.[14] For example, the leftmost line beginning in 1976 reflects autocovariances for $b = 1970$ and values of t ranging from 1976-1990. If heterogeneity in unobserved skill growth is negligible and $t - b$ is large enough such that transitory shocks are uncorrelated, then $\text{Cov}(w_b, w_t) = \mu_t \mu_b \,\text{Var}(\theta_b)$ and following each line over t is directly informative about the evolution of μ_t, while the shifts up or down across lines at a given date t are informative about differences in $\mu_b \,\text{Var}(\theta_b)$. The sharply declining autocovariances over the late-1980s and 1990s (regardless of the base year) suggest that the returns to unobserved skill fell over that period. The time trends for autocovariances were much weaker during earlier and later years, consistent with more stable returns. Finally, the upward shifting lines beginning in the 1980s, coupled with declining or constant skill returns, indicate an increasing variance of unobserved skills.

If heterogeneity in unobserved skill growth is important, then the residual covariances are more

[14]Figure 21 in Appendix E shows that sample attrition due to non-response or aging/retirement does not affect the autocovariance patterns documented in Figure 3.

difficult to interpret, since Cov(θ_b, θ_t) does not generally equal Var(θ_b) (see equation (10)). In this case, it is useful to focus on more experienced workers for whom differences in unobserved skill growth should be negligible. Figure 4 reveals very similar autocovariance patterns to Figure 3 when restricting the sample to men with 16-30 years of experience as of baseline b years.[15]

Altogether, these patterns form the key moments used in our empirical analysis below.

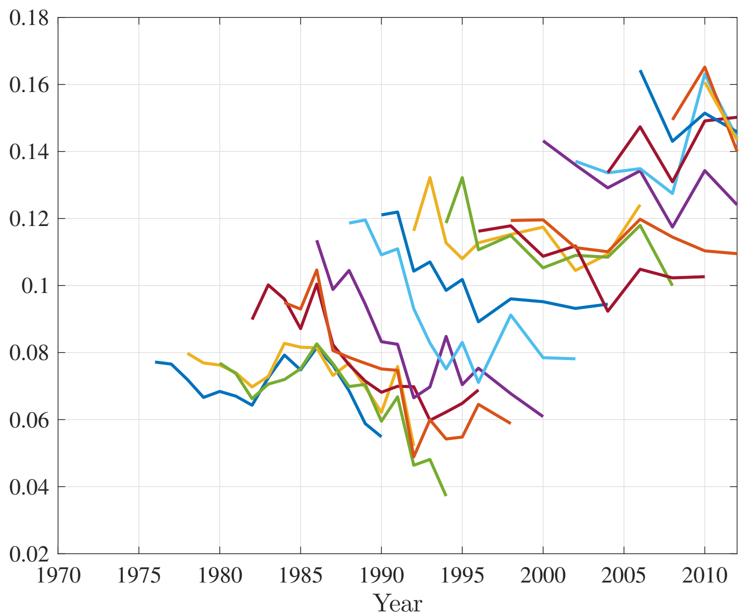


Figure 3: Autocovariances for Log Wages

4 Instrumental Variable Estimation

In this section, we estimate growth rates in the returns to unobserved skill based on the IV strategy described in Section 2. Due to the modest sample sizes of the PSID, we use the full sample of men, implicitly assuming no heterogeneity in unobserved skill growth. As just discussed, however, the autocovariance moments used in this analysis are quite similar for the subsample of more experienced workers. We more formally incorporate unobserved heterogeneity in skill growth among younger workers in Section 5.3

Given our data is only available every other year later in the sample period, we modify the strategy of Section 2 slightly to estimate two-year growth rates, $(\mu_t - \mu_{t-2})/\mu_{t-2}$ using two-stage least squares (2SLS)

[15] Figure 22 in Appendix E shows qualitatively similar autocovariances for men with 1-15 years of experience (i.e. lines generally declining in t over the late 1980s and 1990s, while shifting upwards across base years b); however, the patterns are much more muted. The weaker declines in Cov(w_b, w_t) with changes in t (among less experienced workers) are consistent with a positive correlation between baseline skills θ_b and individual-specific skill growth rates δ for young workers (see equation (10)). The lower levels and more modest upward shifts in the lines across base years b is consistent with less dispersion in skills among less experienced workers.

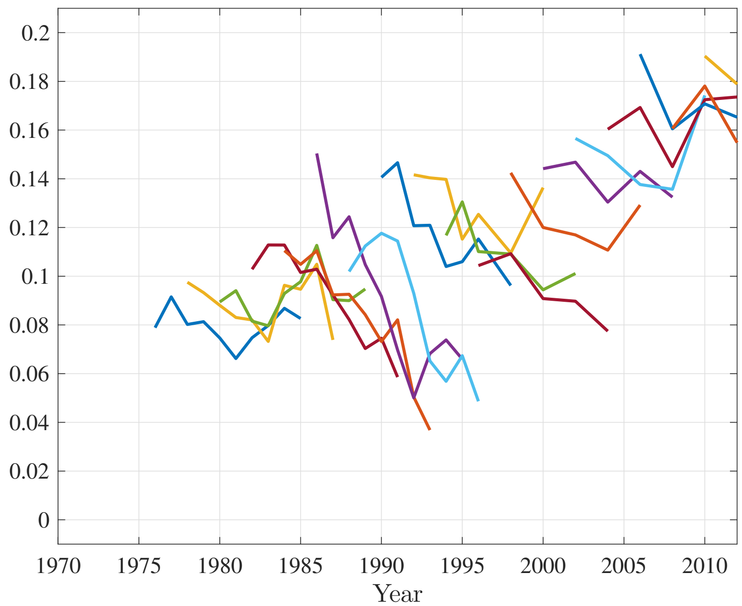


Figure 4: Autocovariances for Log Wages (16-30 Years of Experience)

with multiple lags of log wage residuals as instruments. Substituting in for $\theta_{i,t} = \frac{w_{i,t-2}-\varepsilon_{i,t-2}}{\mu_{t-2}} + \nu_{i,t-1} + \nu_{i,t}$ in equation (2), subtracting $w_{i,t-2}$ from both sides, and re-arranging yields

$$w_{i,t} - w_{i,t-2} = \left(\frac{\mu_t - \mu_{t-2}}{\mu_{t-2}}\right) w_{i,t-2} + \left[\mu_t(\nu_{i,t-1} + \nu_{i,t}) + \varepsilon_{i,t} - \frac{\mu_t}{\mu_{t-2}}\varepsilon_{i,t-2}\right], \tag{11}$$

where the final term in brackets is uncorrelated with sufficiently lagged residuals. So, if $\varepsilon_{i,t}$ is characterized by an MA(q) process (as will be assumed below), we can obtain consistent estimates of $(\mu_t - \mu_{t-2})/\mu_{t-2}$ by estimating equation (11) via 2SLS using lags $w_{i,t-q-3}, w_{i,t-q-4}, \ldots$ as instrumental variables.

Table 1 reports estimates of skill return growth rates using equation (11) for years t covering 1979-1995, assuming that skill return growth rates are constant within two- or three-year periods (i.e. 1979-1980, 1981-1983,...,1993-1995). Allowing for an MA(5) process for $\varepsilon_{i,t}$, we use $(w_{i,t-8}, w_{i,t-9})$ as instruments. Table 2 reports estimates for the later years of the PSID (t covering 1996-2012) when observations become biennial.[16]

Panel A of Tables 1 and 2 reports estimates for the full sample of men in the PSID, while panel B reports estimates for the sample of men with 21-40 years of experience (in year t) when systematic heterogeneity in unobserved skill growth should be negligible. As might be expected from Figures 3 and 4, the estimates are quite similar in both panels, and all are negative. Many estimates in the late 1980s and 1990s are statistically significant. Panels B and C report separate estimates for non-college and college

[16] Estimates in Table 2 assume two-year return growth rates are constant within each of the periods 1996-2000, 2002-2006, and 2008-2012, and use $(w_{i,t-8}, w_{i,t-9})$ as instruments for 1996-2000 and $(w_{i,t-8}, w_{i,t-10})$ thereafter.

men (of all experience levels). Nearly all of these estimates are negative as well, with several statistically significant. Figure 5 combines these estimates to trace out the implied paths for μ_t from 1979-2012. Altogether, these results suggest that the returns to unobserved skill have declined by roughly 50% since the mid-1980s, contrasting sharply with the sizeable increase in residual inequality reported in Figure 1.

Table 1: 2SLS estimates of $(\mu_t - \mu_{t-2})/\mu_{t-2}$ for three-year periods, 1978-1995

	1979-1980	1981-1983	1984-1986	1987-1989	1990-1992	1993-1995
			A. All men			
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.036	-0.044	-0.046	-0.081*	-0.082*	-0.067
	(0.045)	(0.038)	(0.038)	(0.034)	(0.035)	(0.035)
Observations	1,349	2,077	2,188	2,245	2,189	2,095
1st stage F-Statistic	163.09	191.61	114.85	209.42	227.13	286.96
			B. All men with 21-40 years of experience (at year t)			
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.052	-0.088*	-0.031	-0.100*	-0.036	-0.104*
	(0.050)	(0.043)	(0.050)	(0.046)	(0.044)	(0.045)
Observations	928	1,323	1,244	1,211	1,244	1,300
First Stage F-Statistic	117.23	132.19	66.26	130.53	132.83	201.62
			C. Non-college men (all experience levels)			
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.075	0.039	-0.035	-0.127*	-0.062	-0.057
	(0.061)	(0.056)	(0.060)	(0.050)	(0.058)	(0.054)
Observations	740	1,080	997	965	897	851
1st stage F-Statistic	81.85	85.23	39.48	98.34	92.27	91.33
			D. College men (all experience levels)			
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.034	-0.123*	-0.030	-0.028	-0.097*	-0.074
	(0.061)	(0.048)	(0.049)	(0.047)	(0.047)	(0.046)
Observations	508	884	1,046	1,109	1,107	1,242
1st stage F-Statistic	100.95	115.03	123.38	97.29	122.42	208.04

Notes: Estimates from 2SLS regression of $w_{i,t} - w_{i,t-2}$ on $w_{i,t-2}$ using instruments $(w_{i,t-8}, w_{i,t-9})$.

* denotes significance at 0.05 level.

Table 2: 2SLS estimates of $(\mu_t - \mu_{t-2})/\mu_{t-2}$ for four-year periods, 1996-2012

	1996-2000	2002-2006	2008-2012
	A. All men		
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.075*	-0.039	-0.050
	(0.025)	(0.028)	(0.027)
Observations	2,122	2,129	1,968
1st stage F-Statistic	369.09	344.25	341.36
	B. All men with 21-40 years of experience (at year t)		
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.084*	-0.040	-0.058
	(0.030)	(0.032)	(0.031)
Observations	1,427	1,591	1,493
First Stage F-Statistic	295.75	281.91	267.83
	C. Non-college men (all experience levels)		
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.087*	-0.043	0.011
	(0.043)	(0.047)	(0.075)
Observations	862	826	615
1st stage F-Statistic	121.44	142.56	104.92
	D. College men (all experience levels)		
$(\mu_t - \mu_{t-2})/\mu_{t-2}$	-0.070*	-0.041	-0.065*
	(0.031)	(0.034)	(0.029)
Observations	1,252	1,293	1,141
1st stage F-Statistic	260.47	218.64	229.40

Notes: Estimates from 2SLS regression of $w_{i,t} - w_{i,t-2}$ on $w_{i,t-2}$ using instruments (w_{t-8}, w_{t-9}) for 1996-2000 and (w_{t-8}, w_{t-10}) for 2002-2006 and 2008-2012. * denotes significance at 0.05 level.

In Appendix B, we show that analogous GMM estimates to the 2SLS estimates in panel A of Tables 1 and 2 are very similar.[17] More importantly, we calculate Hansen J-statistics to test the validity of our

[17]The GMM estimates exploit the same moments but use the optimal weighting matrix (allowing for heteroskedasticity and serial correlation within individuals).

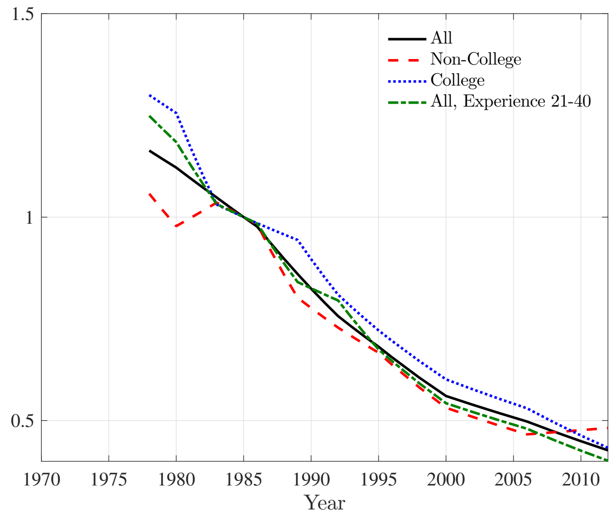


Figure 5: μ_t Implied by IV Estimates ($\mu_{1985} = 1$)

lagged instruments, since we are overidentified when using two instruments to identify a single parameter. In all years, these J-statistics are less than one, and we cannot reject our instruments at conventional levels. We further show that future residuals are invalid instruments (during most of our time periods), highlighting the importance of accounting for variation in skill growth. Finally, comparing estimates using only past vs. only future residuals as instruments, we show that the variance of two-year skill growth relative to prior skill levels, $\frac{\text{Var}(\nu_{t-1}+\nu_t)}{\text{Var}(\theta_{t-2})}$, ranges from 0.16 to 0.29 over our sample period.

It is worth emphasizing that these estimates require no assumptions about the variance of individual skill innovations $\nu_{i,t}$ (or non-skill shocks, $\varepsilon_{i,t}$) over time or across experience groups. The only assumptions are: (i) skill shocks $\nu_{i,t}$ are uncorrelated with past skills; and (ii) non-skill shocks $\varepsilon_{i,t}$ are uncorrelated with non-skill shocks more than five years removed, initial skill levels, and all skill shocks. (Our overidentification tests suggest that these assumptions cannot be rejected.) Equation (7) shows that with modest assumptions on the variance of skills and skill innovations across cohorts, additional moment restrictions can also be used to help identify μ_t in all but the final few years. We have, thus far, used a very limited set of lagged residuals as instruments to keep the specifications similar across years and to allow estimation of skill return growth rates back to 1979. One could certainly add more lags as instruments for most years. Rather than report several sets of 2SLS estimates, the next section employs a minimum distance estimator to estimate the returns to unobserved skills, as well as the variances of unobserved skills and shocks over time, using the full set of available moments in the data. ‘Smoothness’ restrictions on the time and experience patterns for skill and non-skill shock variances are also imposed. The combination of more moments and modest ‘smoothness’ restrictions considerably improves precision, yet it produces very similar estimates for the time sequence of μ_t.

5 Minimum Distance Estimation

We now estimate the full model by choosing parameters to minimize the distance between the sample covariances and the theoretical covariances implied by the model. This requires a more complete specification of the model; however, since we are focused on estimating the evolution of unobserved skill returns and the variances of skills over time, we need not model or use higher order moments.

5.1 Specification and Identification

We assume that individuals enter the labor market with initial skill ψ_i and no other prior shocks, which implies that unobserved skill in year t for individual i who started working in year c_i can be written as

$$\theta_{i,t} = \psi_i + \sum_{j=0}^{t-c_i-1} \nu_{i,t-j}. \tag{12}$$

As in much of the literature, we further assume that the transitory component $\varepsilon_{i,t}$ is a moving average process with order q:

$$\varepsilon_{i,t} = \sum_{j=0}^{\min\{q,t-c_i-1\}} \beta_j \xi_{i,t-j}, \tag{13}$$

where $\beta_0 = 1$.

Given this $MA(q)$ specification for $\varepsilon_{i,t}$, the identifying assumption in Proposition 1 ($\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = 0$ for $|t' - t| \geq k$) holds for $k = q + 1$. This demonstrates identification for all skill returns μ_t (up to one normalization) as well as cohort- and time-specific variances of skills, $\mathrm{Var}(\theta_t|c) = \mathrm{Cov}(w_t, w_{t'}|c)/\mu_t\mu_{t'}$, for cohorts observed in years t and any $t' \geq t+q+1$. Next, $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = \mathrm{Cov}(w_t, w_{t'}|c) - \mu_t\mu_{t'} \mathrm{Var}(\theta_t|c)$ for $t' \geq t$ is identified for (c, t) such that $\mathrm{Var}(\theta_t|c)$ is identified. Then, β_j is identified from some cohort c and period $t = c + 1$ as follows:

$$\frac{\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+j}|c)}{\mathrm{Var}(\varepsilon_t|c)} = \beta_j \frac{\mathrm{Var}(\xi_t|c)}{\mathrm{Var}(\xi_t|c)} = \beta_j.$$

Given β_j's, we can recover $\mathrm{Var}(\xi_t|c)$ for each period by following cohorts over time.[18]

[18] Consider a cohort c. For the initial period $t = c + 1$, $\mathrm{Var}(\xi_t|c) = \mathrm{Var}(\varepsilon_t|c)$. For $t > c + 1$,

$$\mathrm{Var}(\xi_t|c) = \mathrm{Var}(\varepsilon_t|c) - \sum_{j=1}^{\min\{q,t-c-1\}} \beta_j^2 \mathrm{Var}(\xi_{t-j}|c).$$

In our empirical analysis, we assume that the initial skill variance for each cohort $\text{Var}(\psi|c)$ is a cubic polynomial in c. We also assume that the variances of shocks for each period and cohort can be written as products of time trends and experience trends:

$$\text{Var}(\nu_t|c) = \pi(t)\phi(e) \qquad \text{and} \qquad \text{Var}(\xi_t|c) = \omega(t)\kappa(e),$$

where $e = t - c$ is potential work experience and we normalize $\phi(20) = \kappa(20) = 1$. We assume that the experience trends $\phi(e)$ and $\kappa(e)$ are quadratic. The time trend for the skill shock variance $\pi(t)$ is assumed to be cubic, but the time pattern for the variances of transitory shocks $\omega(t)$ is unrestricted. Finally, we assume that the variances of all shocks prior to 1970 are the same as in 1970.

We estimate μ_t for all t (normalizing $\mu_{1985} = 1$), $\{\beta_j\}_{j=1}^{q}$, cohort trends in $\text{Var}(\psi|c)$, time trends in shocks, $\pi(t)$ and $\omega(t)$, and experience trends in shocks, $\phi(e)$ and $\kappa(e)$. Since the relative wages between those who did not attend college and those that did have diverged significantly during the sample period (Katz and Murphy, 1992), we also estimate all parameters separately for the two groups, which we call 'sectors'. Let $s_{i,t}$ be an indicator variable for college attendance.

For a given parameter vector $\mathbf{\Lambda}$, we can compute theoretical counterpart for $\text{Cov}(w_t, w_{t'}|s, c)$ implied by the model (2), (12), and (13) and compare them with the sample covariances. Since some cohort (or, equivalently, experience $e = t - c$) cells have few observations when calculating residual covariances, we partition the experience set $E = \{1, \ldots, 40\}$ into 10-year experience groups E_1, E_2, E_3, and E_4, each corresponding to experiences 1-10, 11-20, 21-30, and 31-40, respectively, and aggregate within experience groups.

The minimum distance estimator $\hat{\mathbf{\Lambda}}$ solves

$$\min_{\mathbf{\Lambda}} \sum_{(s,j,t,t')\in\Gamma} \left\{ \widehat{\text{Cov}}(w_t, w_{t'}|s, E_j) - \text{Cov}(w_t, w_{t'}|s, E_j, \mathbf{\Lambda}) \right\}^2,$$

where $\Gamma = \{s, j, t, t' | 1970 \leq t' \leq t \leq 2012, s \in \{0, 1\}, j \in \{1, 2, 3, 4\}\}$, $\widehat{\text{Cov}}(w_t, w_{t'}|s, E_j)$ is the sample covariance conditional on sector s and experience group E_j in year t, and $\text{Cov}(w_t, w_{t'}|s, E_j, \mathbf{\Lambda})$ is the corresponding theoretical covariance given parameter $\mathbf{\Lambda}$.

5.2 Estimation Results

We discuss results for log wage residuals with MA(5) transitory shocks in the text; however, conclusions are quite similar for log annual earnings and other specifications for the transitory component.[19]

[19] Available upon request.

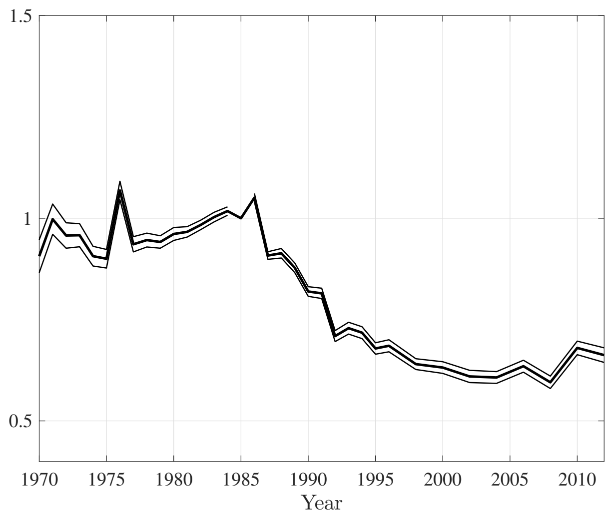


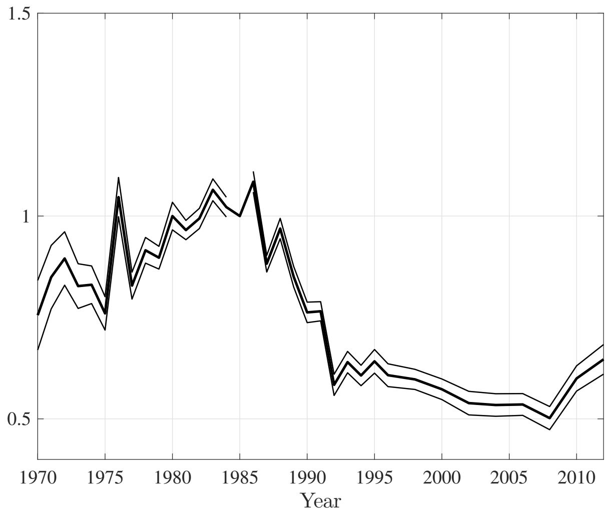
Figure 6: Estimated μ_t (thick lines) with 95% Confidence Interval (thin lines)

Estimated Returns Figure 6 reports the estimated returns over time along with their 95% confidence intervals. The returns are quite similar to those reported in Figure 5 but are much more precisely estimated now. Figure 7 shows separate estimates by education sector. Overall, the returns to unobserved skills fell substantially after 1985 in both sectors despite the increasing residual inequality during the period. Although the returns for non-college and college workers display qualitatively similar time patterns, the return for non-college workers declines about 20 percentage points more between 1985 and 2008 than that of college workers. Indeed, we can reject the hypothesis that the returns to unobserved skills are identical across education groups at the 5% significance level.[20] The relatively larger decline in the returns to unobserved skills for non-college workers after 1985 is in line with the falling relative wages of non-college workers during that period (e.g., Autor, Katz, and Kearney, 2008), suggesting that the changing returns to unobserved skills may also be an important determinant of the evolution of between-group inequality.

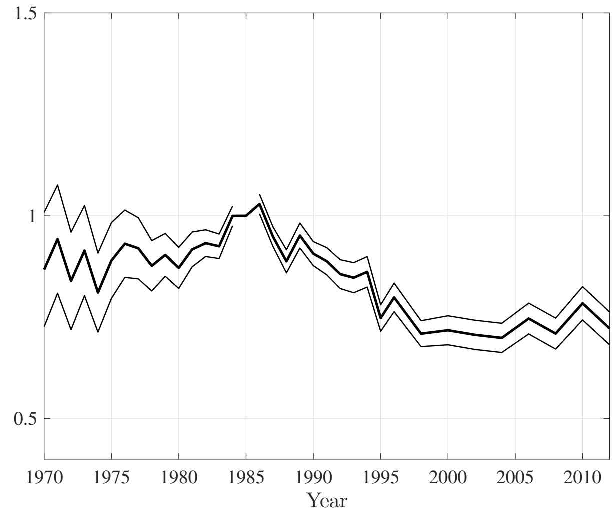
One potential concern is that skill growth shocks may be serially correlated, especially for younger workers who are likely to be investing in their skills through job training. This is less of a concern for older workers, whose wage profiles are relatively flat due to weaker incentives for investment (e.g., Heckman, Lochner, and Taber, 1998; Huggett, Ventura, and Yaron, 2011). Figure 8 shows that estimated returns to unobserved skill are quite similar for workers with low (1-20 years) and high (21-40 years) levels of experience, alleviating concerns that our estimated decline in μ_t over time is due to serially correlated skill shocks.[21] In Subsection 5.3, we further show that the estimated μ_t series is quite similar to that of

[20] The Wald test is used to test the null hypothesis that μ_t is identical across sectors for all $t \neq 1985$.

[21] These estimated return sequences are based on the full sample assuming parameters are the same across education groups (as in Figure 6) but allowing the μ_t returns to depend on worker experience in period t.



(a) Non-College



(b) College

Figure 7: μ_t Estimated Separately by Sector

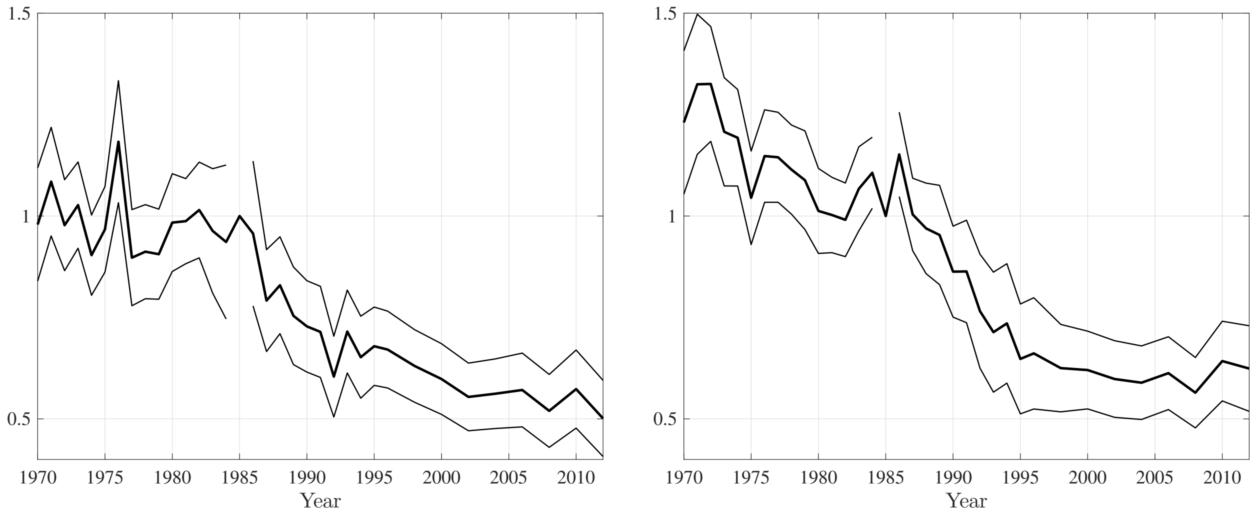
Figure 6 when we incorporate individual heterogeneity in systematic lifecycle wage growth profiles.

Our finding that the returns to unobserved skills have fallen since the mid-1980s is consistent with the work of Castex and Dechter (2014), who showed that the returns to cognitive ability as measured by AFQT scores fell in the 2000s relative to the 1980s by comparing NLSY79 and NLSY97 cohorts. The decreasing returns to unobserved skills differs, however, from the conclusions reached in the CPS-based literature, which implicitly ignores wage shocks and equates changes in the total variance of log wages residuals with changes in the returns to unobserved skills. This literature generally concludes that the returns to unobserved skills increased steadily after the early 1970s (Juhn, Murphy, and Pierce, 1993; Katz and Autor, 1999; Acemoglu, 2002).[22]

The PSID-based literature on earnings dynamics focuses primarily on the relative importance of permanent and transitory shocks, typically ignoring variation in the returns to unobserved skills.[23] Haider (2001) and Moffitt and Gottschalk (2012) are notable exceptions. Haider (2001) estimates that the returns to unobserved skills were stable in the 1970s and then increased throughout the 1980s, while Moffitt and Gottschalk (2012) find that the returns increased until the mid-1980s, stabilized, and then increased again

[22]One exception is Lemieux (2006), who finds that, after controlling for composition effects, the return to unobserved skills declined in the 1970s and 1990s. Autor, Katz, and Kearney (2008) show that the decline in the returns to unobserved skills between 1989 and 2005 is driven by decreasing returns among workers in the lower tail of the wage distribution. Lochner and Shin (2014) explore identification and estimation of the returns to unobserved skills allowing for differential changes in the returns across skill levels.

[23]See, among others, Abowd and Card (1989); Blundell and Preston (1998); Meghir and Pistaferri (2004); Blundell, Pistaferri, and Preston (2008); Heathcote, Perri, and Violante (2010). See MaCurdy (2007) for a review.



(a) Experience 1-20

(b) Experience 21-40

Figure 8: μ_t Estimated Separately by Experience

in the mid-1990s.[24]

Of course, the patterns estimated in the earnings dynamics literature may differ from ours due to labor supply responses and the distinction between annual earnings and hourly wages. This is not the entire story, however. Figure 9 shows that the evolution of estimated returns to skill in Haider (2001) and Moffitt and Gottschalk (2012) differ from ours largely because they restrict the variance of permanent skill shocks to remain constant over time. These estimates are based on an analysis of log hourly wage residuals that pools non-college and college men, estimating a single series for μ_t as in Figure 6 (referred to as 'Baseline' in the figure). The blue line with circles reports estimates assuming that transitory non-skill shocks follow an ARMA(1,1) (rather than MA(5)) process as in their analyses. The estimated time patterns for μ_t are quite similar to our baseline estimates. More generally, different assumptions about the persistence of transitory non-skill shocks all lead to similar conclusions about the returns to skill. The red dashed line reports estimates additionally assuming that the variance of skill shocks is time-invariant, while the green line with + signs further assumes that the initial skill distribution is identical across cohorts. The last model is nearly identical to those estimated in Haider (2001) and Moffitt and Gottschalk (2012)[25] and the estimated series for μ_t is similar to theirs. Moving from the baseline to the final model, the estimated

[24]Other studies exploit different panel data sets on earnings to estimate similar models to Haider (2001) and Moffitt and Gottschalk (2012). DeBacker et al. (2013) use U.S. tax return data from 1987 to 2009, while Baker and Solon (2003) exploit Canadian tax return data from 1976 to 1992. Both studies reach similar conclusions that the returns to skill increased over time.

[25]These studies also allow for heterogeneity in the growth rate of unobserved skills that may be correlated with initial unobserved skills. We incorporate this heterogeneity in Subsection 5.3 and show that the estimated returns to skill are quite similar to the 'Baseline' series.

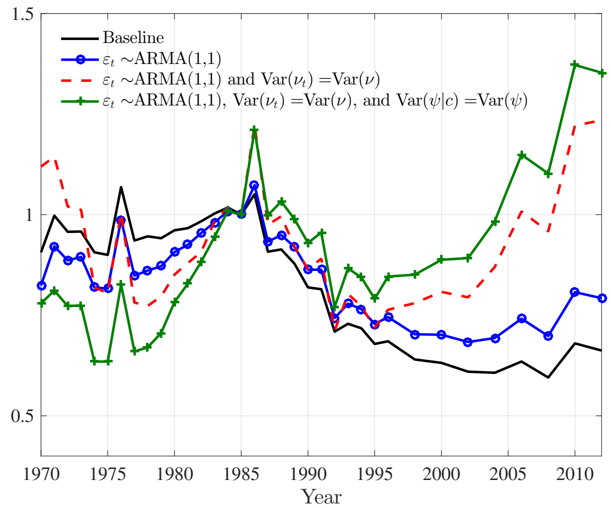


Figure 9: Estimated μ_t under Different Restrictions

returns rotate counter-clockwise, generating strong positive trends both before the mid-1980s and after the mid-1990s. The difference between the $ARMA(1,1)$ model with time-varying vs. time-invariant skill growth shock distributions is dramatic after 1990, highlighting the importance of accounting for the rise in the variance of skill growth innovations. When the variance of skill growth shocks is not allowed to increase over time, the model is 'forced' to explain the increasing residual variance via an increase in the returns to skill.

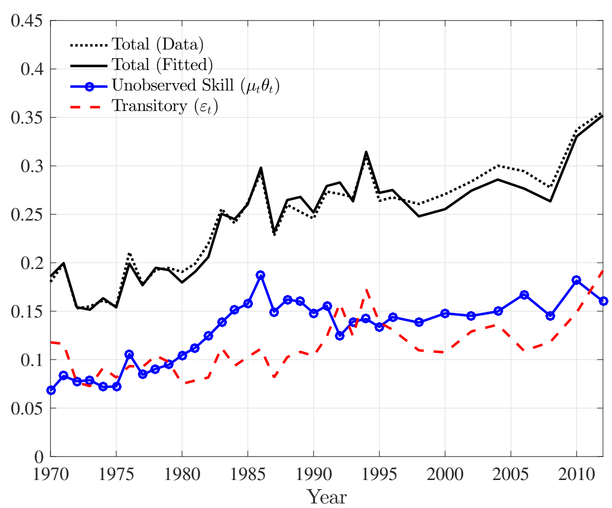
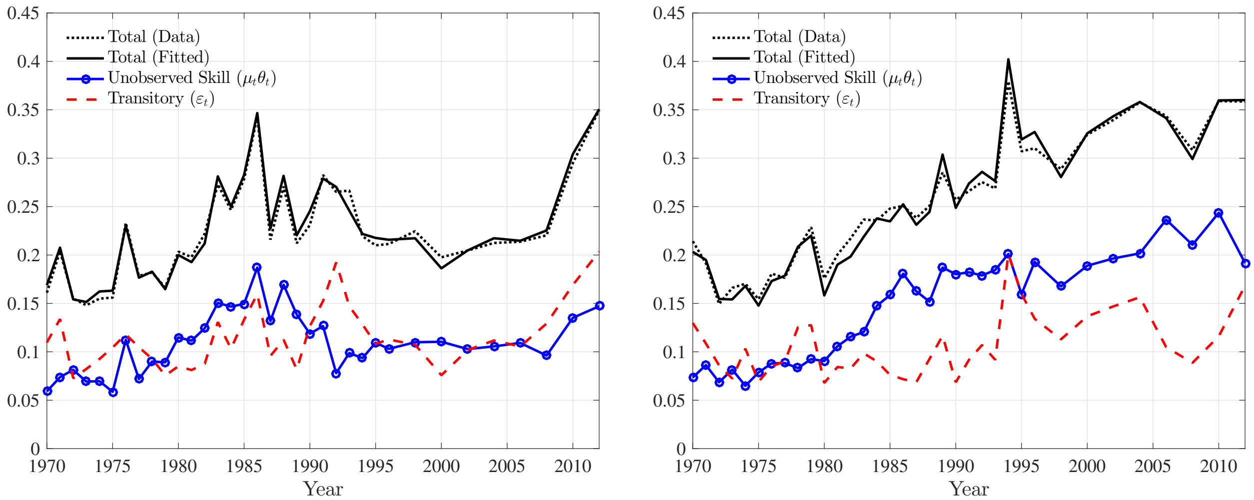


Figure 10: Log Wage Residual Variance Decomposition (Full Sample)



(a) Non-College (b) College

Figure 11: Log Wage Residual Variance Decomposition Separately by Education

Variance of Unobserved Skills and the Rising Residual Variance The fact that estimated returns have evolved quite differently from residual inequality implies that the role of unobserved skills has also changed. Figure 10 decomposes the residual variance into two components: the unobserved skill component ($\mu_t\theta_t$) and transitory component (ε_t).[26] Initially quite low, the variance of the unobserved skill component rises over the 1970s and early 1980s before stabilizing after 1985. The variance of the transitory component rises in the late 1980s and early 1990s. The unobserved skill component explains about 65% of the total residual variance at its peak in the late 1980s with its share decreasing thereafter.

Figure 11 decomposes the residual variance when estimating the model separately for non-college and college men. The figure reveals that the general time patterns for overall residual inequality (within education groups) are largely driven by changes in the variance of the unobserved skill component, $\mu_t\theta_t$. The variance of this component rose for both college and non-college workers over the 1970s and early 1980s; however, it reversed course for non-college workers over the late 1980s and early 1990s. Among college workers, the variance of the unobserved skill component continued to rise throughout the sample period, although there was a slowdown in the growth rate beginning in the mid-1980s. The course reversal for the Var($\mu_t\theta_t$) among non-college workers and the slowdown for college workers beginning in the mid-1980s is driven entirely by the sudden and lasting decline in skill returns μ_t documented in Figure 7. As Figure 12 shows, variation in skill levels θ_t rose continually from the mid-1980s through the early 2000s.

[26]As shown in Section 2, the variances of unobserved skills and transitory components are not nonparametrically identified for the last few years of our panel; however, our 'smoothness' assumptions for time, cohort, and experience trends provide enough restrictions to enable identification.

Figure 12 also decomposes the rising variance in unobserved skills θ_t into the variance of initial skill levels and the accumulation of all permanent skill shocks.[27] The increasing variance of unobserved skills is driven entirely by an increase in the variability of skill growth shocks over time. The estimated time trends for the variance of skill growth shocks, $\pi(t)$, is reported in Figure 13.[28]

Summarizing, we find that U.S. trends in residual wage inequality closely resemble trends in the variance of the unobserved skill component of wages. Variation in the total value of unobserved skill, $\mu_t\theta_t$, has risen considerably since 1970. For both college and non-college men, the increase was strongest before the mid-1980s, after which it increased more slowly for college men and declined for almost a decade among non-college men. The increasing variability in the total value of skill contrasts sharply with the strong decline in returns to unobserved skill, μ_t, from the mid-1980s to late-1990s, especially among non-college men. This decline in returns is largely responsible for the slowdown (college men) and roughly decade-long fall (non-college men) in residual inequality; yet, it was offset by a stronger increase in the variance of unobserved skill θ_t beginning in the early 1980s. Finally, the increasing variance of skills is driven by increasing variation in permanent skill shocks, ν_t, over time rather than an increase in initial skill levels for more recent cohorts. The transitory, non-skill component of wages, ε_t, showed little systematic growth over this period.[29]

5.3 Accounting for Heterogeneity in Systematic Lifecycle Skill Growth

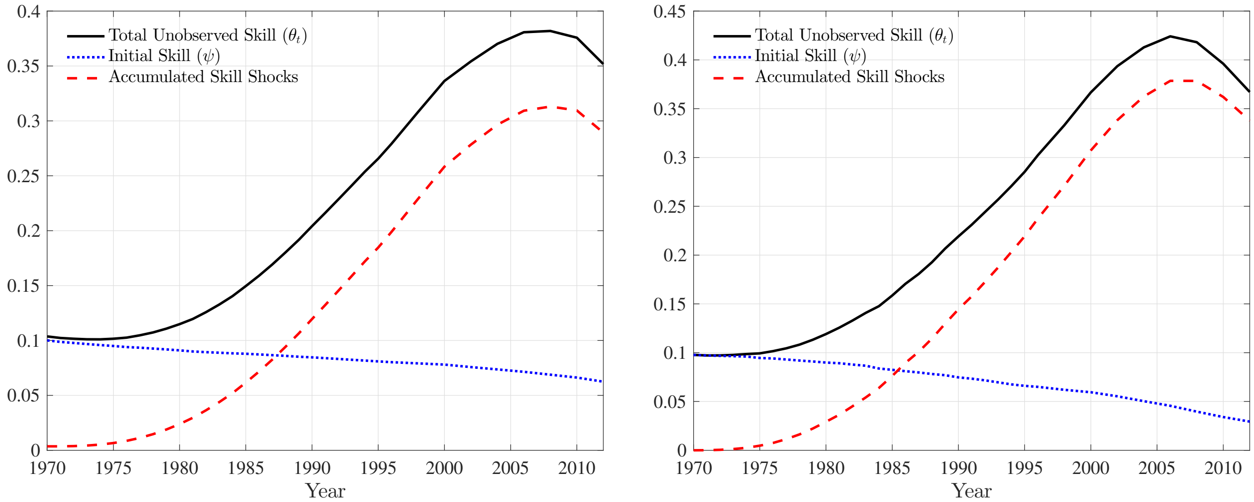
We now incorporate potential heterogeneity in systematic lifecycle skill growth as discussed in Section 2.2.2. In particular, we consider a flexible process governing heterogeneity over the lifecycle and across time: $\nu_{i,t} = \tau_t(c_i)\delta_i + \tilde{\nu}_{i,t}$.

For practical purposes, we make a few additional assumptions on the structure of unobserved skill growth in estimating this framework using the PSID. We begin by assuming that $\tau_t(c)$ is separable in experience and time, so $\tau_t(c) = \chi(t)\eta(e)$. The parameter $\chi(t)$ captures any time-varying differences in systematic skill growth (normalizing $\chi(1985) = 1$), while $\eta(e)$ allows skill growth heterogeneity to vary with experience. We assume that $\chi(t)$ is a cubic polynomial in time, while we assume that $\eta(e)$ declines linearly in experience (a natural assumption given the decline in skill growth over the lifecycle). Specifically, we assume $\eta(e) = \max\{1 - e/\bar{e}, 0\}$ with $\bar{e} = 30$. Finally, we assume that $\mathrm{Var}(\delta|c)$ and $\mathrm{Cov}(\delta, \psi|c)$ are the same across cohorts, while we continue to allow for cohort trends in $\mathrm{Var}(\psi|c)$ as above.

[27]These aggregate across all cohort/experience groups within each period. Initial skills are given by ψ for each cohort, while the accumulation of permanent skills are simply $\theta_t - \psi$ for each cohort.

[28]See Appendix E for estimated cohort trends in the variance of initial skills (Figure 23) and for the experience profiles for the variance of permanent skill shocks, $\phi(e)$ (Figure 24).

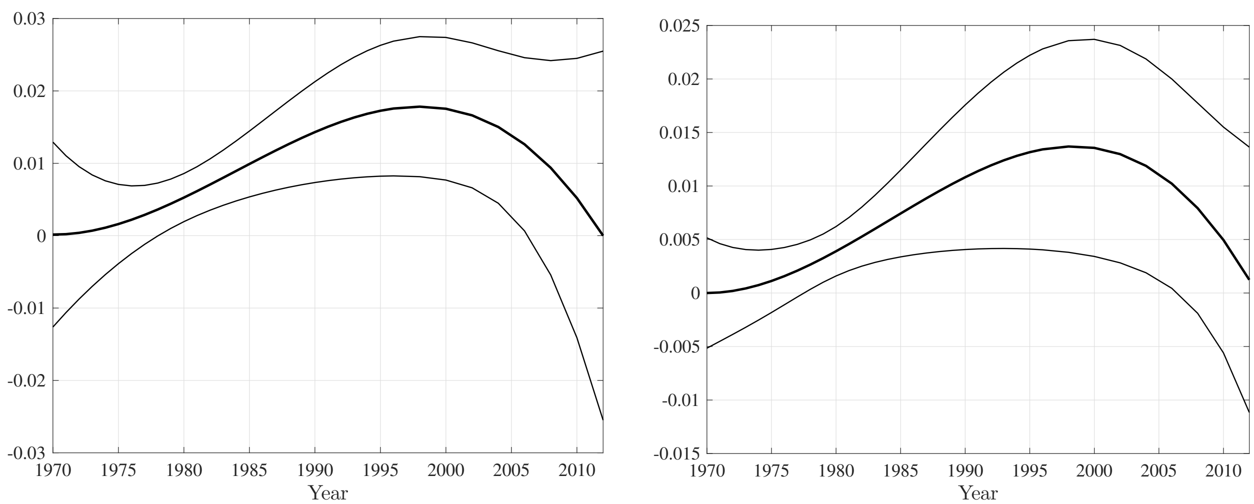
[29]See Figures 25 and 26 in Appendix E for estimated time and experience patterns for transitory shocks.



(a) Non-College

(b) College

Figure 12: Skill Variance Decomposition by Education



(a) Non-College

(b) College

Figure 13: Time Trends in the Variances of Skill Shocks, $\pi(t)$, by Education

We first examine the impacts of introducing systematic skill growth heterogeneity for our estimates of skill returns, keeping $\chi(t) = \chi$ fixed over time. This is similar to previous studies that allow for HIP; however, we continue to allow for time variation in the variance of skill shocks $\tilde{\nu}_t$.[30] Figure 14 shows that incorporating systematic skill growth heterogeneity has very little effect on our estimated μ_t skill return series (compare with Figure 6). Allowing $\chi(t)$ to vary over time (Figure 15), we estimate substantial growth in the variance of systematic skill growth over the 1980s and early 1990s, followed by a strong decline thereafter. These results suggest a slightly stronger decline in skill returns after 1985.

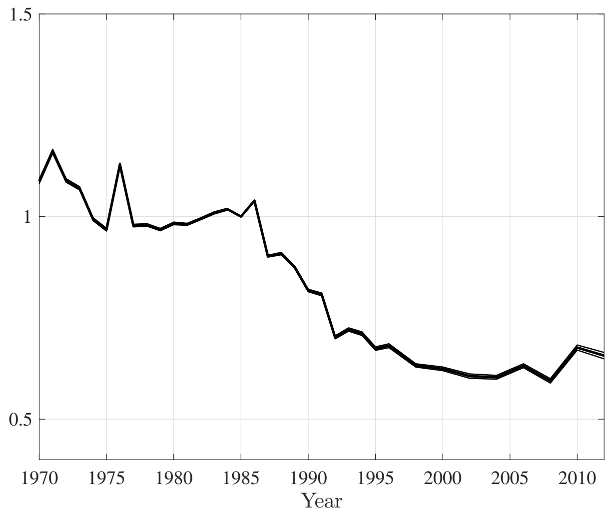


Figure 14: Estimated μ_t Accounting for Heterogeneous Skill Growth with $\chi(t) = \chi$

Finally, Figure 16 shows the dramatic increase in the variance of unobserved skills over time and decomposes this variance into the part due to heterogeneity in initial skills, ψ, which varies across cohorts, the part due to systematic skill growth (including terms related to $\mathrm{Var}(\delta)$ and $\mathrm{Cov}(\psi, \delta)$ in equation (9)), and the part due to accumulated idiosyncratic skill growth shocks, $\tilde{\nu}_t$. Consistent with our earlier results, variation in initial skills plays a minor role in the rising skill variance. Instead, we observe strong increases in both the variance of both systematic and idiosyncratic skill growth innovations.

6 Interpreting the Returns to Skills in a Demand and Supply Framework

Our analysis shows that the changing returns to unobserved skills are important in understanding the evolution of residual inequality. The changes in the returns to skills documented in the previous section

[30]For this analysis, we pool all men and assume parameters are the same across education groups (college and non-college).

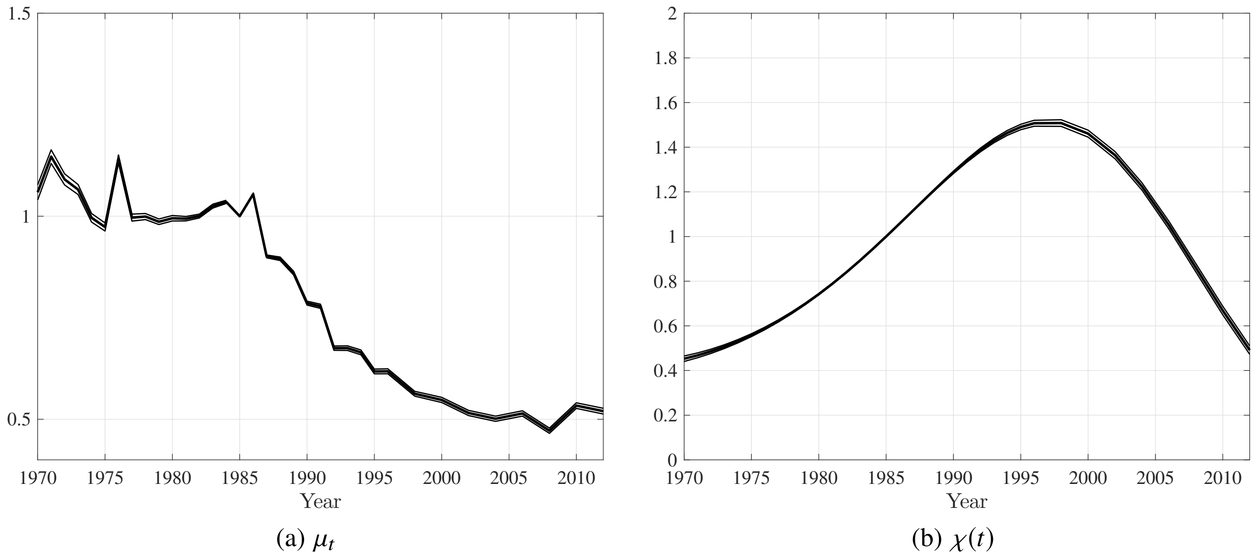


Figure 15: Estimated μ_t and $\chi(t)$ Accounting for Time-Varying Variance of Heterogeneous Skill Growth

could simply reflect shifts in the demand for skills. However, the concurrent changes in the distribution of skills suggests that the supply of skills may also play a role. In this section, we assess the contributions of skill demand and supply in the evolution of returns to skills using an equilibrium framework based on the assignment model of Sattinger (1979). Assignment models, which have been a standard theory of income inequality in labor economics,[31] are useful in studying within-group wage inequality because they generate a hedonic wage function that is non-linear in skill. The theoretical framework highlights economic forces driving the changes in the returns and also provides results that can be used to recover changes in skill demand from the estimated returns and distributions of unobserved skills of the previous section.

In the model, the return to skills is generated by differences in the productivity of skills but is amplified by differences in the productivity of jobs for which skills are employed. Heterogeneous productivity across jobs can be thought of as the different amount of resources under each worker's control, such as the quantity of capital or hierarchy, or different vintages of technology embodied in capital. As in the models of 'span-of-control' (Lucas, 1978) and 'superstar' (Rosen, 1981), production technology determines how much differences in skills are magnified to differences in earnings. In contrast to these models, however, jobs are in fixed supply, so the supply of skills also affect the return to skills through its effect on the equilibrium assignment of skills and jobs.

[31]See Sattinger (1993) for an early review. Recent theoretical and empirical studies of income inequality based on this framework include Terviö (2008); Gabaix and Landier (2008); Costinot and Vogel (2010); Lindenlaub (2017); Burstein, Morales, and Vogel (2015); Ales, Kurnaz, and Sleet (2015); Scheuer and Werning (2015).

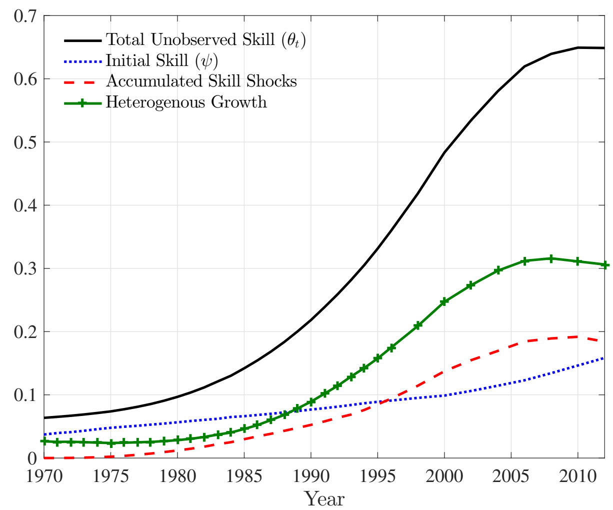


Figure 16: Skill Variance Decomposition Accounting for Time-Varying Variance of Heterogeneous Skill Growth

Since the aim of this analysis is to understand the evolution of skill returns in equilibrium, we do not consider the forces driving changes in the underlying distribution of skills.[32] We also abstract from transitory wage shocks unrelated to skills.[33]

6.1 Assignment Model of Labor Market

Endowment We consider an economy populated by a continuum of measure 1 of workers and jobs. In each period t, there exist workers endowed with heterogeneous skills Θ_t and jobs with heterogeneous productivity Z_t. Worker skills and job productivities are continuously distributed with full support on the real line according to distribution functions $F_t(\Theta_t)$ and $G_t(Z_t)$. The skill Θ_t is observed by all market participants, but is not perfectly observed by an econometrician. The component of the skill that is orthogonal to observed variables, $\theta_t = \Theta_t - \mathbb{E}[\Theta_t|\boldsymbol{x}_t]$, corresponds to the unobserved skill in the empirical model.

[32]There is a literature explaining the trends in between- and within-group wage inequality in the framework of endogenous human capital accumulation with heterogeneous learning ability and imperfect substitution of human capital across groups (e.g., Heckman, Lochner, and Taber, 1998; Guvenen and Kuruscu, 2010). This literature does not take into account idiosyncratic earnings shocks that are unrelated to human capital, equating all changes in residual wage inequality with changes in the inequality of human capital and time spent on training. Using a model that is similar to ours, Jovanovic (1998) theoretically studies the effects of technical changes on long-run inequality when the distributions of skills and jobs endogenously evolve over time.

[33]Violante (2002) develops a theory of how technical changes may affect transitory shocks. An acceleration of technological progress increases the productivity gaps across jobs, which is translated into more volatile earnings in a frictional labor market.

Technology Production takes place through one-to-one matching between workers and jobs. If a worker with skill Θ_t works at a job with productivity Z_t, $Y_t(\Theta_t, Z_t) \geq 0$ units of output is produced. We assume that $Y_t(\cdot, \cdot)$ is twice continuously differentiable and strictly increasing, and it also satisfies the following strict supermodularity condition:

$$\frac{\partial^2 Y_t(\Theta_t, Z_t)}{\partial \Theta_t \partial Z_t} > 0.$$

These assumptions imply that high skill workers are more productive (i.e., they have absolute advantages) than low skill workers at all jobs, but the productivity gap between high and low skill workers is larger at more productive jobs due to complementarity between workers and jobs. Therefore, the efficient assignment that maximizes aggregate output features positive assortative matching where more skilled workers work at more productive jobs (Becker, 1973).

We assume the following constant elasticity of substitution production function:

$$Y_t(\Theta_t, Z_t) = \left[\left(\frac{\lambda_t}{\lambda_t + \gamma_t} \right) \exp(\rho \Theta_t) + \left(\frac{\gamma_t}{\lambda_t + \gamma_t} \right) \exp(\rho Z_t) \right]^{\frac{\lambda_t + \gamma_t}{\rho}}, \tag{14}$$

where $\lambda_t > 0, \gamma_t > 0$, and $\rho < \lambda_t + \gamma_t$. The Cobb-Douglas production function $Y_t(\Theta_t, Z_t) = \exp(\lambda_t \Theta_t + \gamma_t Z_t)$ corresponds to the case where $\rho = 0$.

Profit Maximization All markets are perfectly competitive. Workers with skill Θ_t earn $W_t(\Theta_t)$ and output is sold at price that is normalized to 1. Producers maximize profits, solving

$$\Pi_t(Z_t) \equiv \max_{\Theta_t} \Big\{ Y_t(\Theta_t, Z_t) - W_t(\Theta_t) \Big\}. \tag{15}$$

Denote the solution by $\hat{\Theta}_t(Z_t)$. Because it is strictly increasing in Z_t due to the strict supermodularity of $Y_t(\cdot, \cdot)$, its inverse $\hat{Z}_t(\Theta_t)$, which we call the 'matching function', is well defined. The necessary first order condition for profit maximization is

$$\frac{\partial Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)}{\partial \Theta_t} = \frac{dW_t(\Theta_t)}{d\Theta_t} \tag{16}$$

and the second order condition is also satisfied at optimum due to the supermodularity.

Equation (16) determines the slope of the hedonic wage function. To pin down the level of wage, we assume that both producers and workers have an option not to engage in production, in which case they earn zero. Therefore, $0 \leq W_t(\Theta_t) \leq Y_t(\Theta_t, \hat{Z}_t(\Theta_t))$ should hold for all Θ_t. Since skill and job distributions are unbounded, the production function (14) implies that the least productive worker-job pair produces nothing, that is, $\lim_{\Theta_t \to -\infty} Y_t(\Theta_t, \hat{Z}_t(\Theta_t)) = 0$, which in turn implies $\lim_{\Theta_t \to -\infty} W_t(\Theta_t) = 0$.

Market Clearing The labor market clears if, for all Θ_t, the fraction of producers demanding skills Θ_t or less equals the fraction of workers supplying skills Θ_t or less:

$$F_t(\Theta_t) = G_t\big(\hat{Z}_t(\Theta_t)\big). \tag{17}$$

Note that, with one-to-one matching and exogenous distributions, the equilibrium assignment is entirely characterized by the market clearing condition, independent of technology. We assume that Θ_t and Z_t belong to the same location-scale family distribution,[34] in which case the market clearing condition (17) simplifies to

$$\hat{Z}_t(\Theta_t) = m_{Z_t} + \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}(\Theta_t - m_{\Theta_t}), \tag{18}$$

where m_X and σ_X are the location and scale parameters of a random variable X. The location parameters, which are measures of central tendency, determine the level of the matching function. All workers are matched with better jobs if m_{Z_t} is higher or m_{Θ_t} is lower. The scale parameters, which measure dispersion, determine the slope of the matching function. When $\sigma_{Z_t}/\sigma_{\Theta_t}$ is large, the productivity gap of jobs among workers with similar skill levels is large.

Equilibrium Wage A competitive equilibrium consists of the matching function $\hat{Z}_t(\Theta_t)$ and the wage function $W_t(\Theta_t)$ that satisfy the first order condition for profit maximization (16) and the market clearing condition (17). Therefore, the equilibrium wage for skill Θ_t, $W_t(\Theta_t)$, can be solved as a solution to the differential equation (16) using the production function (14), matching function (18) and the initial condition $\lim_{\Theta_t \to -\infty} W_t(\Theta_t) = 0$.

To better understand how relative wages between high and low skill workers respond to changes in various factors in equilibrium, it is useful to rearrange the first order condition (16) to derive the following

[34]A location-scale family is a family of distributions formed by translation and rescaling of a standard family member. That is, if a random variable X whose distribution function is a member of the family, $\Pr(X \leq x) = \Phi\left(\frac{x-m}{\sigma}\right)$ holds for some location parameter m, scale parameter $\sigma > 0$, and distribution function $\Phi(\cdot)$ of a standard family member. Examples include uniform, normal, logistic, Cauchy, extreme value distribution, and 2-parameter exponential distribution.

expression for the equilibrium return to skill:

$$\underbrace{\frac{d\ln W_t(\Theta_t)}{d\Theta_t}}_{\text{Return to skill}} = \underbrace{\frac{\partial \ln Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)}{\partial \Theta_t}}_{\text{Partial elasticity of output}} \underbrace{\left(\frac{W_t(\Theta_t)}{Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)}\right)^{-1}}_{\text{Labor share}}. \tag{19}$$

The return to skill consists of two components. The first term, which we call the 'partial elasticity of output', is the proportional change of output associated with a marginal skill change, holding productivity of job constant. The second term is the inverse of the worker's share of revenue, or labor share. Therefore, the return to skill for a worker is determined by the worker's marginal contribution to output (relative to a marginally less skilled worker) as well as the share of output accrued to the worker. As we show next, it is important to understand the distinct roles played by each of theses component in the changes in the return to skill.

6.2 Closed Form Formula for Returns to Skill

With the Cobb-Douglas production function ($\rho = 0$), the return to skill (19) is identical across skill levels, consistent with our empirical model. In this case, the scale parameters of the distributions, thus the slope of the matching function, play important roles in determining the return to skill. The following Proposition shows that the return to skill can be derived in a closed form. The proofs for this and subsequent results are provided in Appendix C.

Proposition 2. *If $\rho = 0$, then*

$$\frac{d\ln W_t(\Theta_t)}{d\Theta_t} = \lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}. \tag{20}$$

To derive (20), recall that, $\ln Y_t(\Theta_t, Z_t) = \lambda_t \Theta_t + \gamma_t Z_t$ when $\rho = 0$. More skilled workers produce more at all jobs, which is captured by the partial elasticity of output λ_t, but they also work at jobs with higher productivity. Taking into account this sorting effect, the proportional increase in output for higher skilled workers, which we call the 'total elasticity of output' is

$$\frac{d\ln Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)}{d\Theta_t} = \frac{\partial \ln Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)}{\partial\Theta_t} + \frac{\partial \ln Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)}{\partial Z_t}\frac{d\hat{Z}_t(\Theta_t)}{d\Theta_t} = \lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}, \tag{21}$$

which is equal to (20). Wages and output increase in skill at the same proportion, because wages are a constant fraction of output.[35] By combining Equations (19) and (20), we get the following formula for

[35] Constant labor share results from constant marginal rate of technical substitution between Θ_t and Z_t.

the labor share:

$$\frac{W_t(\Theta_t)}{Y(\Theta_t, \hat{Z}_t(\Theta_t))} = \alpha_t, \quad \text{where } \alpha_t \equiv \frac{\lambda_t}{\lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}}. \tag{22}$$

The labor share is the ratio between the partial and total elasticity of output, reflecting the share of a worker's marginal contribution out of the total output increase for a more productive worker-job pair.

The formula (20), first derived by Sattinger (1980) and Gabaix and Landier (2008) under specific distributional assumptions,[36] reveals that technology, λ_t and γ_t, as well as the relative heterogeneity between workers and jobs is important for the equilibrium return. When workers are relatively homogeneous compared to jobs, the matching function is steep, and a slightly more skilled worker is assigned to a much more productive job. This generates a large proportional changes in output and wage, amplifying the innate skill differences across workers. On the other hand, when jobs are relatively homogeneous compared to workers, the sorting effect is small, and the differences in wage mainly reflects the skill differences.

The Cobb-Douglas case provides an intuitive mechanism for the main drivers of the return to skill and also provides a clear mapping to our empirical model in which log wages are linear in skills. In Subsections 6.3 and 6.4, we use Equations (20) and (22) to account for the time patterns of the estimated return to skill.

6.3 Recovering Demand and Supply Factors

In this subsection, we combine our minimum distance estimation results from Section 5 and the assignment model assuming a Cobb-Douglas production function to recover changes in demand and supply factors. The strategy is as follows. First, we note that Equation (20) implies

$$\mu_t = \lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}. \tag{23}$$

Next, the labor share parameter α_t is recovered from additional data on the labor share. Then, we can use Equation (22) to recover $\lambda_t = \alpha_t \mu_t$ from the measured labor share and the estimated return to skill. By assuming Θ_t and Z_t are normally distributed, the scale parameter of the skill distribution, σ_{Θ_t}, is recovered from the estimated skill variance. Then, the scale parameter of the skill distribution, along with μ_t and λ_t, gives $\gamma_t \sigma_{Z_t} = (\mu_t - \lambda_t)\sigma_{\Theta_t}$, although we cannot separately identify γ_t and σ_{Z_t} without additional information. We assume that the labor market is segmented between a sector for college workers and another sector for non-college workers, and that all parameters can differ across sectors. Therefore, we

[36] Sattinger (1980) assumes normal distribution and Gabaix and Landier (2008) use 2-parameter exponential distribution.

use the baseline results estimated separately by sector (i.e., μ_t from Figure 7 and $\mathrm{Var}(\theta_t)$ from Figure 12).

Supply Factor We assume that skills are correlated with the workers' observed characteristics but are not fully observed by an econometrician. Specifically, skill for individual i in year t is

$$\Theta_{i,t} = g_t(\boldsymbol{x}_{i,t}) + \theta_{i,t}.$$

From (20), the equilibrium wage function can be written as $\ln W_t(\Theta_t) = \zeta_t + \mu_t \Theta_t$ for some constant ζ_t. Assuming that $\ln W_{i,t}$ is the sum of a skill-related component that reflects the equilibrium wage function and a transitory non-skill component, $\varepsilon_{i,t}$, it can be written as

$$\ln W_{i,t} = \zeta_t + \mu_t \Theta_{i,t} + \varepsilon_{i,t} = \underbrace{\zeta_t + \mu_t g_t(\boldsymbol{x}_{i,t})}_{\equiv f_t(\boldsymbol{x}_{i,t})} + \underbrace{\mu_t \theta_{i,t} + \varepsilon_{i,t}}_{= w_{i,t}},$$

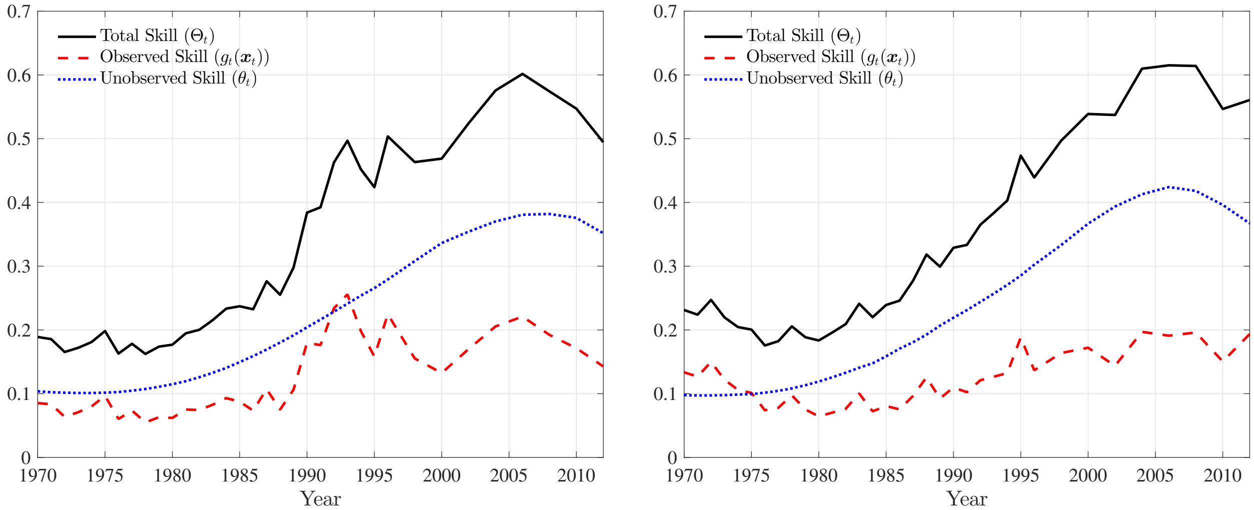
which is equivalent to or empirical log wage equation (1). The scale of the skill distribution is easily identified from the variance:

$$\sigma^2_{\Theta_t} = \mathrm{Var}(\Theta_t) = \frac{\mathrm{Var}(\ln W_t) - \mathrm{Var}(\varepsilon_t)}{\mu_t^2}.$$

Therefore, the second stage estimates of μ_t and $\mathrm{Var}(\varepsilon_t)$ can be used to compute the variance of skills in each period. Figure 17 shows the time trends in skill variances for college and non-college workers. The skill variance increased steeply from the early 1980s to the mid-2000s for both college and non-college workers, with most of the increase driven by unobserved skills.

Demand Factors Recovering the output elasticity from the factor income share is a commonly used method, but implementing it in this case is complicated by the need to know the education sector-specific labor shares. The difficulty arises from the fact that data on value added by workers with different levels of education is not available. To overcome this difficulty, we assume that the labor share in our model can be approximated by average industry-level labor share for workers in each sector. Suppose that there are $j = 1, \ldots, J$ industries with value added $V_{j,t}$ and labor compensation $L_{j,t}$ in year t. Let $N_{j,s,t}$ be the number of workers in industry j, sector s, and year t, and let $N_{s,t} = \sum_{j=1}^{J} N_{j,s,t}$ be the total number of workers in sector s and year t in all industries. Then, the average labor share in sector s and year t is

$$\sum_{j=1}^{J} \frac{N_{j,s,t}}{N_{s,t}} \frac{L_{j,t}}{V_{j,t}}.$$



(a) Non-College (b) College

Figure 17: Skill Variances

We use data on labor share and number of college and non-college workers in the U.S. by 31 ISIC[37] industries from the World KLEMS dataset,[38] which is available until 2010. The number of workers is computed by counting male workers ages 25-64. Figure 18 shows the average labor share by sector. Overall, average labor shares in the two sectors tend to comove. The average labor share is slightly lower for college workers until the late 1990s, but the gap closes afterwards.

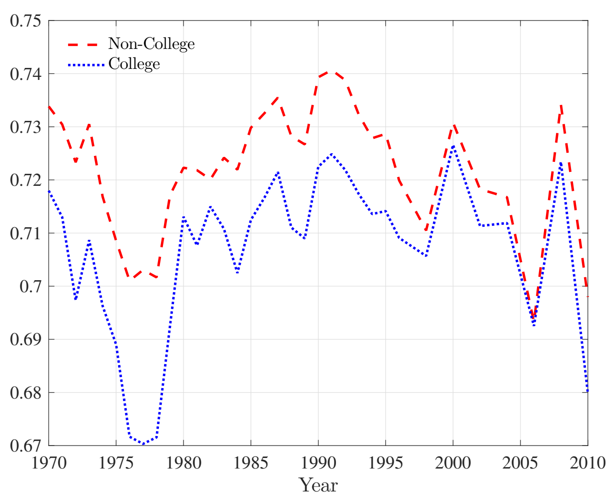
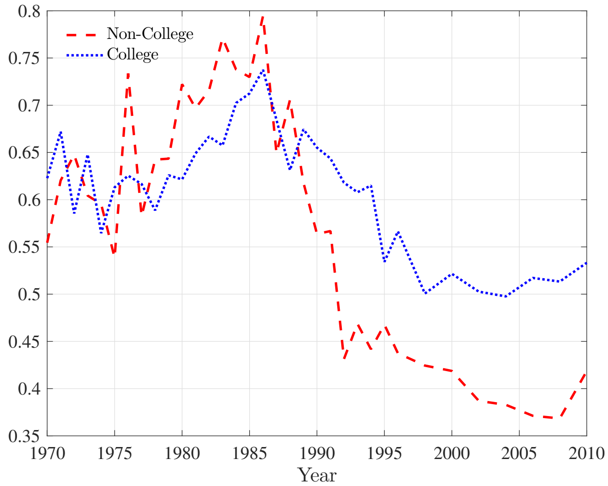


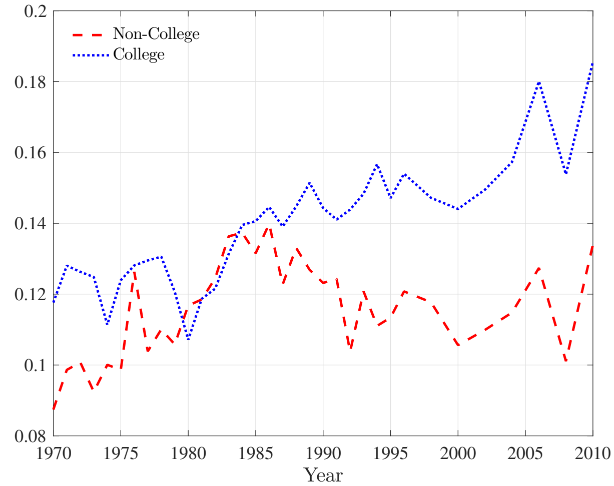
Figure 18: Average Labor Share

[37]The International Standard Industrial Classification of All Economic Activities (ISIC) is a United Nations industry classification system.

[38]See Jorgenson, Ho, and Samuels (2012) for description.



(a) λ_t



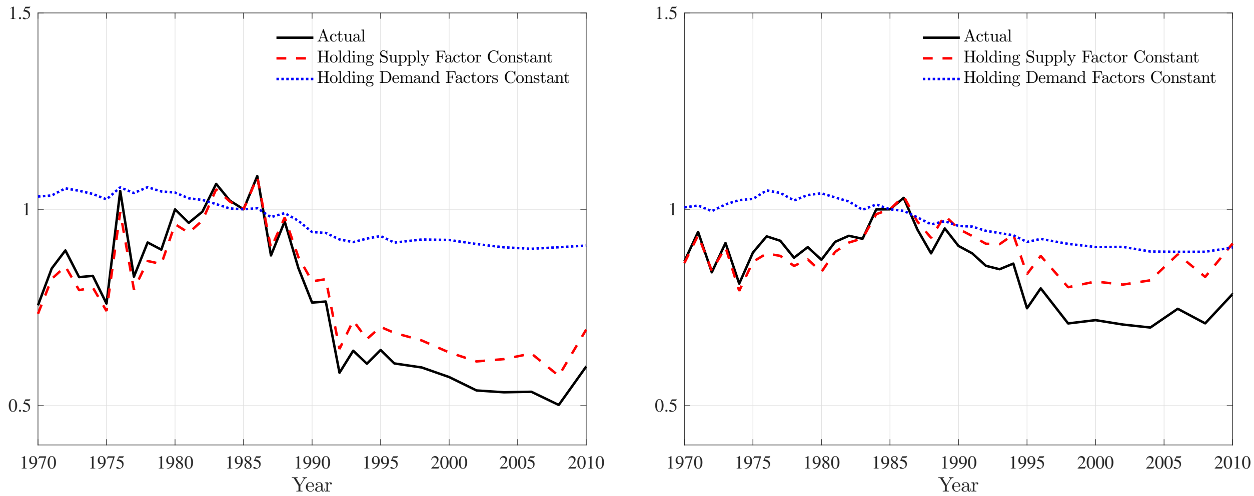
(b) $\gamma_t \sigma_{Z_t}$

Figure 19: Demand Factors

Figure 19 shows the demand factors calculated from the average labor share and the returns to skills. The time patterns of the partial elasticity of output λ_t are similar to those of the returns to skills μ_t, because the time variation in the average labor share is smaller than the variation in μ_t. For both college and non-college workers, λ_t increases between the mid-1970s and the mid-1980s, then decrease until the early 2000s. The rise and fall of λ_t is much more pronounced for non-college workers. The time patterns of the other demand factor, $\gamma_t \sigma_{Z_t}$, is quite different from those of λ_t: it shows sustained growth since the early 1980s for college workers while it falls modestly for non-college workers after 1985. During the 1970s, $\gamma_t \sigma_{Z_t}$ only increases for non-college workers.

6.4 Effects of Demand and Supply Factors on Returns to Skill

We assess the contributions of each factor to the evolution of the returns to skills. Figure 20 shows the actual returns as well as counterfactual returns when the demand factors (λ_t and $\gamma_t \sigma_{Z_t}$) or the supply factor (σ_{Θ_t}) are held constant at their 1985 values. When the supply factor is held constant, the counterfactual return moves closely with the actual return in both sectors, suggesting that a large part of observed change in μ_t is demand-driven. The supply factor plays a relatively minor role, especially for non-college workers, and its role is reversed around the mid-1980s: it weakens the demand-driven growth in the return to skill before the mid-1980s, and then it reinforces the downward trend afterwards, explaining half of the overall decrease in μ_t between 1985 and 2010 for college workers.



(a) Non-College

(b) College

Figure 20: Actual and Counterfactual Returns (μ_t)

6.5 Other Factors Driving the Changes in the Returns to Skills

Shifts in the Level of Skills or Job Productivity One unfortunate consequence of the Cobb-Douglas assumption is that the location parameters are irrelevant for the return to skill, as they lead to proportional changes in wages for all workers, keeping relative wages constant. When $\rho \neq 0$, a change in the intercept of the matching function induced by shifts in the location parameters m_{Z_t} or m_{Θ_t} affects the return to skill through a non-proportional change in output. In this case, the return to skill is heterogeneous across skill levels, so log wages are no longer linear in skill as in our empirical specification. However, the qualitative predictions of the theoretical model are still useful to interpret the empirical results in a broader context.

Proposition 3. *An increase in m_{Z_t} or a decrease in m_{Θ_t} raises $d \ln W_t(\Theta_t)/d\Theta_t$ for all Θ_t if and only if $\rho(\sigma_{Z_t} - \sigma_{\Theta_t}) > 0$.*

To simplify discussion, we focus on the case $\rho < 0$ where the worker-job complementarity is stronger than the Cobb-Douglas production function, which is equivalent to the following strict log-supermodularity condition:

$$\frac{\partial^2 \ln Y_t(\Theta_t, Z_t)}{\partial \Theta_t \partial Z_t} > 0.$$

From the matching function (18), it is easy to see that an increase in m_{Z_t} (or a decrease in m_{Θ_t}) shifts up $\hat{Z}_t(\Theta_t)$ for all workers, which raises the partial elasticity of output due to the log-supermodularity. Therefore, an increase in m_{Z_t} leads to larger returns to skill for all workers as long as its effect on the labor share is small enough, which is the case when $\sigma_{Z_t} < \sigma_{\Theta_t}$. On the other hand, when $\sigma_{Z_t} > \sigma_{\Theta_t}$, an increase in m_{Z_t} raises the the labor share and completely offsets the effects on the partial elasticity of output.

This mechanism, first introduced by Costinot and Vogel (2010),[39] resembles classical theory of wage determination where relative wages are determined by relative demand and supply. When relative demand for high skill workers rises (or relative supply decreases), their jobs go to workers with lower skill.[40] Although all workers end up with better jobs, high skills workers gain more than proportionally because of strong worker-job complementarity.

Thus, the fall in the return to skill since the mid-1980s could have come from an increasingly skilled workforce or the disappearance of good jobs.[41] Quantifying this mechanism is challenging, however, because the location of the skill distribution is not identified without additional assumptions when skills are unobserved.

Minimum Wage In addition to demand and supply factors, institutional factors such as unions and minimum wages can also be important determinants of wage inequality (e.g., Fortin and Lemieux, 1997). Consider the effects of the minimum wage on the return to skill in our assignment model.

Proposition 4. *Suppose that employers must pay at least the minimum wage $\underline{W}_t > 0$. Then, (i) there exists $\underline{\Theta}_t > -\infty$ such that workers with skill Θ_t are employed if and only if $\Theta_t \geq \underline{\Theta}_t$; (ii) an increase in $\underline{W}_t$ raises $\underline{\Theta}_t$ and reduces $d \ln W_t(\Theta_t)/d\Theta_t$ for all $\Theta_t \geq \underline{\Theta}_t$. Moreover, if $\rho = 0$, then, for $\Theta_t \geq \underline{\Theta}_t$,*

$$\frac{d \ln W_t(\Theta_t)}{d\Theta_t} = \frac{\lambda_t}{\alpha_t + (1 - \alpha_t) \exp\left(\frac{\lambda_t}{\alpha_t}(\underline{\Theta}_t - \Theta_t)\right)}. \tag{24}$$

Introducing the minimum wage destroys low skill jobs where output is too low to cover the minimum wage, but it does not affect the matching function, and, therefore, has no effects on the slope of the wage function. Thus, an increase in the minimum wage just shifts the wage function up everywhere by the same amount, keeping the difference in wage levels the same across skill levels (above the minimum), which means the slope of log wages as a function of skill declines.

[39]Costinot and Vogel (2010) and Sampson (2014) also derive Equation (19) in a one-to-many matching model where the labor share is constant. The constant labor share implies that the matching function becomes a sufficient statistic for the return to skill, but the matching function is no longer independent of technology due to the one-to-many matching assumption.

[40]More precisely, an increase in the relative demand for high skill workers is a monotone likelihood ratio dominant change in the distribution of Z_t. Since it implies a first order stochastic dominant change, by Equation (17), the matching function $\hat{Z}_t(\Theta_t)$ increases for all Θ_t. An increase in the location parameter does not necessarily imply a monotone likelihood ratio dominant change, even though it holds for normal distribution.

[41]The changes in the distribution of skills and jobs could reflect actual distributional shifts in the domestic economy, but it could also reflect the effects of globalization. For example, when jobs can move freely across countries, the return to skill is determined by the world distribution of skills and jobs.

The return to skill in the Cobb-Douglas case, Equation (24), shows that the effects of the minimum wage are concentrated among low skill workers. The labor share (denominator) is 1 for the minimum wage worker, and it decreases and converges to the labor share without the minimum wage law, α_t, for higher skill workers (i.e., as $\Theta_t \to \infty$). Likewise, the return to skill is λ_t for the minimum wage worker, and it converges to the return to skill without the minimum wage law, λ_t/α_t, for higher skill workers.

The negative relationship between the minimum wage and wage inequality, especially at the lower tail of the wage distribution, is consistent with empirical findings of DiNardo, Fortin, and Lemieux (1996) and Lee (1999), who study the effects of the decline in the real value of the federal minimum wage in the U.S. during the 1980s. However, it is somewhat inconsistent with the time patterns of our estimates of the return to skill. Although our minimum distance results suggest that the return to skill increased during the first half of the 1980s, especially for non-college workers; The return then decreased sharply during the second half of the 1980s despite a continued decline in the minimum wage. Moreover, the 2SLS results suggest that the return to skill continuously declined throughout the 1980s.

7 Conclusion

Economists have struggled to determine the underlying causes of rising wage inequality over the past few decades. Most efforts have relied on large repeated samples of cross-sectional data, attributing the growth in residual inequality to rising returns to unobserved skill while assuming that the distribution of unobserved skills has remained constant over time. More recent studies have often attempted to incorporate additional measures of worker skills or job tasks, continuing to rely on data drawn from different samples of workers over time. While these efforts have yielded important insights, they are not without limitations.

This paper takes a very different approach, demonstrating that traditional panel data sets can be used to separately identify changes in the returns to unobserved skill from changes in the distributions of unobserved skill and in the distribution of transitory non-skill shocks. Based on transparent identifying assumptions, we show that a simple 2SLS strategy can be used to estimate the returns to unobserved skill over time, even when lifecycle skill growth varies across individuals due to systematic unobserved heterogeneity and idiosyncratic shocks. Once skill returns have been identified, it is straightforward to identify and estimate the evolution of skill (and skill growth) distributions as well as distributions of transitory non-skill shocks. None of this requires measuring the tasks workers perform or efforts to directly measure worker skill levels.

Using panel data on the wages of American men from the PSID, we show that accounting for changes in the distributions of skills and the volatility of wages is critical in estimating the evolution of returns to unobserved skills. Our estimates reveal that these returns were fairly stable or increasing in the 1970s and early 1980s, but then fell sharply after 1985, especially among non-college workers. The decline in returns was offset by a stronger increase in the variance of unobserved skill beginning in the early 1980s, driven by increasing variation in lifecycle skill growth. These conclusions stand in stark contrast to the prevailing view, which attributes rising residual inequality to rising returns.

To understand why the returns to skill have fallen since the mid-1980s, we develop an assignment model of the labor market in which workers of heterogeneous skill levels are matched to different jobs. We further show conditions under which this framework produces equilibrium wage functions like those commonly assumed in the empirical literature on wage inequality. Combining labor shares with our estimated skill returns and skill distributions, we identify changes in production technology and decompose changes in skill returns into supply vs. demand effects. Our estimates suggest that the fall in demand for skill explains most of the decline in returns for non-college men, while both supply and demand shifts are important for college men.

Our findings are broadly consistent with the growing polarization literature, while highlighting the challenges faced by the simplest SBTC hypothesis (e.g., Card and DiNardo, 2002), which argues that a steady increase in demand for skills has driven the increase in inequality. Further work is required to understand why skill demand declined during the 1990s, precisely when technological progress appears to have accelerated (cum, 2002). Equally important, our results suggest that more attention should be devoted to understanding the dramatic increase in unobserved skill inequality, stemming largely from growing differences in lifecycle (post-school) skill growth across workers with similar experience and education levels.

References

2002. "Investment-Specific Technical Change in the United States (1947âĂŞ2000): Measurement and Macroeconomic Consequences." *Review of Economic Dynamics* 5 (2):243–284.

Abowd, John M. and David Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica* 57 (2):411–445.

Acemoglu, Daron. 1999. "Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence." *American Economic Review* 89 (5):1259–1278.

———. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40 (1):7–72.

Acemoglu, Daron and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, vol. 4, edited by O. Ashenfelter and D. Card, chap. 12. Elsevier, 1043–1171.

Ales, Laurence, Musab Kurnaz, and Christopher Sleet. 2015. "Technical Change, Wage Inequality, and Taxes." *American Economic Review* 105 (10):3061–3101.

Autor, David and Michael Handel. 2013. "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages." *Journal of Labor Economics* 31 (S1):S59–S96.

Autor, David, Frank Levy, and Richard Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118 (4):1279–1333.

Autor, David H. and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (5):1553–97.

Autor, David H, Lawrence F Katz, and Melissa S Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90 (2):300–323.

Baker, Michael and Gary Solon. 2003. "Earnings Dynamics and Inequality among Canadian Men, 1976âĂŞ1992: Evidence from Longitudinal Income Tax Records." *Journal of Labor Economics* 21 (2):289–321.

Becker, Gary. 1964. *Human Capital*. New York: Columbia University Press.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4):813–846.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Lifecycle of Earnings." *Journal of Political Economy* 75:352–365.

Blundell, Richard, Michael Graber, and Magne Mogstad. 2015. "Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family." *Journal of Public Economics* 127:58 – 73.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *American Economic Review* 98 (5):1887–1921.

Blundell, Richard and Ian Preston. 1998. "Consumption Inequality and Income Uncertainty." *Quarterly Journal of Economics* 113 (2):603–640.

Bonhomme, Stephane and Jean-Marc Robin. 2010. "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics." *Review of Economic Studies* 77 (2):491–533.

Bowlus, Audra J. and Chris Robinson. 2012. "Human Capital Prices, Productivity, and Growth." *American Economic Review* 102 (7):3483–3515.

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel. 2015. "Accounting for Changes in Between-Group Inequality." Working Paper 20855, National Bureau of Economic Research.

Caines, Colin, Florian Hoffmann, and Gueorgui Kambourov. 2017. "Complex-Task Biased Technological Change and the Labor Market." *Review of Economic Dynamics* 25:298–319.

Card, David and John DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20 (4):733–783.

Card, David and Thomas Lemieux. 1996. "Wage Dispersion, Returns to Skill, and Black-White Wage Differentials." *Journal of Econometrics* 74 (2):319–361.

Carroll, Christopher D. 1992. "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence." *Brookings Papers on Economic Activity* 1992 (2):61–156.

Caselli, Francesco. 1999. "Technological Revolutions." *American Economic Review* 89 (1):78–102.

Castex, Gonzalo and Evgenia Kogan Dechter. 2014. "The Changing Roles of Education and Ability in Wage Determination." *Journal of Labor Economics* 32 (4):685–710.

Costinot, Arnaud and Jonathan Vogel. 2010. "Matching and Inequality in the World Economy." *Journal of Political Economy* 118 (4):747–786.

DeBacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos. 2013. "Rising inequality: transitory or persistent? New evidence from a panel of US tax returns." *Brookings Papers on Economic Activity* 2013 (1):67–142.

Deming, David. forthcoming. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics* .

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64 (5):1001–1044.

Fortin, Nicole M. and Thomas Lemieux. 1997. "Institutional Changes and Rising Wage Inequality: Is There a Linkage?" *Journal of Economic Perspectives* 11 (2):75–96.

Friedman, Milton and Simon Kuznets. 1954. *Income from Independent Professional Practice*. National Bureau of Economic Research.

Gabaix, Xavier and Augustin Landier. 2008. "Why has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123 (1):49–100.

Galor, Oded and Omer Moav. 2000. "Ability-Biased Technological Transition, Wage Inequality, and Economic Growth." *Quarterly Journal of Economics* 115 (2):469–497.

Galor, Oded and Daniel Tsiddon. 1997. "Technological Progress, Mobility, and Economic Growth." *American Economic Review* 87 (3):363–382.

Gottschalk, Peter and Robert Moffitt. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers on Economic Activity* 25 (2):217–272.

———. 2009. "The Rising Instability of US Earnings." *Journal of Economic Perspectives* 23 (4):3–24.

Guvenen, Fatih. 2009. "An Empirical Investigation of Labor Income Processes." *Review of Economic dynamics* 12 (1):58–79.

Guvenen, Fatih and Burhanettin Kuruscu. 2010. "A Quantitative Analysis of the Evolution of the US Wage Distribution, 1970-2000." In *NBER Macroeconomics Annual 2009, Volume 24*. University of Chicago Press, 227–276.

Haider, Steven J. 2001. "Earnings Instability and Earnings Inequality of Males in the United States: 1967âĂŞ1991." *Journal of Labor Economics* 19 (4):799–836.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. “Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967âĂŞ2006.” *Review of Economic Dynamics* 13 (1):15 – 51. Special issue: Cross-Sectional Facts for Macroeconomists.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2010. “The Macroeconomic Implications of Rising Wage Inequality in the United States.” *Journal of Political Economy* 118 (4):681–722.

Heckman, James J, Lance Lochner, and Christopher Taber. 1998. “Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents.” *Review of Economic Dynamics* 1 (1):1 – 58.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. “Estimating Vector Autoregressions with Panel Data.” *Econometrica* 56 (6):1371–1395.

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2011. “Sources of Lifetime Inequality.” *American Economic Review* 101 (7):2923–2954.

Jorgenson, Dale W, Mun S Ho, and Jon D Samuels. 2012. “A Prototype Industry-Level Production Account for the United States, 1947-2010.” Second World KLEMS Conference, Harvard University, August 9, 2012.

Jovanovic, Boyan. 1998. “Vintage Capital and Inequality.” *Review of Economic Dynamics* 1 (2):497–530.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. “Wage Inequality and the Rise in Returns to Skill.” *Journal of Political Economy* 101 (3):410–42.

Katz, Lawrence F. and David H. Autor. 1999. “Changes in the Wage Structure and Earnings Inequality.” In *Handbook of Labor Economics*, *Handbook of Labor Economics*, vol. 3, edited by O. Ashenfelter and D. Card, chap. 26. Elsevier, 1463–1555.

Katz, Lawrence F. and Kevin M. Murphy. 1992. “Changes in Relative Wages, 1963-1987: Supply and Demand Factors.” *Quarterly Journal of Economics* 107 (1):35–78.

Lee, David S. 1999. “Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?” *Quarterly Journal of Economics* 114 (3):977.

Lemieux, Thomas. 2006. “Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?” *American Economic Review* 96 (3):461–498.

Lindenlaub, Ilse. 2017. “Sorting Multidimensional Types: Theory and Application.” *Review of Economic Studies* Forthcoming.

Lochner, Lance and Youngki Shin. 2014. "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks." Working Paper 20068, National Bureau of Economic Research.

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics* 9 (2):508–523.

MaCurdy, Thomas. 2007. "A Practitioner's Approach to Estimating Intertemporal Relationships Using Longitudinal Data: Lessons from Applications in Wage Dynamics." In *Handbook of Econometrics*, *Handbook of Econometrics*, vol. 6, Part A, edited by James J. Heckman and Edward E. Leamer, chap. 62. Elsevier, 4057 – 4167.

Meghir, Costas and Luigi Pistaferri. 2004. "Income Variance Dynamics and Heterogeneity." *Econometrica* 72 (1):1–32.

Moffitt, Robert A. and Peter Gottschalk. 2002. "Trends in the Transitory Variance of Earnings in the United States." *Economic Journal* 112 (478):C68–C73.

———. 2012. "Trends in the Transitory Variance of Male Earnings Methods and Evidence." *Journal of Human Resources* 47 (1):204–236.

Rosen, Sherwin. 1981. "The Economics of Superstars." *The American Economic Review* 71 (5):845–858.

Sampson, Thomas. 2014. "Selection into Trade and Wage Inequality." *American Economic Journal: Microeconomics* 6 (3):157–202.

Sattinger, Michael. 1979. "Differential Rents and the Distribution of Earnings." *Oxford Economic Papers* 31 (1):60–71.

———. 1980. *Capital and the Distribution of Labor Earnings*. North-Holland.

———. 1993. "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature* 31 (2):831–880.

Scheuer, Florian and Iván Werning. 2015. "The Taxation of Superstars." Working Paper 21323, National Bureau of Economic Research.

Spitz-Oener, Alexandra. 2006. "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure." *Journal of Labor Economics* 24 (2):235–270.

Terviö, Marko. 2008. "The Difference That CEOs Make: An Assignment Model Approach." *American Economic Review* 98 (3):642–668.

Violante, Giovanni L. 2002. "Technological Acceleration, Skill Transferability, and the Rise in Residual Inequality." *Quarterly Journal of Economics* 117 (1):297–338.

Appendix

A Identification Results

A.1 Identification with $\varepsilon_{i,t} \sim ARMA(1,q)$

We demonstrate identification for the model in Section 2.2.1 generalized so that the transitory component $\varepsilon_{i,t}$ follows an $ARMA(1,q)$ process. That is, $\varepsilon_{i,t} = \xi_{i,t}$ for $t = c_i + 1$ and, for $t > c_i + 1$,

$$\varepsilon_{i,t} = \rho \varepsilon_{i,t-1} + \sum_{j=0}^{\min\{q,t-c_i-1\}} \beta_j \xi_{i,t-j},$$

where $\beta_0 = 1$.

Identification of ρ Let $k = q + 1$. Then for (c, t, t') such that $c < t$ and $t' - t \geq k$,

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'} - \rho \varepsilon_{t'-1} | c) = \mathrm{Cov}\left(\varepsilon_t, \sum_{j=0}^{\min\{q,t-c-1\}} \beta_j \xi_{t'-j} \middle| c \right) = 0.$$

Therefore, for $t' - t \geq k$,

$$\mathrm{Cov}(w_t, w_{t'}|c) - \rho \,\mathrm{Cov}(w_t, w_{t'-1}|c) = \mathrm{Cov}(w_t, w_{t'} - \rho w_{t'-1}|c) = \mu_t(\mu_{t'} - \rho \mu_{t'-1}) \,\mathrm{Var}(\theta_t|c).$$

By taking the ratio between cohort c and $\tilde{c}$, we get

$$\frac{\mathrm{Cov}(w_t, w_{t'}|c) - \rho \,\mathrm{Cov}(w_t, w_{t'-1}|c)}{\mathrm{Cov}(w_t, w_{t'}|\tilde{c}) - \rho \,\mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})} = \frac{\mathrm{Var}(\theta_t|c)}{\mathrm{Var}(\theta_t|\tilde{c})}.$$

Similarly, for $t'' - t \geq k$,

$$\frac{\mathrm{Cov}(w_t, w_{t''}|c) - \rho \,\mathrm{Cov}(w_t, w_{t''-1}|c)}{\mathrm{Cov}(w_t, w_{t''}|\tilde{c}) - \rho \,\mathrm{Cov}(w_t, w_{t''-1}|\tilde{c})} = \frac{\mathrm{Var}(\theta_t|c)}{\mathrm{Var}(\theta_t|\tilde{c})}.$$

By combining these two equations, we get

$$\frac{\mathrm{Cov}(w_t, w_{t'}|c) - \rho \,\mathrm{Cov}(w_t, w_{t'-1}|c)}{\mathrm{Cov}(w_t, w_{t'}|\tilde{c}) - \rho \,\mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})} = \frac{\mathrm{Cov}(w_t, w_{t''}|c) - \rho \,\mathrm{Cov}(w_t, w_{t''-1}|c)}{\mathrm{Cov}(w_t, w_{t''}|\tilde{c}) - \rho \,\mathrm{Cov}(w_t, w_{t''-1}|\tilde{c})}, \tag{25}$$

which is a second-order equation in ρ. Therefore, ρ is identified if there exists ρ that solves (25).

A simple sufficient condition for ρ to be identified is that there exists $(t, c, \tilde{c})$ such that $\mathrm{Var}(\theta_t|c) = \mathrm{Var}(\theta_t|\tilde{c})$ and $\mathrm{Cov}(w_t, w_{t'-1}|c) \neq \mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})$. Then ρ is given by

$$\rho = \frac{\mathrm{Cov}(w_t, w_{t'}|c) - \mathrm{Cov}(w_t, w_{t'}|\tilde{c})}{\mathrm{Cov}(w_t, w_{t'-1}|c) - \mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})}.$$

Identification of μ_t For $t' - t \geq k$, suppose that there exists $(c, \tilde{c})$ such that $\mathrm{Var}(\theta_{t-1}|c) \neq \mathrm{Var}(\theta_{t-1}|\tilde{c})$ and $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|\tilde{c})$. Then

$$\begin{aligned}
&\frac{\left[\mathrm{Cov}(w_t, w_{t'}|c) - \rho\,\mathrm{Cov}(w_t, w_{t'-1}|c)\right] - \left[\mathrm{Cov}(w_t, w_{t'}|\tilde{c}) - \rho\,\mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})\right]}{\left[\mathrm{Cov}(w_{t-1}, w_{t'}|c) - \rho\,\mathrm{Cov}(w_{t-1}, w_{t'-1}|c)\right] - \left[\mathrm{Cov}(w_{t-1}, w_{t'}|\tilde{c}) - \rho\,\mathrm{Cov}(w_{t-1}, w_{t'-1}|\tilde{c})\right]}\\
=&\frac{\mu_t(\mu_{t'} - \rho\mu_{t'-1})\left[\mathrm{Var}(\theta_t|c) - \mathrm{Var}(\theta_t|\tilde{c})\right]}{\mu_{t-1}(\mu_{t'} - \rho\mu_{t'-1})\left[\mathrm{Var}(\theta_{t-1}|c) - \mathrm{Var}(\theta_{t-1}|\tilde{c})\right]}\\
=&\frac{\mu_t}{\mu_{t-1}}
\end{aligned}$$

In this way, we can identify μ_t for $t \leq \bar{t} - k$.

For $t > \bar{t} - k$, consider $t' < t - k$. Then

$$\frac{\mathrm{Cov}(w_{t'}, w_t|c) - \rho\,\mathrm{Cov}(w_{t'}, w_{t-1}|c)}{\mathrm{Cov}(w_{t'}, w_{t-1}|c) - \rho\,\mathrm{Cov}(w_{t'}, w_{t-2}|c)} = \frac{\mu_{t'}(\mu_t - \rho\mu_{t-1})\,\mathrm{Var}(\theta_{t'}|c)}{\mu_{t'}(\mu_{t-1} - \rho\mu_{t-2})\,\mathrm{Var}(\theta_{t'}|c)} = \frac{\mu_t - \rho\mu_{t-1}}{\mu_{t-1} - \rho\mu_{t-2}} \tag{26}$$

Since ρ and μ_t for $t \leq \bar{t} - k$ are identified, we can sequentially identify μ_t for $t > \bar{t} - k$ using the above equation.

Identification of $\mathrm{Var}(\theta_t|c)$ For $t' - t \geq k$,

$$\mathrm{Var}(\theta_t|c) = \frac{\mathrm{Cov}(w_t, w_{t'}|c) - \rho\,\mathrm{Cov}(w_t, w_{t'-1}|c)}{\mu_t(\mu_{t'} - \rho\mu_{t'-1})}.$$

Therefore, $\mathrm{Var}(\theta_t|c)$ is identified for all $t \leq \bar{t} - k$.

Identification of β_j First, note that the ARMA(1,q) process can be written as an MA($t - c + 1$) process:

$$\varepsilon_{i,t} = \sum_{j=0}^{t-c_i-1} \tilde{\beta}_j \xi_{i,t-j},$$

where $\tilde{\beta}_j = 1$ for $j = 0$, $\tilde{\beta}_j = \rho\tilde{\beta}_{j-1} + \beta_j$ for $1 \leq j \leq q$, and $\tilde{\beta}_j = \rho\tilde{\beta}_{j-1}$ for $j > q$.

Then, for $t = c + 1$ and $j \geq 0$,

$$\frac{\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+j}|c)}{\mathrm{Var}(\varepsilon_t|c)} = \tilde{\beta}_j \frac{\mathrm{Var}(\xi_t|c)}{\mathrm{Var}(\xi_t|c)} = \tilde{\beta}_j.$$

Therefore, $\tilde{\beta}_j$'s are identified from cohort-specific autocovariances of ε_t, which can be obtained from

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+j}|c) = \mathrm{Cov}(w_t, w_{t+j}|c) - \mu_t \mu_{t+j} \mathrm{Var}(\theta_t|c).$$

Given ρ and $\tilde{\beta}_j$, β_j for $1 \leq j \leq q$ is identified as follows:

$$\beta_j = \tilde{\beta}_j - \rho \tilde{\beta}_{j-1}.$$

Identification of $\mathrm{Var}(\xi_t|c)$ For the initial period $t = c + 1$ for cohort c, $\mathrm{Var}(\xi_t|c) = \mathrm{Var}(\varepsilon_t|c)$. For $c + 1 < t \leq \bar{t} - k$,

$$\mathrm{Var}(\xi_t|c) = \mathrm{Var}(\varepsilon_t|c) - \sum_{j=1}^{\min\{q, t-c-1\}} \tilde{\beta}_j^2 \mathrm{Var}(\xi_{t-j}|c).$$

A.2 Identification with Heterogeneous Skill Growth Rates

We demonstrate identification for the model in Section 2.2.2 with systematic heterogeneity in lifecycle skill growth:

$$\theta_{i,t} = \theta_{i,t-1} + \tau_t(c_i)\delta_i + \tilde{\nu}_{i,t},$$

where $\tau_t(c) = 0$ for $e = t - c \geq \overline{e}$.

Identification of μ_t Note that μ_t / μ_{t-1} for $t > \underline{t} + k$ is identified if there exists some cohort c such that (i) the cohort is more than or equal to experience level $\overline{e}$ in some year $t' < t - k$ and (ii) the cohort is observed in years t', $t - 1$, and t. These require that $\overline{e} + k < 40$, which holds under our assumption $\overline{e} = 30$ and $k = 6$.

Moreover, μ_t / μ_{t-1} for $t \leq \underline{t} + k$ is identified if there exist two cohorts c and $\tilde{c}$ such that (i) both cohorts are more than or equal to experience level $\overline{e}$ in year $t - 1$, (ii) both cohorts are observed in years $t - 1$, t, and some year $t' \geq t + k$, and (iii) $\mathrm{Var}(\theta_{t-1}|c) \neq \mathrm{Var}(\theta_{t-1}|\tilde{c})$ and $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|\tilde{c})$. For the first two conditions to be satisfied, we need $\overline{e} + k < 39$. But we may need a stronger assumption to ensure that the third condition also holds.

Identification of τ_t By dividing the residual by μ_t, we get

$$\frac{w_{i,t}}{\mu_t} = \theta_{i,t} + \frac{\varepsilon_{i,t}}{\mu_t}.$$

If we take a first difference,

$$\Delta\left(\frac{w_{i,t}}{\mu_t}\right) = \Delta\theta_{i,t} + \Delta\left(\frac{\varepsilon_{i,t}}{\mu_t}\right) = \tau_t(c_i)\delta_i + \tilde{\nu}_{i,t} + \Delta\left(\frac{\varepsilon_{i,t}}{\mu_t}\right).$$

For (c, t, t') such that $\mathrm{Cov}(\Delta\varepsilon_t, \Delta\varepsilon_{t'}|c) = \mathrm{Cov}(\Delta\varepsilon_{t-1}, \Delta\varepsilon_{t'}|c) = 0$, we have

$$\begin{aligned}
\mathrm{Cov}\left(\Delta\left(\frac{w_{t-1}}{\mu_{t-1}}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)\middle|c\right) &= \mathrm{Cov}(\Delta\theta_{t-1}, \Delta\theta_{t'}|c) = \tau_{t-1}(c)\tau_{t'}(c)\,\mathrm{Var}(\delta|c), \\
\mathrm{Cov}\left(\Delta\left(\frac{w_{t}}{\mu_{t}}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)\middle|c\right) &= \mathrm{Cov}(\Delta\theta_{t}, \Delta\theta_{t'}|c) = \tau_{t}(c)\tau_{t'}(c)\,\mathrm{Var}(\delta|c).
\end{aligned}$$

By combining these two, we can identify changes in $\tau_t(c)$:

$$\frac{\tau_t(c)}{\tau_{t-1}(c)} = \frac{\mathrm{Cov}\left(\Delta\left(\frac{w_t}{\mu_t}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)|c\right)}{\mathrm{Cov}\left(\Delta\left(\frac{w_{t-1}}{\mu_{t-1}}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)|c\right)}.$$

Identification of $\mathrm{Var}(\delta|c)$ Once $\tau_t(c)$'s have been identified (up to a normalization $\tau_{t^*(c)}(c) = 1$ for some $t^*(c)$), $\mathrm{Var}(\delta|c)$ is also identified from

$$\mathrm{Var}(\delta|c) = \frac{\mathrm{Cov}\left(\Delta\left(\frac{w_t}{\mu_t}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)|c\right)}{\tau_t(c)\tau_{t'}(c)}.$$

Identification of $\mathrm{Cov}(\psi, \delta|c)$ For (c, t, t') such that $t' - t \geq k + 1$, $\mathrm{Cov}(\varepsilon_t, \Delta\varepsilon_{t'}|c) = 0$ and we get

$$\mathrm{Cov}\left(\frac{w_t}{\mu_t}, \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)\middle|c\right) = \mathrm{Cov}(\theta_t, \Delta\theta_{t'}|c) = \tau_{t'}(c)\,\mathrm{Cov}(\theta_t, \delta|c),$$

where

$$\mathrm{Cov}(\theta_t, \delta|c) = \mathrm{Cov}(\psi, \delta|c) + \mathrm{Var}(\delta|c)\sum_{j=0}^{t-c-1}\tau_{t-j}(c). \tag{27}$$

Therefore,

$$\mathrm{Cov}(\psi,\delta|c)=\frac{\mathrm{Cov}\left(\frac{w_t}{\mu_t},\Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)|c\right)}{\tau_{t'}(c)}-\mathrm{Var}(\delta|c)\sum_{j=0}^{t-c-1}\tau_{t-j}(c).$$

Identification of $\mathrm{Var}(\theta_t|c)$ For (c,t,t') such that $t'-t\geq k$, write

$$\theta_{i,t'}=\theta_{i,t}+\sum_{j=0}^{t'-t-1}\left[\tau_{t'-j}(c_i)\delta_i+\tilde{\nu}_{i,t'-j}\right].$$

Then

$$\mathrm{Cov}\left(\frac{w_t}{\mu_t},\frac{w_{t'}}{\mu_{t'}}\middle|c\right)=\mathrm{Cov}(\theta_t,\theta_{t'}|c)=\mathrm{Var}(\theta_t|c)+\mathrm{Cov}(\theta_t,\delta|c)\sum_{j=0}^{t'-t-1}\tau_{t'-j}(c).$$

Therefore,

$$\mathrm{Var}(\theta_t|c)=\mathrm{Cov}\left(\frac{w_t}{\mu_t},\frac{w_{t'}}{\mu_{t'}}\middle|c\right)-\mathrm{Cov}(\theta_t,\delta|c)\sum_{j=0}^{t'-t-1}\tau_{t'-j}(c).$$

Identification of $\mathrm{Var}(\tilde{\nu}_t|c)$ Note that

$$\mathrm{Var}(\theta_{t+1}|c)=\mathrm{Var}(\theta_t|c)+\mathrm{Var}(\delta|c)\tau_{t+1}(c)^2+2\,\mathrm{Cov}(\theta_t,\delta|c)\tau_{t+1}(c)+\mathrm{Var}(\tilde{\nu}_{t+1}|c).$$

Therefore,

$$\mathrm{Var}(\tilde{\nu}_{t+1}|c)=\mathrm{Var}(\theta_{t+1}|c)-\mathrm{Var}(\theta_t|c)-\mathrm{Var}(\delta|c)\tau_{t+1}(c)^2-2\,\mathrm{Cov}(\theta_t,\delta|c)\tau_{t+1}(c).$$

B GMM Estimates of Skill Returns, Over-Identification Tests, and Variance of Skill Growth

In this appendix, we report GMM estimates for the returns to skill using the same model and moments (i.e. lagged residuals serve as instruments) as with our 2SLS approach in Section 4 along with J-statistics to test for overidentification. We also report analogous GMM estimates that use both past and future wage residuals as instruments, reporting J-statistics to test the validity of the latter. Finally, we combine estimates using past vs. future residuals as instruments to estimate the variance of skill growth relative to lagged skill levels.

To begin, rewrite the two-period wage growth equation (11) as follows:

$$w_{i,t} - w_{i,t-2} = \left(\frac{\mu_t - \mu_{t-2}}{\mu_{t-2}}\right) w_{i,t-2} + u_{i,t}, \tag{28}$$

where $u_{i,t} = \varepsilon_{i,t} - \frac{\mu_t}{\mu_{t-2}}\varepsilon_{i,t-2} + \mu_t(\nu_{i,t-1} + \nu_{i,t})$.

Our model with skill serially uncorrelated skill shocks (Section 2.2.1) implies the following moment condition:

$$\mathbb{E}[w_{t'}u_t] = 0, \qquad \text{for } t' \leq t - 2 - k. \tag{29}$$

Under the stronger assumption that $\mathrm{Var}(\nu_t) = 0$ for all t, the following additional moment condition holds:

$$\mathbb{E}[w_{t''}u_t] = 0, \qquad \text{for } t'' \geq t + k. \tag{30}$$

Equation (30) will not hold when $\mathrm{Var}(\nu_{t-1}) + \mathrm{Var}(\nu_t) > 0$, and the IV estimate using future residuals as instruments is asymptotically biased with probability limit

$$\frac{\mathrm{Cov}(w_t - w_{t-2}, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})} = \left(\frac{\mu_t - \mu_{t-2}}{\mu_{t-2}}\right) + \frac{\mu_t}{\mu_{t-2}}\left(\frac{\mathrm{Var}(\nu_{t-1}) + \mathrm{Var}(\nu_t)}{\mathrm{Var}(\theta_{t-2})}\right) > \frac{\mu_t - \mu_{t-2}}{\mu_{t-2}}, \qquad \text{for } t' \geq t + k.$$

The difference between estimates using future and past residuals as instruments identifies the magnitude of the skill shock variance relative to the skill variance: for $t' \leq t - 2 - k$ and $t'' \geq t + k$,

$$\frac{\mathrm{Var}(\nu_{t-1}) + \mathrm{Var}(\nu_t)}{\mathrm{Var}(\theta_{t-2})} = \left(\frac{\mathrm{Cov}(w_t - w_{t-2}, w_{t''})}{\mathrm{Cov}(w_{t-2}, w_{t''})} - \frac{\mathrm{Cov}(w_t - w_{t-2}, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right)\left(1 + \frac{\mathrm{Cov}(w_t - w_{t-2}, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right)^{-1}. \tag{31}$$

B.1 Overidentification Tests

We begin by testing the moments in equation (30) using Hansen's J-test, assuming $k = 6$ and using the two nearest valid instruments. This amounts to using $w_{i,t-8}$ and $w_{i,t-9}$ (or $w_{i,t-10}$) for equation (29) and the first two available out of $w_{i,t+6}, w_{i,t+7}, w_{i,t+8}, w_{i,t+9}$ for (30).

Table 3 reports the two-step optimal GMM estimates (allowing for heteroskedasticity and serial correlation within individual) for the coefficient on $w_{i,t-2}$ along with Hansen's J-statistics when estimating the wage growth equation (28). Panel A reports estimates when moments from both equations (29) and (30) are used (i.e., lags and leads), while Panel B reports estimates when only the moment condition from equation (29) is used (i.e., lags only). The sample is restricted to be the same in both panels.[42]

Comparing the J-statistics in Panels A and B in Table 3, we can test the validity of using leads as instruments (i.e. moments in equation (30)). Since the differences are greater than 5.991 ($\chi^2_{0.05}$ with 2 degrees of freedom) except for 1979-1980 and 2002-2004, we reject the 'leads' moments in equation (30) at 5% significance level for 1981-2000. Moreover, all J-statistics in Panel B are smaller than 3.841 ($\chi^2_{0.05}$ with 1 degree of freedom), implying that we cannot reject the lags as instruments (i.e. moments in equation (29)) at the 5% level. Altogether, these results suggest that the lagged residuals are valid instruments, while the leads are not (in most years).

Also note that the estimates using both leads and lags as instruments are always greater than their counterparts using only the lags, consistent with the positive bias induced from using leads when there are skill growth shocks.

Table 3: GMM Estimates of Skill Return Growth using Leads and Lags as Instruments (Balanced Samples)

	1979-1980	1981-1983	1984-1986	1987-1989	1990-1992	1993-1995	1996-2000	2002-2004
A. 2 Nearest Valid Lags and 2 Nearest (Potentially Valid) Leads as Instruments								
Coeff. on $w_{i,t-2}$	-0.019	0.088*	0.053	0.007	-0.030	0.026	0.008	0.022
	(0.053)	(0.044)	(0.046)	(0.034)	(0.038)	(0.035)	(0.0235)	(0.035)
Observations	818	1,251	1,325	1,356	1,313	1,311	1,375	777
J-Statistic	4.400	10.392	11.743	9.579	9.461	6.991	8.922	1.646
B. 2 Nearest Valid Lags as Instruments								
Coeff. on $w_{i,t-2}$	-0.070	-0.010	-0.065	-0.057	-0.103*	-0.025	-0.041	-0.003
	(0.056)	(0.053)	(0.055)	(0.040)	(0.046)	(0.039)	(0.029)	(0.0389)
Observations	818	1,251	1,325	1,356	1,313	1,311	1,375	777
J-Statistic	0.009	0.187	0.632	0.869	0.064	0.238	0.107	0.016

Notes: GMM estimates for a regression of $(w_{i,t} - w_{i,t-2})$ on $w_{i,t-2}$. Panel A uses as instruments the 2 nearest available lags from $(w_{t-8}, w_{t-9}, w_{t-10})$ and 2 nearest available leads from $(w_{t+6}, ..., w_{t+9})$. Panel B uses only the 2 lags as instruments.
* denotes significance at 0.05 level.

B.2 Inferring Relative Magnitude of Skill Shocks

Table 4 reports GMM estimates using only lags or leads as instruments where all available observations are used (i.e. samples are not restricted to be the same across specifications). Panel A reports estimates when only the moments in equation (29) are used (i.e. 2 nearest valid lags). These results are analogous to

[42] Because use of both leads and lags requires observations that are as many as 19 years apart, this restriction reduces the sample size substantially relative to that used in our baseline 2SLS analysis (see Tables 1 and 2). Panel A of Table 4 below reports GMM estimates when this sample selection is not imposed. Those results are directly comparable and quite similar to those in Tables 1 and 2.

the 2SLS estimates in Tables 1 and 2, using the same samples. Comparing estimates across the tables, we see that they are quite similar. Panel B reports GMM estimates when only the moments in equation (30) are used (i.e., 2 nearest potentially valid leads), also based on all available observations. Finally, we compare the estimates in Panels A and B using equation (31) to estimate the relative importance of skill growth shocks. These estimates are reported in Panel C. The variance of (two-year) skill growth relative to the variance of prior skill levels ranges from 0.16 to 0.29 over our entire sample period.

Table 4: GMM Estimates of Skill Return Growth using Leads vs. Lags as Instruments and Relative Skill Shock Variance (Unbalanced Samples)

	1979-1980	1981-1983	1984-1986	1987-1989	1990-1992	1993-1995	1996-2000	2002-2004
A. 2 Nearest Valid Lags as Instruments								
Coeff. on $w_{i,t-2}$	-0.033	-0.045	-0.044	-0.084*	-0.083*	-0.067	-0.076*	-0.090*
	(0.045)	(0.038)	(0.038)	(0.033)	(0.035)	(0.035)	(0.025)	(0.035)
Observations	1,349	2,077	2,188	2,245	2,189	2,095	2,122	1,377
B. 2 Nearest (Potentially Valid) Leads as Instruments								
Coeff. on $w_{i,t-2}$	0.165*	0.229*	0.193*	0.099*	0.067	0.087*	0.073*	0.115*
	(0.059)	(0.053)	(0.047)	(0.042)	(0.043)	(0.038)	(0.028)	(0.039)
Observations	1,500	2,229	2,159	2,100	2,042	1,994	2,178	1,249
C. Estimated Shock Variances Relative to Skill Variances								
$\frac{\text{Var}(\nu_{t-1})+\text{Var}(\nu_t)}{\text{Var}(\theta_{t-2})}$	.204	0.287	0.248	0.200	0.163	0.166	0.161	0.225

Notes: GMM estimates for a regression of $(w_{i,t} - w_{i,t-2})$ on $w_{i,t-2}$. Panel A uses 2 nearest available lags as instruments from $(w_{t-8}, w_{t-9}, w_{t-10})$. Panel B uses 2 nearest available leads as instruments from $(w_{t+6}, ..., w_{t+9})$. Panel C reports estimates of skill growth shock variance relative to skill variance based on equation (31). * denotes significance at 0.05 level.

C Proofs and Analytical Details for Assignment Model

C.1 Proof of Proposition 2

With the matching function (18), the first order condition (16) becomes

$$\frac{dW_t(\Theta_t)}{d\Theta_t} = \lambda_t \exp\left(\lambda_t \Theta_t + \gamma_t \hat{Z}_t(\Theta)\right) = \lambda_t \exp\left(\gamma_t \left(m_{Z_t} - \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}} m_{\Theta_t}\right) + \left(\lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}\right)\Theta_t\right).$$

By integrating the above equation, we get

$$W_t(\Theta_t) = \int_{-\infty}^{\Theta_t} \frac{dW_t(\Theta_t')}{d\Theta_t} d\Theta_t' = \left(\frac{\lambda_t}{\lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}}\right) \exp\left(\gamma_t \left(m_{Z_t} - \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}} m_{\Theta_t}\right) + \left(\lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}\right)\Theta_t\right).$$

By taking logs and differentiating with respect to Θ_t, we get (20).

C.2 Proof of Proposition 3

We can write the inverse of $d \ln W_t(\Theta_t)/d\Theta_t$ as follows:

$$\begin{aligned} W_t(\Theta_t)\left(\frac{dW_t(\Theta_t)}{d\Theta_t}\right)^{-1} &= \int_{-\infty}^{\Theta_t} \frac{dW_t(\Theta_t')}{d\Theta_t}\left(\frac{dW_t(\Theta_t)}{d\Theta_t}\right)^{-1} d\Theta_t' \\ &= \int_{-\infty}^{\Theta_t} \Psi\left(m_{Z_t}, m_{\Theta_t}, \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}, \Theta_t', \Theta_t\right)^{\frac{\lambda_t+\gamma_t}{\rho}-1} \exp\left((\lambda_t+\gamma_t)(\Theta_t' - \Theta_t)\right) d\Theta_t', \end{aligned}$$

where

$$\Psi\left(m_{Z_t}, m_{\Theta_t}, \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}, \Theta_t', \Theta_t\right) = \frac{\lambda_t + \gamma_t \exp\left(\rho\left[m_{Z_t} - \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}} m_{\Theta_t} + \left(\frac{\sigma_{Z_t}-\sigma_{\Theta_t}}{\sigma_{\Theta_t}}\right)\Theta_t'\right]\right)}{\lambda_t + \gamma_t \exp\left(\rho\left[m_{Z_t} - \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}} m_{\Theta_t} + \left(\frac{\sigma_{Z_t}-\sigma_{\Theta_t}}{\sigma_{\Theta_t}}\right)\Theta_t\right]\right)}.$$

The sign of the derivative of $\Psi(m_{Z_t}, m_{\Theta_t}, \sigma_{Z_t}/\sigma_{\Theta_t}, \Theta_t', \Theta_t)$ with respect to m_{Z_t} is identical to the sign of

$$\rho\left\{1 - \exp\left(\rho\left(\frac{\sigma_{Z_t} - \sigma_{\Theta_t}}{\sigma_{\Theta_t}}\right)(\Theta_t - \Theta_t')\right)\right\}.$$

Because ρ and $(\lambda_t+\gamma_t)/\rho - 1$ have always same signs and $\Theta_t \geq \Theta_t'$, $W_t(\Theta_t)\,(dW_t(\Theta_t)/d\Theta_t)^{-1}$ is decreasing in m_{Z_t} if and only if $\rho(\sigma_{Z_t} - \sigma_{\Theta_t}) > 0$, which implies that $d \ln W_t(\Theta_t)/d\Theta_t$ is increasing m_{Z_t} if and only if $\rho(\sigma_{Z_t} - \sigma_{\Theta_t}) > 0$. The effect of m_{Θ_t} can be shown similarly.

C.3 Proof of Proposition 4

Since the equilibrium assignment is exogenously determined, it is unaffected by the minimum wage. However, because $\lim_{\Theta_t \to -\infty} Y_t(\Theta_t, \hat{Z}_t(\Theta_t)) = 0 < \underline{W}_t$ and employers have an option not to operate, only sufficiently productive worker-job pairs engage in production. Because profit is increasing in Z_t, there exists a threshold job with productivity $\hat{Z}_t(\underline{\Theta}_t)$ that earns zero profit:

$$\Pi_t\left(\hat{Z}_t(\underline{\Theta}_t)\right) = Y_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right) - \underline{W}_t = 0. \tag{32}$$

By differentiating Equation (32) with respect to $\underline{W}_t$, we get

$$\frac{d\underline{\Theta}_t}{d\underline{W}_t} = \left(\frac{dY_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right)}{d\Theta_t}\right)^{-1}. \tag{33}$$

The equilibrium wage for $\Theta_t \geq \underline{\Theta}_t$ is

$$W_t(\Theta_t) = \underline{W}_t + \int_{\underline{\Theta}_t}^{\Theta_t} \frac{dW_t(\Theta'_t)}{d\Theta_t} d\Theta'_t. \tag{34}$$

Note that the slope of the wage function $dW_t(\Theta_t)/d\Theta_t = \partial Y_t(\Theta_t, \hat{Z}_t(\Theta_t))/\partial\Theta_t$ for $\Theta_t \geq \underline{\Theta}_t$ is not affected by $\underline{W}_t$. Therefore, by taking derivative of both sides of Equation (34) and using Equation (33), we get

$$\frac{dW_t(\Theta_t)}{d\underline{W}_t} = 1 - \frac{\partial Y_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right)}{\partial\Theta_t}\frac{d\underline{\Theta}_t}{d\underline{W}_t} = \frac{\partial Y_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right)}{\partial Z_t}\frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}\left(\frac{dY_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right)}{d\Theta_t}\right)^{-1} > 0,$$

which implies that an increase in the minimum wage raises the wage level $W_t(\Theta_t)$ for all workers while it reduces the return to skill $d\ln W_t(\Theta_t)/d\Theta_t$.

When $\rho = 0$, Equation (34) can be written as

$$\begin{aligned}
W_t(\Theta_t) =& \underline{W}_t + \left(\frac{\lambda_t}{\lambda_t + \gamma_t\frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}}\right)\exp\left(\gamma_t\left(m_{Z_t} - \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}m_{\Theta_t}\right)\right)\left[\exp\left(\left(\lambda_t + \gamma_t\frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}\right)\Theta_t\right) - \exp\left(\left(\lambda_t + \gamma_t\frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}\right)\underline{\Theta}_t\right)\right] \\
=& Y_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right) + \left(\frac{\lambda_t}{\lambda_t + \gamma_t\frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}}\right)\left[Y_t\left(\Theta_t, \hat{Z}_t(\Theta_t)\right) - Y_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right)\right] \\
=& \left(\frac{\lambda_t\sigma_{\Theta_t}}{\lambda_t\sigma_{\Theta_t} + \gamma_t\sigma_{Z_t}}\right)Y_t\left(\Theta_t, \hat{Z}_t(\Theta_t)\right) + \left(\frac{\gamma_t\sigma_{Z_t}}{\lambda_t\sigma_{\Theta_t} + \gamma_t\sigma_{Z_t}}\right)Y_t\left(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\right).
\end{aligned}$$

Then the labor share for $\Theta_t \geq \underline{\Theta}_t$ is

$$\begin{aligned}
\frac{W_t(\Theta_t)}{Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)} &= \left(\frac{\lambda_t \sigma_{\Theta_t}}{\lambda_t \sigma_{\Theta_t} + \gamma_t \sigma_{Z_t}}\right) + \left(\frac{\gamma_t \sigma_{Z_t}}{\lambda_t \sigma_{\Theta_t} + \gamma_t \sigma_{Z_t}}\right) \frac{Y_t\big(\underline{\Theta}_t, \hat{Z}_t(\underline{\Theta}_t)\big)}{Y_t\big(\Theta_t, \hat{Z}_t(\Theta_t)\big)} \\
&= \left(\frac{\lambda_t \sigma_{\Theta_t}}{\lambda_t \sigma_{\Theta_t} + \gamma_t \sigma_{Z_t}}\right) + \left(\frac{\gamma_t \sigma_{Z_t}}{\lambda_t \sigma_{\Theta_t} + \gamma_t \sigma_{Z_t}}\right) \exp\left(\left(\lambda_t + \gamma_t \frac{\sigma_{Z_t}}{\sigma_{\Theta_t}}\right)(\underline{\Theta}_t - \Theta_t)\right) \\
&= \alpha_t + (1 - \alpha_t) \exp\left(\frac{\lambda_t}{\alpha_t}(\underline{\Theta}_t - \Theta_t)\right). \qquad (35)
\end{aligned}$$

Combining Equation (14) with $\rho = 0$, Equation (19), and Equation (35) gives Equation (24).

D Calculating Standard Errors

Let $m = 1, \dots, M$ be the index of moments. Let $d_{i,m}$ be the indicator of whether individual i contributes to the m^{th} moment $\text{Cov}(w_t, w_{t'}|s, E_j)$. That is, both $w_{i,t}$ and $w_{i,t'}$ are non-missing and $s_{i,t} = s_{i,t'} = s$ and $e_{i,t} \in E_j$. Also let $p_m(\boldsymbol{\Lambda}) = \text{Cov}(w_t, w_{t'}|s, E_j, \boldsymbol{\Lambda})$. Then we can write

$$h_m(z_i, \boldsymbol{\Lambda}) = d_{i,m}\big[w_{i,t} w_{i,t'} - p_m(\boldsymbol{\Lambda})\big],$$

where z_i includes $w_{i,t}$ $d_{i,m}$ for all t and m for individual i. Let $\boldsymbol{h}(z, \boldsymbol{\Lambda}) = (h_1(z, \boldsymbol{\Lambda}), \dots, h_M(z, \boldsymbol{\Lambda}))$. Then the following moment condition holds for the true parameter $\boldsymbol{\Lambda}_0$:

$$\mathbb{E}[\boldsymbol{h}(z, \boldsymbol{\Lambda}_0)] = 0.$$

The minimum distance estimator $\hat{\boldsymbol{\Lambda}}$ is equivalent to the GMM estimator that solves

$$\min_{\boldsymbol{\Lambda}} \left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{h}(z_i, \boldsymbol{\Lambda})\right]' \boldsymbol{W} \left[\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{h}(z_i, \boldsymbol{\Lambda})\right],$$

where $\boldsymbol{W} = \text{diag}[\frac{N^2}{N_1^2}, \dots, \frac{N^2}{N_M^2}]$ and $N_m = \sum_{i=1}^{N} d_{i,m}$.

The GMM estimator $\hat{\boldsymbol{\Lambda}}$ is asymptotically normal with a variance matrix $\boldsymbol{V} = (\boldsymbol{H}'\boldsymbol{W}\boldsymbol{H})^{-1}(\boldsymbol{H}'\boldsymbol{W}\Omega\boldsymbol{W}\boldsymbol{H})(\boldsymbol{H}'\boldsymbol{W}\boldsymbol{H})^{-1}$ where $\boldsymbol{H}$ is the Jacobian of the vector of moments, $\mathbb{E}[\partial \boldsymbol{h}(z, \boldsymbol{\Lambda}_0)/\partial \boldsymbol{\Lambda}']$, and $\boldsymbol{\Omega} = \mathbb{E}[\boldsymbol{h}(z, \boldsymbol{\Lambda}_0)\boldsymbol{h}(z, \boldsymbol{\Lambda}_0)']$.

Both expectations are replaced by sample averages and evaluated at the estimated parameter:

$$\hat{\boldsymbol{H}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \boldsymbol{h}(z_i,\hat{\boldsymbol{\Lambda}})}{\partial \boldsymbol{\Lambda}'} = \boldsymbol{W}^{-\frac{1}{2}}\frac{\partial \boldsymbol{p}(\hat{\boldsymbol{\Lambda}})}{\partial \boldsymbol{\Lambda}'},$$

$$\hat{\boldsymbol{\Omega}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}(z_i,\hat{\boldsymbol{\Lambda}})\boldsymbol{h}(z_i,\hat{\boldsymbol{\Lambda}})',$$

where $\boldsymbol{W}^{-\frac{1}{2}} = \text{diag}[\frac{N_1}{N},\ldots,\frac{N_M}{N}]$.

We can test r linear parameter restrictions $H_0 : \boldsymbol{R\Lambda} = 0$ using Wald test statistic:

$$\text{Wald} = N(\boldsymbol{R}\hat{\boldsymbol{\Lambda}})'(\boldsymbol{R}\hat{\boldsymbol{V}}\boldsymbol{R}')^{-1}\boldsymbol{R}\hat{\boldsymbol{\Lambda}} \xrightarrow{d} \chi^2_r.$$

E Additional Empirical Results

To examine whether attrition affects the residual autocovariances reported in Figure 3, Figure 21 shows the autocovariances, $\text{Cov}(w_b, w_t)$ for $6 \leq t - b \leq 16$, where the samples for each line (representing different base years, b) are restricted to those individuals observed in the base year as well as at least one of the last two years used for that line (i.e. $t - b = 15$ or 16 in early years or $t - b = 14$ or 16 in later years with biannual surveys). Comparing Figures 3 and 21, the autocovariance patterns are quite similar, indicating little effect of sample attrition (due to non-response or retirement) on the key moments used in our analysis.

Figure 22 shows the residual autocovariances for individuals with 1-15 years of experience in the base years. Regardless of the base year, the autocovariances are typically declining from late 1980s through the 1990s as in Figures 3 (full sample) and 4 (men with 16-30 years experience) in the text. The lines also shift upwards over time, consistent with rising skill variances.

For the model estimated separately for college and non-college men in Section 5.2, Figures 23-26 report estimates (with 95% confidence intervals) for the variance of initial skills by cohort (Figure 23), experience patterns for the variance of skill shocks (Figure 24), and time trends (Figure 25) and experience patterns (Figure 26) for the variance of transitory non-skill shocks.

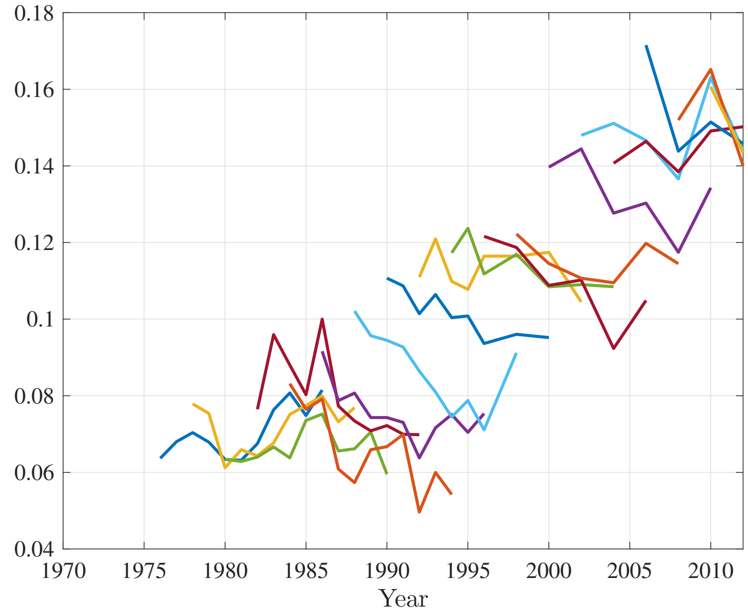


Figure 21: Log Wage Residual Autocovariances ('Balanced' Sample)

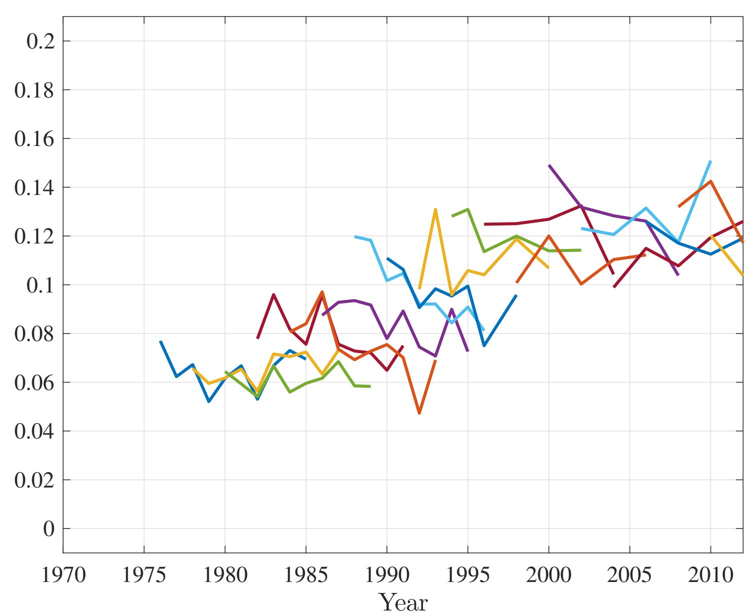
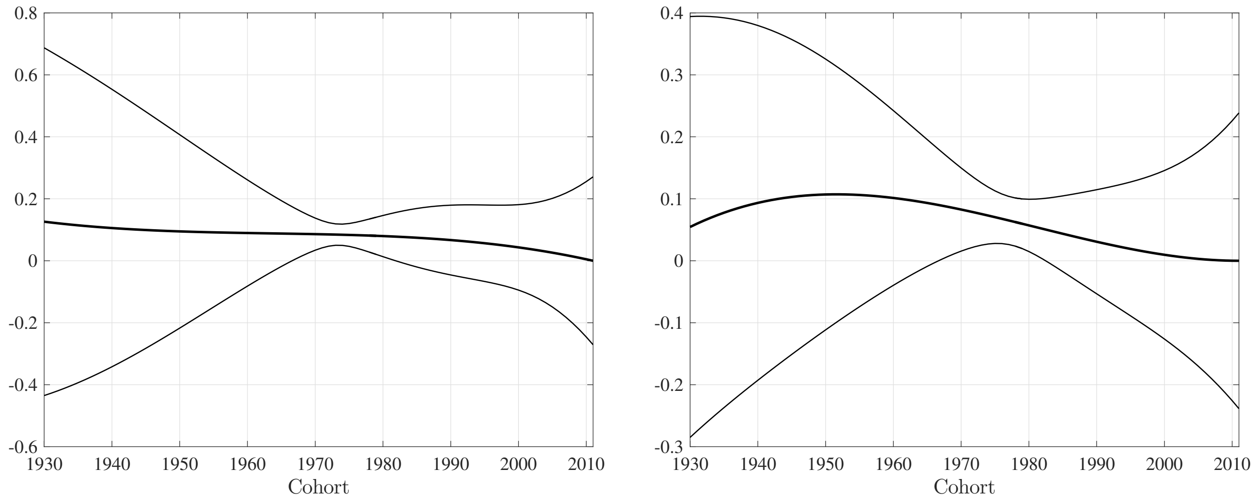
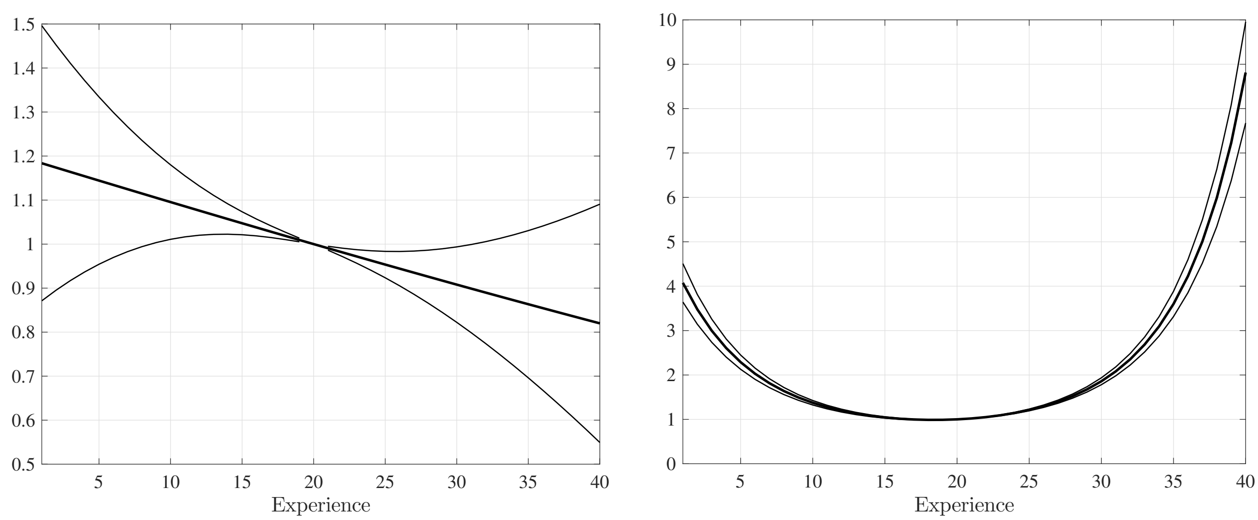


Figure 22: Autocovariances for Log Wages (1-15 Years of Experience)



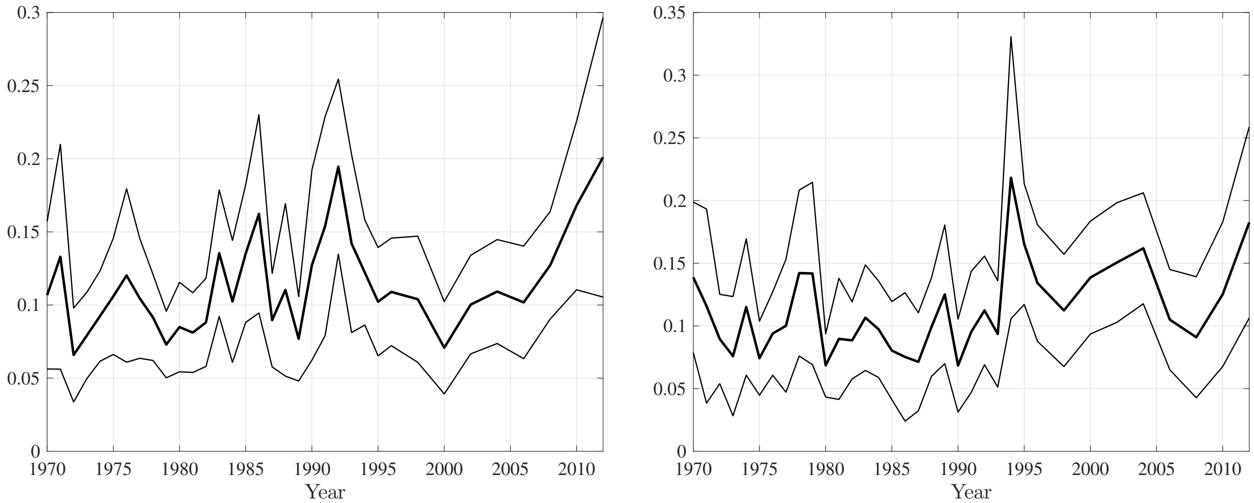
(a) Non-College (b) College

Figure 23: Variance of Initial Skill by Education



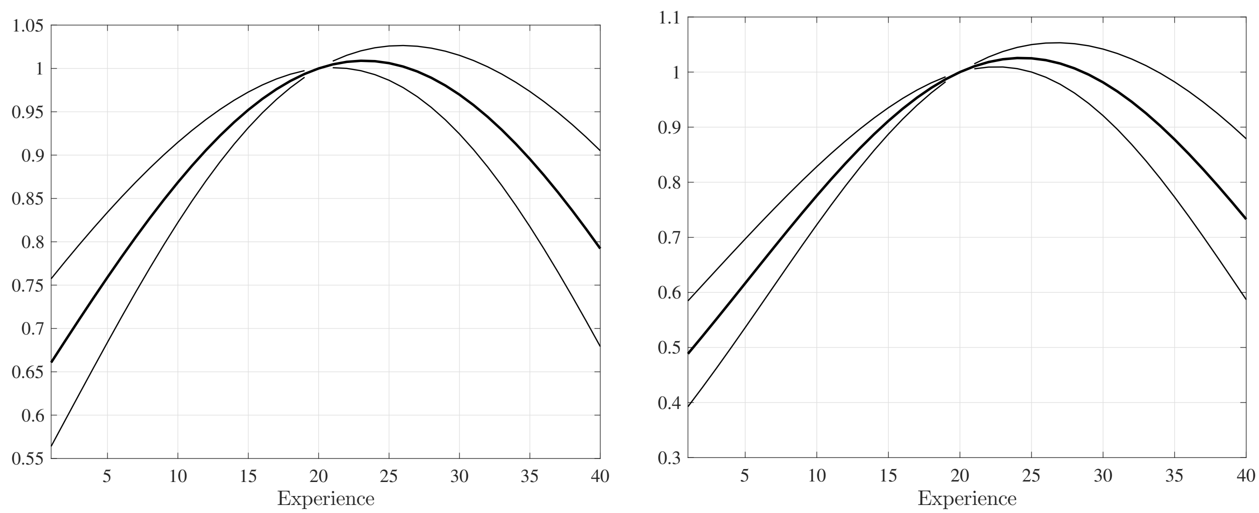
(a) Non-College (b) College

Figure 24: Experience Patterns for the Variances of Skill Shocks, $\phi(e)$, by Education



(a) Non-College (b) College

Figure 25: Time Trends in the Variances of Transitory Non-Skill Shocks, $\omega(t)$, by Education



(a) Non-College (b) College

Figure 26: Experience Patterns for the Variances of Transitory Non-Skill Shocks, $\kappa(e)$, by Education