Staff Working Paper/Document de travail du personnel—2017-61

Last updated: November 15, 2024

The Evolution of Unobserved Skill Returns in the U.S.: A New Approach Using Panel Data

by Lance Lochner,[1] Youngmin Park[2] and Youngki Shin[3]

[1] University of Western Ontario

[2] Canadian Economic Analysis Department
Bank of Canada
ypark@bankofcanada.ca

[3] McMaster University

DOI: https://doi.org/10.34989/swp-2017-61 | **ISSN** 1701-9397

Acknowledgements

An earlier (and much different) version of this paper circulated under the title "Wage Dynamics and Returns to Unobserved Skill." For valuable comments on this line of research, we thank Patrick Bayer, Magne Mogstad, Terry Moon, Fabrizio Perri, Gianluca Violante and Thomas Lemieux as well as numerous seminar and conference participants over the years. Lance Lochner acknowledges generous support from the Social Sciences and Humanities Research Council. The views expressed here are those of the authors and do not necessarily reflect those of the Bank of Canada. Parts of this analysis were first performed using the SIPP Synthetic Beta (SSB) on the Synthetic Data Server housed at Cornell University, which is funded by NSF Grant #SES-1042181. These data are for public use and may be accessed by researchers outside secure Census facilities. For more information, visit https://www.census.gov/programs-surveys/sipp/guidance/sipp-synthetic-beta-data-product.html. Final results for this paper were obtained by Census Bureau staff using the authors' programs and the SIPP Completed Gold Standard Files. This does not imply endorsement by the Census Bureau of any methods, results, opinions, or views presented in this paper.

Abstract
Economists disagree about the factors driving the substantial increase in residual wage inequality in the United States over the past few decades. To identify changes in the returns to unobserved skills, we make a novel assumption about the dynamics of skills (especially among older workers) rather than about the stability of skill distributions across cohorts, as is standard. We show that this assumption is supported by data on test score dynamics for older workers in the Health and Retirement Study. Using survey data from the Panel Study of Income Dynamics and administrative data from the Internal Revenue Service and the Social Security Administration, we estimate that the returns to unobserved skills declined substantially since the mid-1980s despite a sizable increase in residual inequality. Instead, the variance of skills rose over this period due to increasing variability in idiosyncratic lifecycle skill growth. We extend our framework to consider occupational differences in returns to skill and multiple unobserved skills and show that returns to skill display similar patterns for workers employed in cognitive, routine and social occupations.

Topics: Econometric and statistical methods; Labour markets
JEL codes: C, C2, C23, J, J2, J24, J3, J31

Résumé
Les économistes ne s'entendent pas sur les facteurs qui sont à l'origine de l'augmentation importante des inégalités résiduelles de salaire observée aux États-Unis depuis les dernières décennies. Pour identifier les changements relativement au rendement des compétences non observées, nous faisons une hypothèse originale au sujet de la dynamique des compétences (en particulier chez les travailleurs âgés) plutôt qu'au sujet de la stabilité de la distribution des compétences entre les cohortes, comme c'est habituellement le cas. Nous montrons que cette hypothèse est appuyée par des données sur la dynamique des résultats de tests de travailleurs âgés, tirées de la Health and Retirement Study (une étude américaine sur la santé et la retraite). À partir des données d'enquête de la Panel Study of Income Dynamics (une étude par panel sur la dynamique des revenus) et de données administratives de deux agences fédérales américaines (l'Internal Revenue Service et la Social Security Administration), nous estimons que le rendement des compétences non observées a enregistré un recul considérable depuis le milieu des années 1980 en dépit d'une hausse appréciable des inégalités résiduelles. La variance des compétences a plutôt augmenté au cours de la période en raison d'une plus grande variabilité de la croissance idiosyncrasique des compétences durant le cycle de vie. Nous élargissons notre cadre pour prendre en considération l'effet des différences professionnelles sur le rendement des compétences ainsi que de nombreuses compétences non observées, et nous montrons que le rendement des compétences est semblable qu'un emploi exige des compétences cognitives, sociales ou autres.

Sujets : Méthodes économétriques et statistiques; Marchés du travail
Codes JEL : C, C2, C23, J, J2, J24, J3, J31

1 Introduction

Wage inequality has risen considerably in the United States since the 1960s. The long-term increases in wage differentials by education and experience are widely attributed to rising returns to skill (Bound and Johnson, 1992; Katz and Murphy, 1992). In addition to these trends, wage inequality within narrowly defined groups (e.g. by race, education, and age/experience) also rose substantially. Figure 1 reports these trends for men based on data from the Panel Study of Income Dynamics (PSID) used in this study.[1]

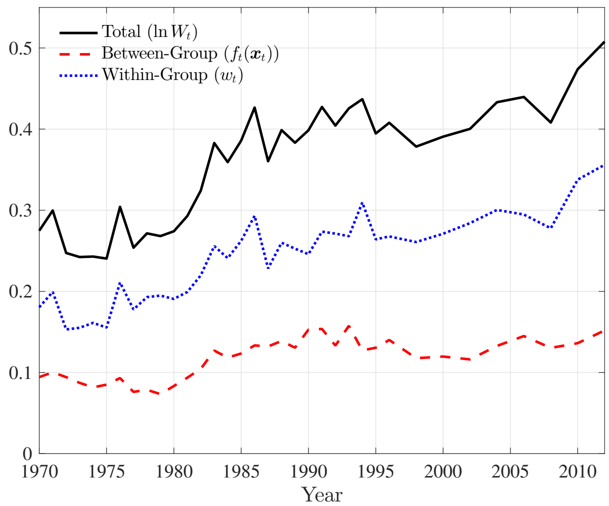


Figure 1: Between- and Within-Group Variances of Log Wages

Since the seminal work of Juhn, Murphy, and Pierce (1993), economists have often equated 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). This interpretation, along with the rising returns to 'observable' skills (i.e, education, experience), motivated an enormous and still influential literature on skill-biased technical change (SBTC).[2] More recent task-based models of the labor market also explore the influence of automation and globalization on wage and employment inequality between and within groups by altering the demand for both observed and unobserved skills (e.g., Autor, Levy, and Murnane, 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2022; Acemoglu and Loebbing, 2022).

In an important challenge to 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

[1]In obtaining between- and within-group log wage variances, we condition log wages on potential experience, race/ethnicity, and 7 educational attainment categories, separately by year and college vs. non-college status. See Section 3.1 for details.

[2]Many early theoretical 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).

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 wage volatility have also contributed to rising residual inequality. Whether the rise in residual inequality reflects an increase in returns to unobserved skills, growing unobserved skill inequality, or increased wage volatility unrelated to skills (or measurement error) is critical to our understanding of both the economic causes and welfare consequences of rising inequality. This paper provides a new approach for disentangling the importance of these distinct economic forces.

Several recent studies have turned to richer data to incorporate additional measures of skills 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 during adolescence by the Armed Forces Qualifying Test (AFQT), declined substantially between the late-1980s and late-2000s in the United States. Deming (2017) confirms this finding but further estimates that the returns to social skills have risen across these two cohorts.[3] 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 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 cohorts. For example, Castex and Dechter (2014) and Deming (2017) 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.[4]

Studies of long-term changes in residual wage inequality or the returns to unobserved skills largely rely on repeated samples of cross-section data, making it difficult to distinguish changes in skill returns

[3] Edin et al. (2022) estimate relatively stable returns to cognitive skills and rising returns to a measure of teamwork and leadership skills in Sweden.

[4] Autor and Handel (2013) find that person-level job tasks vary systematically across demographic groups within occupations, suggesting that worker skills and job tasks are likely to change within occupations as the workforce composition changes. Comparing skill requirements by occupation across editions of the Dictionary of Occupational Titles (DOT) and O*NET, Cavounidis et al. (2021) and Cortes, Jaimovich, and Siu (2023) document within-occupation changes in the skill/task content/requirements of jobs in the U.S. Using data with individual-level measures of job tasks, Spitz-Oener (2006) shows that most task changes in Germany over the 1980s and 1990s occurred within occupations.

from changes in the distributions of skills. As we show, panel data are naturally more useful. Intuitively, if heterogeneity in skills is important, then workers earning a high wage one one year should continue, on average, to earn a high wage many years later (even after the influence of transitory wage shocks has faded). As such, heterogeneity in unobserved skills implies that differences in wage residuals across workers should be predictive of long-term future residual differences. Moreover, an increase in the return to unobserved skills should lead to divergence in average log wage residuals across workers with different initial wage residuals. This is not what we observe during the late-1980s and 1990s.

Categorizing workers based on their log wage residual quartile in three different base years ($b = 1970$, 1980, and 1990), Figure 2 reports their average residuals 6–20 years later (i.e., years $t = b + 6, \ldots, b + 20$). Consistent with an important role for unobserved skills, those with higher wage residuals in any given base year also earn more, on average, up to 20 years later.[5] As discussed further below, the sharp convergence in lines over the late-1980s and 1990s suggests that the returns to skill *fell* over those years, despite modest growth in residual inequality at the time. Returns appear to be more stable in earlier and later years.

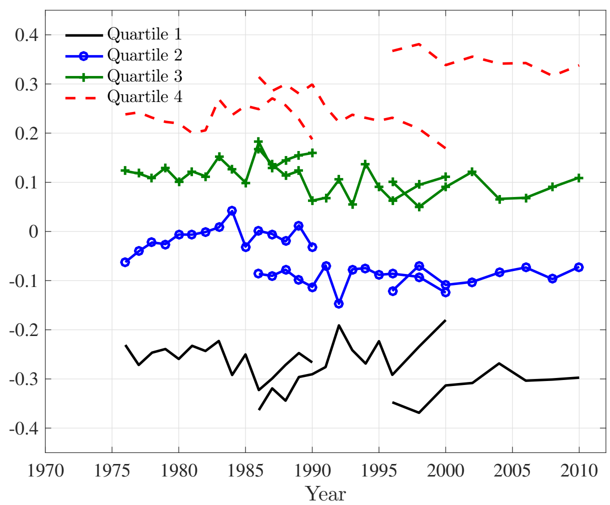


Figure 2: Average Predicted Log Wage Residuals by Baseline Residual Quartile

Notes:

We show that if unobserved skill growth is uncorrelated with sufficiently lagged skill levels and if non-skill wage shocks exhibit limited persistence, then a simple instrumental variable (IV) strategy can be used to estimate growth in skill returns over time. While endogenous skill investments may raise concerns about the skill-growth assumption for young workers, it is much more natural for older workers whose skill investments are likely to be negligible (Becker, 1964; Ben-Porath, 1967). Indeed, we show that panel data on cognitive test scores for older men in the Health and Retirement Study (HRS) support

[5]Differences in levels across lines for any given quartile are due to differences in base year wage distributions.

this assumption, as do several other specification tests. We also show that this assumption, along with the assumption of limited persistence in non-skill wage shocks, can be relaxed; although, we can no longer rely on a simple IV estimator and instead turn to more general moment-based approaches using the full autocovariance structure for log wage residuals.[6] Once the returns to skill have been estimated (from, e.g., older workers), it is straightforward to estimate the distribution of unobserved skills, skill growth, and non-skill shocks over time. Importantly, there is no need to observe independent measures of skills or what workers do on their jobs, enabling application of our approach in widely available panel data sets.

Using PSID data on log hourly wages, we estimate the evolution of returns to unobserved skills for American men from 1970 to 2012. Our main finding is that the returns to unobserved skills were relatively stable from 1970 to the mid-1980s, then *fell* considerably through the late-1980s and 1990s, stabilizing thereafter. The drop in estimated returns reflects the sharp convergence in predicted wage residuals conditional on earlier differences as documented in Figure 2 and is robust to different estimation strategies and assumptions about the dynamics of skills and non-skill wage shocks. The decline in skill returns appears to be stronger 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 flip side of declining returns is that the variance of unobserved skill has risen substantially since the early-1980s, driving the increase in residual wage inequality.[7] Consistent with stability of AFQT distributions among teenagers across NLSY cohorts (Altonji, Bharadwaj, and Lange, 2012), this increase is not driven by growth in the variance of early-career skill levels across cohorts. Instead, we find that the growing skill dispersion reflects an increase in the variance of idiosyncratic skill growth innovations, consistent with the notion of growing economic turbulence studied by Ljungqvist and Sargent (1998). We find little evidence of heterogeneity in systematic lifecycle skill growth, as studied by Lillard and Weiss (1979), MaCurdy (1982), Baker and Solon (2003), Guvenen (2009), and Moffitt and Gottschalk (2012).

A growing literature highlights differences in pay across occupations and the potential for different trends in the returns to heterogeneous skills (e.g., Autor, Levy, and Murnane, 2003; Acemoglu and Autor, 2011; Autor and Dorn, 2013; Deming, 2017). To explore these issues, we extend our analysis to consider occupation-specific wage schedules over a multi-dimensional skill vector. We show that our IV estimator identifies a weighted-average of returns across different (unobserved) skills and use occupation-stayers to

[6]The voluminous literature on earnings dynamics uses a similar moment-based estimation approach but focuses on a different set of questions 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).

[7]The widening skill growth distributions within education and experience groups are not accounted for in the composition adjustments of Lemieux (2006).

estimate the evolution of occupation-specific skill returns. Based on the PSID, we estimate very similar long-run declines in the returns to skills within routine, cognitive, and social occupations.

Given recent concerns about differences in the dynamics of log earnings residuals between survey and administrative data (see, e.g., Sabelhaus and Song, 2010), we reproduce key results using earnings measures from W-2 forms (collected by the the Internal Revenue Service, IRS) linked with several panels of the Survey of Income and Program Participation (SIPP). This analysis also indicates substantial long-run declines in average returns to unobserved skills; however, it suggests weaker (though not statistically different) declines in skill returns within cognitive relative to routine occupations since 1990.

This paper proceeds as follows. Section 2 describes our baseline assumptions used to identify and estimate the returns to skill over time using panel data on wages, contrasting these assumptions with those used in prior work (e.g. Juhn, Murphy, and Pierce, 1993; Lemieux, 2006; Castex and Dechter, 2014). We also test our main assumption on unobserved skill dynamics using cognitive test scores in the HRS. Section 3 describes the PSID data used for most of our empirical analysis and reports estimated returns to unobserved skill in the U.S. since the late-1970s. Sections 4 discusses identification of unobserved skill distributions, separately from the distributions of non-skill shocks, and provides estimates of these distributions over time. We also decompose the variance of skills into contributions from heterogeneity in initial skills and variation due to idiosyncratic lifecycle skill growth. Section 5 extends our analysis to account for differences across occupations and multiple skills. Finally, we confirm our PSID-based empirical findings with administrative data in Section 6 before concluding in Section 7.

2 Identifying and Estimating the Returns to Unobserved Skills

We consider the following specification for log wages motivated by the literature on unobserved skills (e.g., Juhn, Murphy, and Pierce, 1993; Card and Lemieux, 1994; Chay and Lee, 2000; Lemieux, 2006):

$$\ln W_{i,t} = f_t(\boldsymbol{x}_{i,t}) + w_{i,t} \tag{1}$$

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

where $W_{i,t}$ reflects wages for individual $i = 1, ..., N$ in period $t = \underline{t}, ..., \bar{t}$, $f_t(\boldsymbol{x}_{i,t})$ reflects the time-varying influence of observed characteristics $\boldsymbol{x}_{i,t}$ (e.g. education, race, experience), and $w_{i,t}$ is the log wage "residual" satisfying $\mathrm{E}[\theta_t|\boldsymbol{x}_t] = \mathrm{E}[\varepsilon_t|\boldsymbol{x}_t] = 0$.[8] The residual $w_{i,t}$ reflects the contributions of unobserved skill (equivalently, worker productivity) $\theta_{i,t}$ and idiosyncratic non-skill shocks $\varepsilon_{i,t}$, which may

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

include measurement error.[9] Note that average unobserved skill growth, which may vary by observable characteristics, is subsumed in changes in $f_t(\boldsymbol{x}_{i,t})$.[10] Individuals may come from different cohorts (i.e. different years of labor market entry), which we discuss further below.

As emphasized by Juhn, Murphy, and Pierce (1993), the returns to unobserved skills, reflected in $\mu_t > 0$, may evolve quite differently over time from the returns to observed measures of skills, reflected in $f_t(\cdot)$. Our analysis focuses on the log wage residual of equation (2) with the aim of identifying and estimating the returns to unobserved skill over time.[11] We also use the residual $w_{i,t}$ to identify and estimate the evolution of unobserved skill variation over time.

A few recent studies (e.g., Castex and Dechter, 2014; Deming, 2017) take advantage of skill measurements, or test scores, to aid in identification of the returns to skill. To facilitate discussion of these studies and to test our own assumptions, consider a (potentially) repeated skill measurement, $T_{i,t}$, in period t:

$$T_{i,t} = g_t(\boldsymbol{x}_{i,t}) + \tau\theta_{i,t} + \eta_{i,t}. \tag{3}$$

This specification allows test scores to vary with both observed factors and unobserved skills.[12] We assume that unobserved skills have the same effect on scores for the same test regardless of when the test is taken (i.e., τ is time-invariant). We also assume throughout our analysis that test measurement errors are serially uncorrelated and are uncorrelated with other observed variables $\boldsymbol{x}_t$, unobserved skills θ_t, and non-skill shocks ε_t.[13] It is useful to define test score residuals, $\tilde{T}_{i,t} \equiv T_{i,t} - g_t(\boldsymbol{x}_{i,t}) = \tau\theta_{i,t} + \eta_{i,t}$.

2.1 Prior Assumptions in the Literature

We briefly consider the strategies and assumptions previously employed in the literature on skill returns.

[9] Chay and Lee (2000), Card and Lemieux (1994), and Lemieux (2006) consider the same log wage residual decomposition; however, they assume that the variances of skills within observable groups (e.g. education, experience, race) are time invariant. Thus, their approaches only account for changes in the overall variance of unobserved skills due to changes in the composition of workers across observable types. In estimating the importance of these composition changes, Lemieux (2006) ignores any variation in the transitory component, $\varepsilon_{i,t}$; although, he also provides a separate analysis documenting increases in log wage measurement error over time. Our use of panel data facilitates a more general analysis.

[10] The assumption of separability between $\boldsymbol{x}_{i,t}$ and $\theta_{i,t}$ is both common and convenient, though not necessary. One can condition the analysis that follows on $\boldsymbol{x}_{i,t}$. Our empirical analysis separately studies non-college and college educated workers.

[11] Equations (1) and (2) imply wage *levels* that 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 this would imply log wages functions that are additively separable in 'prices' and skills. Instead, this non-linearity is consistent with assignment and task-based models of the labor market (see, e.g., Sattinger, 1993; Costinot and Vogel, 2010; Acemoglu and Autor, 2011). See Lochner, Park, and Shin (2018) for assumptions on the production technology and distribution of skills and firm productivity that yield wage functions given by equations (1) and (2) in a standard assignment model.

[12] Note that $g_{j,t}(\cdot)$ may reflect differential measurement quality across groups or differences in skills across groups (e.g., total skills measured by the test may be given by $\tilde{g}_{j,t}(\boldsymbol{x}_t) + \theta_t$, in which case $g_{j,t}(\boldsymbol{x}_t) = \tau_j \tilde{g}_{j,t}(\boldsymbol{x}_t)$).

[13] Specifically, we assume that $\mathrm{E}[\eta_t|\boldsymbol{x}_t] = \mathrm{Cov}(\eta_t, \theta_t|\boldsymbol{x}_t) = 0$ for all t, $\mathrm{Cov}(\eta_t, \theta_{t'}) = \mathrm{Cov}(\eta_t, \varepsilon_{t'}) = 0$ for all t, t', and $\mathrm{Cov}(\eta_t, \eta_{t'}) = 0$ for all $t \neq t'$.

2.1.1 Juhn, Murphy, and Pierce (1993)

By equating the increase in residual inequality with an increase in skill returns, μ_t, Juhn, Murphy, and Pierce (1993) assume that the cross-sectional distributions of unobserved skills and non-skill shocks are time-invariant. To "test" this assumption, they compare growth in the variance of residuals when following cohorts vs. experience groups over the period they examine (1963–1989). Unfortunately, this comparison is not very informative about the evolution of skill distributions or returns over time. To see why, let c reflect a cohort's year of labor market entry and assume for simplicity that $\mathrm{Cov}(\varepsilon_t, \theta_t|c) = 0$ and $\mathrm{Var}(\varepsilon_t|c) = \mathrm{Var}(\varepsilon_t)$ for all $t \geq c$. Then,

$$\begin{aligned}\Big[\underbrace{\mathrm{Var}(w_{t+\ell}|c) - \mathrm{Var}(w_t|c)}_{\text{within-cohort}}\Big] - \Big[\underbrace{\mathrm{Var}(w_{t+\ell}|c+\ell) - \mathrm{Var}(w_t|c)}_{\text{within-experience}}\Big] &= \mathrm{Var}(w_{t+\ell}|c) - \mathrm{Var}(w_{t+\ell}|c+\ell) \\ &= \mu_{t+\ell}^2 \left[\mathrm{Var}(\theta_{t+\ell}|c) - \mathrm{Var}(\theta_{t+\ell}|c+\ell)\right],\end{aligned}$$

which equals zero if the variance of skills (in period $t+\ell$) does not differ across cohorts.[14] Notice that $\mathrm{Var}(\theta_{t+\ell}|c) = \mathrm{Var}(\theta_{t+\ell}|c+\ell)$ for all $\ell \geq 0$ implies that the variance of skills within each period is the same across all cohorts, but it does not say anything about the evolution of skill variation or returns over time. As discussed further in Appendix A.1, equal within-cohort and within-experience growth in the variance of residuals is consistent with growth in skill returns, the variance of skill growth, or the variance of non-skill wage shocks.[15]

2.1.2 Lemieux (2006)

Assuming that the variance of skills conditional on observed characteristics is time-invariant, $\mathrm{Var}(\theta_t|\boldsymbol{x}_t = \boldsymbol{x}) = \sigma^2(\boldsymbol{x})$, $\forall(t, \boldsymbol{x})$, Lemieux (2006) shows that the variance of skills increased over time due to compositional shifts in the labor market. To estimate changes in returns to unobserved skill over time, he implicitly ignores time variation in $\mathrm{Var}(\varepsilon_t)$ and re-weights $\mathrm{Var}(w_t|\boldsymbol{x}_t)$ each year to account for composition shifts in $\boldsymbol{x}_t$.[16]

Since Lemieux's assumption is not related to the dynamics of skill, it can be tested using repeated

[14] This would arise if, for example, growth in the variance of skills accumulated via labor market experience was exactly offset by growth in the variance of initial skills across cohorts.

[15] Empirically, Juhn, Murphy, and Pierce (1993) estimate very similar within-experience and within-cohort growth in *log wage* variation between 1964 and 1989; yet, their results suggest 20–30% stronger within-cohort (relative to within-experience) growth in *log wage residual* variation from 1970 to 1988. These results are, therefore, consistent with stronger growth in unobserved skill variation over the lifecycle than across cohorts during the 1970s and 1980s, but they say little about changes in the population-wide distribution of unobserved skill over this period.

[16] A separate analysis in Lemieux (2006) shows that measurement error in wages increased over time, at least in the widely used March Current Population Survey (CPS).

cross-section data with the same skill measurement over time. To see this, notice that equation (3) implies

$$\mathrm{Var}(T_t|\boldsymbol{x}_t) = \mathrm{E}[\tilde{T}_t^2|\boldsymbol{x}_t] = \tau^2\,\mathrm{Var}(\theta_t|\boldsymbol{x}_t) + \mathrm{Var}(\eta_t|\boldsymbol{x}_t).$$

Assuming the variance of test score measurement error, $\mathrm{Var}(\eta_t|\boldsymbol{x}_t)$, does not change over time, time-invariance of $\mathrm{Var}(\theta_t|\boldsymbol{x}_t)$ implies that $\mathrm{Var}(T_t|\boldsymbol{x}_t)$ should also be constant over time.

Using data on men with 30–50 years of experience in the 1996–2018 HRS, we test whether the variance of cognitive memory scores conditional on $\boldsymbol{x}_t$ has changed over time by regressing the squared residual of memory scores on indicators for race, educational attainment, experience, and year.[17] If $\mathrm{Var}(\theta_t|\boldsymbol{x}_t = \boldsymbol{x})$ is time-invariant, then the coefficients on year indicators should all be equal. Our results, shown in Appendix A.2, strongly reject time-invariance for highly experienced men in the HRS, indicating changes in within-group skill inequality over time.

2.1.3 Castex and Dechter (2014) and Deming (2017)

A few recent studies incorporate direct skill measurements in estimating the returns to (traditionally unobserved) skills over time (e.g., Castex and Dechter, 2014; Deming, 2017). Regressing log wages of workers in their late-20s on adolescent skill measures for different cohorts, these studies identify changes in the effects of adolescent skills on adult earnings (10–15 years later), confounded by any changes in measurement reliability. Even ignoring idiosyncratic errors in skill measurements, these estimates do not necessarily identify the evolution of returns to contemporaneous skills, μ_t, because they are confounded by any cross-cohort changes in the relationship between adolescent skills and adult skills.

In our context, OLS regression of log wage residuals in year $t+\ell$ on test residuals in year t identifies

$$\hat{\beta}_{t,t+\ell} \xrightarrow{p} \frac{\mathrm{Cov}(w_{t+\ell},\tilde{T}_t)}{\mathrm{Var}(\tilde{T}_t)} = \frac{\mu_{t+\ell}}{\tau}\underbrace{\left[1+\frac{\mathrm{Cov}(\theta_{t+\ell}-\theta_t,\theta_t)}{\mathrm{Var}(\theta_t)}\right]}_{\text{Skill Dynamics}}\underbrace{\left[\frac{\tau^2\,\mathrm{Var}(\theta_t)}{\tau^2\,\mathrm{Var}(\theta_t)+\mathrm{Var}(\eta_t)}\right]}_{\text{Test Reliability Ratio } (R_t)},$$

where $\ell \geq 0$ reflects the years between wage measurement and the time tests were administered. Under ideal conditions, this regression identifies returns to skill in any period $t+\ell$ up to the test score scale: $\mu_{t+\ell}/\tau$. Unfortunately, the two terms in brackets complicate identification of skill return growth.

Following a single cohort over time (i.e., varying ℓ for fixed t) will confound systematic heterogeneity

[17] The cognitive memory measure is a combination of immediate and delayed recall with raw scores ranging from 0 to 20. While the HRS contain other skill measures, they are either discrete (with very few values) or available for limited years. The memory recall measure we use has a correlation of roughly 0.3 with other cognitive tests focused on math skills, roughly 0.25 with log wages, and 0.07 with log wage residuals. See Appendix F for details on these measures and our HRS data.

in skill growth with changes in the returns to skill, as reflected in the "Skill Dynamics" term.[18] Instead of following the same cohort over time, Castex and Dechter (2014) compare estimates $\hat{\beta}_{t,t+\ell}$ and $\hat{\beta}_{t',t'+\ell}$ across the NLSY79 and NLSY97 cohorts where $t \approx 1980$ and $t' \approx 1997$, respectively. They use the same cognitive test measure, AFQT, in both periods with wages reported roughly 10–15 years after the tests were administered.[19] Comparing across cohorts, the "Skill Dynamics" term may differ due to cohort differences in the dynamics of skill accumulation between the year tests are taken (t and t') and the year wages are measured ($t + \ell$ and $t' + \ell$). If the variance of test score measurement error is time-invariant, then the "Test Reliability Ratio", R_t, will be the same across cohorts if and only if $\mathrm{Var}(\theta_t) = \mathrm{Var}(\theta_{t'})$, which then implies that the variance of test score residuals should also be the same across cohorts. For $\ell \geq 1$, $\hat{\beta}_{t',t'+\ell}/\hat{\beta}_{t,t+\ell}$ is unlikely to identify growth in unobserved skill returns, $\mu_{t'+\ell}/\mu_{t+\ell}$, if (i) the process for unobserved skill dynamics (over the first ℓ years after tests are measured) differs across cohorts or (ii) "initial" skill distributions differ across cohorts.[20]

Given modest changes in the distribution of AFQT scores across cohorts (Altonji, Bharadwaj, and Lange, 2012), the "Test Reliability Ratio" term is likely to be very similar across NLSY cohorts. By contrast, there are good reasons to think that skill dynamics during early-adulthood have changed. For example, Ashworth et al. (2021) document increases in work experience throughout high school and college, coupled with a rise in time to college degree for the NLSY97 cohort. Additionally, Appendix B documents substantial changes across NLSY cohorts in the types of occupational experience accumulated over ages 17–26. Most notably, experience accumulated in sales positions nearly tripled, while experience in manager and professional positions increased by 23% and 54%, respectively. Increases in management and professional experience were particularly strong at the high end of the AFQT distribution, while increases in sales and service experience were more uniform or concentrated at the low end.

Even ignoring these concerns, Castex and Dechter (2014) are only able to estimate changes in the returns to skill across two snapshots in time, from the late-1980s to around 2010. These estimates, as well as similar estimates for AFQT by Deming (2017), suggest that the returns to math and reading skills *fell* by roughly half over this 20-year period. Our estimated returns to skill presented below imply a similar drop, indicating that much of the decline occurred during the late-1980s and 1990s with relative stability in the 2000s.

[18] See Murnane, Willett, and Levy (1995) and Cawley, Heckman, and Vytlacil (2001) for efforts to sort out the rising importance of schooling vs. cognitive ability for earnings using the NLSY79.

[19] The NLSY79 (NLSY97) surveyed youth born 1957–1964 (1980–1984) administering a battery of tests to all respondents. The AFQT tests measure math and reading skills and were administered in 1980 (NLSY79) and 1997 (NLSY97) for most respondents in the two cohorts. In practice, the Armed Services Vocational Aptitude Battery (ASVAB) underlying the AFQT was taken via pencil and paper in the NLSY79, while it was administered in computer-adaptive form for the NLSY97.

[20] If wages are observed during the same years skills are measured for both cohorts (i.e., $\ell = 0$), then the "Skill Dynamics" term equals one, and consistent estimation of growth in skill returns, $\mu_{t'}/\mu_t$, depends only on time-invariance of the test reliability across cohorts.

This discussion has assumed that the same skill measurement is used for both cohorts; otherwise, cross-cohort comparisons of $\hat{\beta}_{t,t+\ell}$ and $\hat{\beta}_{t',t'+\ell}$ would also be confounded by differences in the mapping between skills and their measurement (i.e., differences in τ across tests). This is an additional challenge faced by Deming (2017), who aims to estimate changes in the return to social skills across NLSY cohorts. Unfortunately, he is forced to use different measures of social skills across cohorts, which means that his results cannot distinguish between differences in the "strength" of those measures vs. changes in skill returns, even if one is willing to assume that initial social skill distributions and their accumulation through high school, college, and early work-years remained the same across NLSY cohorts – a questionable assumption given the aforementioned cross-cohort increases in experience working in service, professional, and management occupations.[21]

2.2 Identification using Panel Data on Wages

Previous efforts to estimate the returns to unobserved skills rely on assumptions about the stability of skill distributions across cohorts. Using panel data, we introduce a very different approach based primarily on an assumption about lifecycle skill dynamics. 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 reflecting 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 negligible.[22] We begin with the following assumption.

Assumption 1. *For known $k \geq 1$ and for all $t - t' \geq k$: (i)* $\mathrm{Cov}(\Delta\theta_t, \theta_{t'}) = 0$*; (ii)* $\mathrm{Cov}(\Delta\theta_t, \varepsilon_{t'}) = 0$*; (iii)* $\mathrm{Cov}(\varepsilon_t, \theta_{t'}) = 0$*; and (iv)* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$*.*

Condition (i) assumes that skill growth is uncorrelated with sufficiently lagged skill levels. This allows for both fully permanent and transitory skill innovations. Condition (ii) allows for non-skill shocks to influence skill growth in the short-term but not in the long-term. For example, family illness or short-term work disruptions (including transitory firm-level productivity disruptions) may impact skill growth in the same year or even over the next $k - 1$ years. Condition (iii) is satisfied if skill levels are uncorrelated with non-skill shocks k or more years later, while condition (iv) requires that non-skill shocks have limited

[21] Deming (2017) normalizes his available measures of social skills to have a standard deviation of one in both cohorts before estimating $\hat{\beta}_{t,t+\ell}$ and $\hat{\beta}_{t',t'+\ell}$ (for social skills); however, this does not eliminate bias coming from differences in τ across measurements. See Appendix A.3 for details. Edin et al. (2022) take advantage of more consistent measures of cognitive and social/leadership skills across cohorts in Sweden, estimating modest reductions in returns to cognitive skills and increases in returns to social/leadership skills. While scores need not be re-normalized for each cohort, this analysis still relies on the assumption that early-career skill dynamics are identical across cohorts, as well as stability in measurement reliability ratios over time.

[22] Also see Carroll (1992) and Moffitt and Gottschalk (2011), who make similar assumptions ensuring that "long" autocovariances for log earnings residuals reflect a permanent component.

persistence (e.g., they may follow an MA(q) process where $q \leq k-1$). We discuss all of these conditions in greater detail below, empirically testing or relaxing those most central to identifying skill returns.

Our analysis assumes a sufficiently long panel with length satisfying $\bar{t} - \underline{t} \geq k+1$. Let Δ reflect the first-difference operator.

Proposition 1. *Assumption 1 implies that for all $t - t' \geq k+1$, the following instrumental variable (IV) estimator identifies skill return growth rates:*

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\Delta\mu_t}{\mu_{t-1}}. \tag{4}$$

For $\bar{t} - \underline{t} \geq k+1$ and normalizing $\mu_{t^*} = 1$ for some $t^* \geq \underline{t}+k$, all other $\mu_{\underline{t}+k}, \mu_{\underline{t}+k+1}, \ldots, \mu_{\bar{t}}$ are identified.

Proof: For all $t - t' \geq k$,

$$\begin{aligned}
\mathrm{Cov}(w_t, w_{t'}) &= \mathrm{Cov}(\mu_t\theta_t + \varepsilon_t, \mu_{t'}\theta_{t'} + \varepsilon_{t'}) \\
&= \mu_t \,\mathrm{Cov}(\theta_t, \mu_{t'}\theta_{t'} + \varepsilon_{t'}) + \mu_{t'}\,\mathrm{Cov}(\varepsilon_t, \theta_{t'}) + \mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}) \\
&= \mu_t \,\mathrm{Cov}(\theta_t, \mu_{t'}\theta_{t'} + \varepsilon_{t'}) && \text{[Assum 1(iii)–(iv)]} \\
&= \mu_t \,\mathrm{Cov}(\theta_{t'+k-1} + \Delta\theta_{t'+k} + \Delta\theta_{t'+k+1} + \ldots + \Delta\theta_t, \mu_{t'}\theta_{t'} + \varepsilon_{t'}) \\
&= \mu_t \underbrace{\left[\mu_{t'}\,\mathrm{Cov}(\theta_{t'+k-1}, \theta_{t'}) + \mathrm{Cov}(\theta_{t'+k-1}, \varepsilon_{t'})\right]}_{\equiv \Omega_{t'}} && \text{[Assum 1(i)–(ii)].}
\end{aligned} \tag{5}$$

Thus, for $t - t' \geq k+1$,

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\Delta\mu_t\Omega_{t'}}{\mu_{t-1}\Omega_{t'}} = \frac{\Delta\mu_t}{\mu_{t-1}}.$$

□

Proposition 1 shows that $\Delta\mu_t/\mu_{t-1}$ can be estimated by regressing $\Delta w_{i,t}$ on $w_{i,t-1}$ using sufficiently lagged $w_{i,t'}$ as an instrument. This IV approach is intuitive, since wage residuals can be thought of as 'noisy' measures of skill levels. To further this line of reasoning, follow the approach of Holtz-Eakin, Newey, and Rosen (1988), using $\theta_{i,t} = \theta_{i,t-1} + \Delta\theta_{i,t}$ and $\theta_{i,t-1} = (w_{i,t-1} - \varepsilon_{i,t-1})/\mu_{t-1}$ to obtain an expression for $\Delta w_{i,t}$ in terms of $w_{i,t-1}$:

$$\Delta w_{i,t} = \left[\mu_t\left(\frac{w_{i,t-1} - \varepsilon_{i,t-1}}{\mu_{t-1}} + \Delta\theta_{i,t}\right) + \varepsilon_{i,t}\right] - w_{i,t-1} = \frac{\Delta\mu_t}{\mu_{t-1}}w_{i,t-1} + \left(\varepsilon_{i,t} - \frac{\mu_t}{\mu_{t-1}}\varepsilon_{i,t-1} + \mu_t\Delta\theta_{i,t}\right). \tag{6}$$

This suggests that lagged residuals $w_{i,t-1}$, much like a test score, might serve as a proxy for unobserved skills. However, $w_{i,t-1} = \mu_{t-1}\theta_{i,t-1} + \varepsilon_{i,t-1}$ is a 'noisy' measure of unobserved skill, so it is correlated with the error $\varepsilon_{i,t-1}$, as well as $\varepsilon_{i,t}$ if $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1}) \neq 0$. Simply regressing $\Delta w_{i,t}$ on $w_{i,t-1}$ would, therefore, produce a biased estimate of $\Delta\mu_t/\mu_{t-1}$. To address this problem, lagged wage residuals from the distant past (i.e. any $w_{i,t'}$ for $t' \leq t-k-1$) can be used as instrumental variables in 2SLS estimation, since they are correlated with $w_{i,t-1}$ (through unobserved skills) but uncorrelated with $\varepsilon_{i,t-1}$, $\varepsilon_{i,t}$, and $\Delta\theta_{i,t}$ (under Assumption 1).

Future wage residuals are not valid instruments in equations (4) or (6), because skill growth has lasting effects on future skills, generating a correlation between future wage residuals and $\Delta\theta_{i,t}$. This correlation biases the IV estimator (for $\Delta\mu_t/\mu_{t-1}$) and makes it challenging to estimate skill returns during early sample periods. Appendix C.2 discusses conditions under which different cohorts may be used to eliminate the bias, enabling estimation of skill returns over the full sample period. Given the lengthy period covered by many panel data sets and stronger identification requirements for early skill returns, we focus on identification and estimation for periods $t \geq \underline{t} + k$.

The evolution of returns to skill are also directly related to predicted differences in wages across workers given any prior differences. Strengthening Assumption 1 to mean independence, $\mathrm{E}[\varepsilon_t|\theta_{t'}, \varepsilon_{t'}] = \mathrm{E}[\Delta\theta_t|\theta_{t'}, \varepsilon_{t'}] = 0$ for $t - t' \geq k$, implies that

$$\mathrm{E}[w_t|w_{t'}] = \mu_t \underbrace{\left(\frac{w_{t'} - \mathrm{E}[\varepsilon_{t'}|w_{t'}]}{\mu_{t'}} + \mathrm{E}[\theta_{t'+k} - \theta_{t'}|w_{t'}]\right)}_{\equiv \Psi_{t'}(w_{t'})}, \quad \text{for all } t \geq t' + k.$$

Because wages are increasing in skills and skills are persistent, workers with a high wage in one period will also tend to have a high wage in the future, even after the influence of transitory non-skill shocks has disappeared.[23]

More importantly for our purposes, for any given differences in year t' residuals across workers, long-term differences in expected future residuals, $\mathrm{E}[w_t|w_{t'}]$, will increase (decrease) over time as the returns to skill μ_t increase (decrease):

$$\mathrm{E}[w_t|w_{t'} = w^H] - \mathrm{E}[w_t|w_{t'} = w^L] = \mu_t\big(\Psi_{t'}(w^H) - \Psi_{t'}(w^L)\big), \quad \text{for all } t \geq t' + k. \tag{7}$$

Thus, the strong convergence in predicted future log wage residuals by prior residual quartiles over the late 1980s and 1990s shown in Figure 2 indicates a sharp decline in the returns to skill over those years.

We now make several additional observations on identification of skill returns.

Transitory skill shocks. Proposition 1 also applies if ε_{it} shocks are considered a component of skills, i.e., if $w_{i,t} = \mu_t(\theta_{i,t} + \varepsilon_{i,t})$.[24] Whether transitory shocks are assumed to be related or unrelated to skills

[23] This discussion assumes that $\Psi_{t'}(w_{t'})$ is an increasing function, as observed empirically.

[24] In this case, $\Omega_{t'} = \mu_{t'}[\mathrm{Cov}(\theta_{t'+k-1}, \theta_{t'}) + \mathrm{Cov}(\theta_{t'+k-1}, \varepsilon_{t'})]$.

has no effect on identification or IV estimation of the returns to skill under Assumption 1. Conceptually, it seems natural to think that transitory wage innovations have little to do with skills, so we continue with residuals as defined in equation (2).

Serially correlated non-skill shocks. Our identification strategy has, thus far, relied on the assumption that non-skill shocks, ε_t, become serially uncorrelated when observations are far enough apart. This is not critical; although, identification is most transparent in this case. Appendix C.4 shows identification of skill returns when the 'transitory' component ε_t contains an autoregressive component, such that the serial correlation in non-skill shocks depreciates over time but never fully disappears. Section 3.6 shows that estimates assuming ε_t contains an AR(1) component are quite similar to those obtained under our baseline Assumption 1.

Time-invariant skills. If skills are heterogeneous but time-invariant (i.e., $\theta_{i,t} = \theta_i$ with $\text{Cov}(\varepsilon_t, \theta) = 0$ for all t), then

$$\text{Cov}(w_t, w_{t'}) = \mu_t \mu_{t'} \text{Var}(\theta), \quad \text{for all } |t - t'| \geq k. \tag{8}$$

In this case, $\Delta\mu_t/\mu_{t-1}$ could be identified and estimated using the IV estimator in equation (4) with sufficiently *lagged* or *future* log wage residuals (i.e., $w_{t'}$ satisfying $t' \leq t - k - 1$ or $t' \geq t + k$) as instruments. For panel length satisfying $\bar{t} - \underline{t} \geq 2k$ and a single normalization (e.g., $\mu_{\underline{t}+k} = 1$), all $\mu_{\underline{t}}, ..., \mu_{\bar{t}}$ would be identified along with $\text{Var}(\theta)$. Comparing IV estimates using past vs. future wage residuals as instruments, our empirical analysis below provides strong evidence against fixed unobserved skills over the lifecycle.

Conditioning on observable subgroups. Assumption 1 can be modified so that all conditions (and results) hold for any observable subgroup, including specific cohort, age, or experience groups. For example, it is natural to condition on older (or more experienced) workers for whom endogenous human capital investments are likely to be negligible (Becker, 1964; Ben-Porath, 1967).[25] Our empirical analysis below pays particular attention to experienced workers, estimating returns to skill based on this subgroup.

[25]Appealing to Becker (1964) and Ben-Porath (1967), previous studies rely on the assumption of zero skill growth among older workers to identify the evolution of additively separable (log) skill prices (e.g. Heckman, Lochner, and Taber, 1998; Bowlus and Robinson, 2012) or the distribution of skill shocks (e.g. Huggett, Ventura, and Yaron, 2011). This assumption is stronger than needed in our context, where condition (i) of Assumption 1 only rules out persistent unobserved heterogeneity in skill growth among experienced workers. Heterogeneity in skill growth based on observable characteristics is accounted for through $f_t(\boldsymbol{x}_t)$ in obtaining residuals.

2.3 Testing our Assumption on Skill Dynamics

Even if endogenous skill investments become negligible as workers approach the end of their careers, skill growth rates may still be correlated with past skill levels for older workers due to other factors (e.g., heterogeneous skill depreciation). We explore this possibility using the same skill measure and sample of men with 30–50 years of experience in the 1996–2018 HRS data used earlier in Section 2.1.2.

We more formally test whether $\mathrm{Cov}(\Delta_2\theta_{t+2}, \theta_{t-\ell}) = 0$ using the following moments:

$$\mathrm{E}\left[(\Delta_2\tilde{T}_{t+2} - \varrho\tilde{T}_t)\tilde{T}_{t-\ell}\right] = 0, \quad \text{for } \ell \geq k, \tag{9}$$

where Δ_2 reflects the two-period time difference given our use of biennial data from the HRS. (These moments are consistent with 2SLS regression of $\Delta_2\tilde{T}_{t+2}$ on $\tilde{T}_t$ using $\tilde{T}_{t-\ell}$ as an instrument.) We test whether Assumption 1(i) holds for various k values. The first four columns of Table 1 test this assumption for $k = 2$ by testing whether $\varrho = 0$ when using instruments of lags $\ell \geq 2$. Columns 5 and 6 test the assumption for $k = 4$ and $k = 6$, respectively, using only longer lags as instruments. Panel A of Table 1 reports GMM estimates of ϱ using residualized memory recall scores. Although we reject $\varrho = 0$ at the 5% significance level when instruments of lags $\ell \leq 4$ are used, the estimated ϱ values are quite small. For perspective, if skills follow a simple autoregressive process (i.e., $\theta_t = \rho\theta_{t-1} + \nu_t$ with $\mathrm{Cov}(\theta_t, \nu_{t'}) = 0$ for all $t' \geq t + 1$), then $\varrho = \rho^2 - 1$. The reported estimates in columns 1–5 would all imply ρ values of 0.97–1.02, very close to a random walk. The last column of Table 1 reports an estimated ϱ of 0.018 when using lags $\ell = 6, 8$. This estimate is not significantly different from zero and suggests that Assumption 1(i) is satisfied for $k = 6$. That is, skill growth rates $\Delta\theta_t$ are uncorrelated with skill levels (at least) 6 years earlier, θ_{t-6}.

Panel B of Table 1 reports estimates of ϱ when also including lagged log wage residuals, $w_{t-\ell}$, as additional instruments. In this case, $\varrho = 0$ implies that both conditions (i) and (ii) of Assumption 1 are satisfied for the relevant k. These estimates are nearly identical to those using only lagged memory test score residuals as instruments in Panel A, indicating that condition (ii) is likely to be satisfied. Altogether, the estimates reported in Table 1 suggest that, for older men at least, conditions (i) and (ii) of Assumption 1 are satisfied for $k = 6$, while violations of those conditions are quite modest for k as small as 2. We conduct most of our analysis assuming that these conditions are satisfied for $k = 6$; however, we reach very similar conclusions when relaxing condition (i) in Section 3.5.

Table 1: GMM estimates of ϱ in equation (9) using $(\tilde{T}_{i,t-\ell}, w_{i,t-\ell})$ as instruments

	$\ell = 2$	$\ell = 2,4$	$\ell = 2,4,6$	$\ell = 2,4,6,8$	$\ell = 4,6,8$	$\ell = 6,8$
A. Instruments: $\tilde{T}_{i,t-\ell}$						
Estimated ϱ	0.045*	-0.040*	-0.029*	-0.031*	-0.057*	0.018
	(0.020)	(0.012)	(0.011)	(0.010)	(0.014)	(0.022)
Implied $\rho = \sqrt{1+\varrho}$	1.022	0.980	0.985	0.984	0.971	1.009
B. Instruments: $\tilde{T}_{i,t-\ell}, w_{i,t-\ell}$						
Estimated ϱ	0.044*	-0.040*	-0.030*	-0.031*	-0.059*	0.018
	(0.019)	(0.011)	(0.011)	(0.010)	(0.013)	(0.022)
Implied $\rho = \sqrt{1+\varrho}$	1.022	0.980	0.985	0.984	0.970	1.009

Notes: $\tilde{T}_t$ are residuals from regressions of word recall on experience, cohort, race, and education dummies. w_t are residuals from year-specific regressions on the same covariates. Uses 1996–2018 HRS data for men ages 50–70 with 30–50 years of experience. Estimated via two-step optimal GMM with cluster-robust weighting matrix. * denotes significance at 0.05 level.

3 New Evidence on Returns to Unobserved Skill

Our primary objective is estimation of returns to unobserved skill over time. We mainly exploit data from the PSID; however, we replicate key results using administrative data on earnings in Section 6. This section briefly describes the PSID before turning to estimated returns to skill based on these data.

3.1 Panel Study of Income Dynamics (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.

Our sample is based on male heads of households from the core (SRC) sample restricted to years these men were ages 16–64, had potential experience of 1–40 years, reported positive wage and salary income, had positive hours worked, and were not enrolled as students. Our sample is 92% white with an average age of 39 years. Roughly half of our sample completed more than 12 years of schooling, which we refer to as "college workers". The wage measure used in our analysis divides annual earnings by annual hours worked, trimming the top and bottom 1% of all wages within year and college/non-college status by ten-year experience cells. The resulting sample contains 3,766 men and 44,547 person-year observations – roughly 12 observations for each individual. See Appendix E for further details.

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/ethnicity, and 7 educational attainment categories, along with interactions between a cubic in experience and both race and education indicators.

3.2 Underlying Trends

Log wage inequality has increased substantially since 1970, with particularly strong increases in the early-1980s and after 2000 (see Figure 1). The evolution of residual inequality closely mirrors this pattern, explaining a larger share of the total variance than the between-group variance.

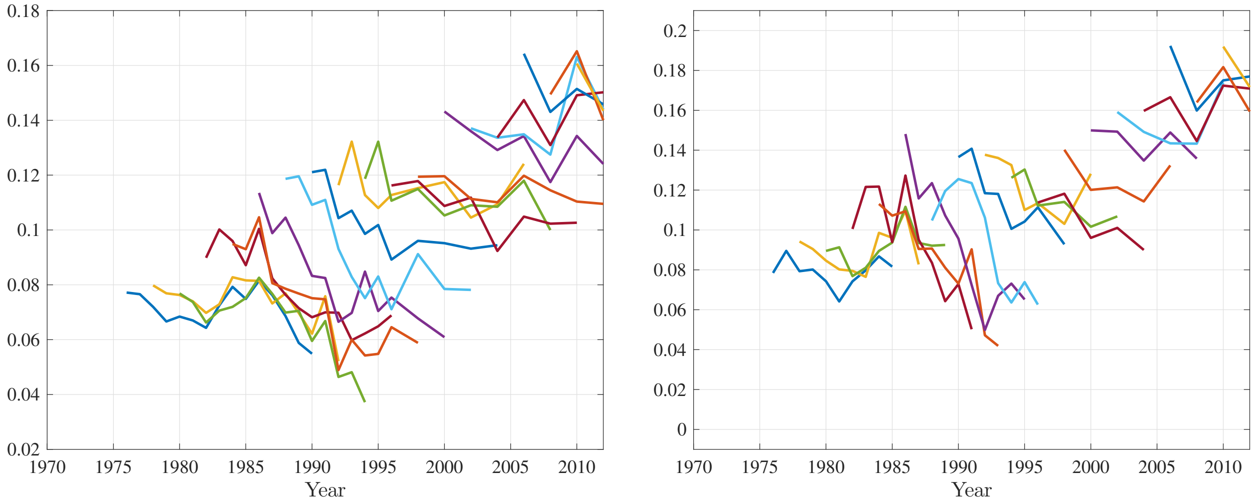
Consistent with an important role for unobserved skills, Figure 2 shows that those with higher wage residuals in any given year also have higher wage residuals, on average, up to 20 years later.[26] The sharp convergence in log wage residuals (across quartiles of earlier residual levels) over the late-1980s and 1990s indicates that the returns to skill fell over that period (see equation (7)), despite modest growth in residual inequality at the time.

Section 2.2 suggests that long autocovariances in wage residuals offer a more direct way to identify changes in the returns to unobserved skill. Figure 3(a) reports $\mathrm{Cov}(w_b, w_t)$ for $t = b + 6, \ldots, b + 20$ with each line reporting autocovariances for a different 'base' year b and 15 subsequent years.[27] For example, the leftmost line beginning in 1976 reflects autocovariances for $b = 1970$ and values of t ranging from 1976–1990. If systematic differences in unobserved skill growth are negligible and $t - b$ is large enough such that transitory shocks are uncorrelated, then $\mathrm{Cov}(w_b, w_t) = \mu_t \Omega_b$ (see equation (5)) and following each line over t is directly informative about the evolution of μ_t. (We discuss the shifts up or down across lines below.) 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 or even increasing returns.

If there are persistent differences in unobserved skill growth, then the residual covariances are more difficult to interpret, since $\mathrm{Cov}(\theta_b, \Delta\theta_t)$ will not generally equal zero as Assumption 1(i) requires. In this case, it is useful to focus on more experienced workers for whom differences in unobserved skill growth should be negligible (and idiosyncratic) due to diminished investment incentives (Becker, 1964; Ben-Porath, 1967). Figure 3(b) reveals very similar autocovariance patterns to Figure 3(a) when restricting the sample to men with 15–30 years of experience as of baseline b years (21+ years of experience in years

[26]For comparison, Appendix Figure E-1 shows average log wage residuals for each quartile over time without conditioning on prior residual levels.

[27]Appendix Figure E-2 shows that sample attrition due to non-response or aging/retirement does not affect the autocovariance patterns documented here.



(a) All Men

(b) Men with 15–30 Years of Experience in Year b

Figure 3: Autocovariances for Log Wage Residuals

$t \geq b + 6$).[28]

As emphasized by 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), wage inequality has evolved differently for non-college and college-educated workers. Figure 4 shows that the rise in residual inequality over the early-1980s was stronger among non-college workers, before falling and then quickly stabilizing in the mid-1980s, while it continued to increase among college workers throughout our sample period. It is natural to ask whether these different trends reflect differences in the evolution of returns to skill by educational attainment, as is often implicitly assumed.

Figure 5 reports long autocovariances separately for non-college and college educated men. The time patterns are qualitatively similar for both education groups with two noteworthy differences: First, the autocovariance lines continue declining for non-college men throughout the early-2000s when they flatten out for college men. This suggests that the returns to skill continued falling for non-college men several years after they stabilized for college men. Second, the lines generally shift upward over time, with particularly strong increases over the late-1990s and early-2000s for college men. As discussed in Section 4, these shifts reflect rising skill dispersion, partially muted by declining skill returns.

3.3 2SLS Estimation of Skill Returns

In this subsection, we directly estimate growth rates in the returns to unobserved skill based on the IV strategy described in Section 2.2. Because our data is only available every other year later in the sample

[28] Appendix Figure E-3 shows qualitatively similar, though flatter, autocovariance patterns for less-experienced men.

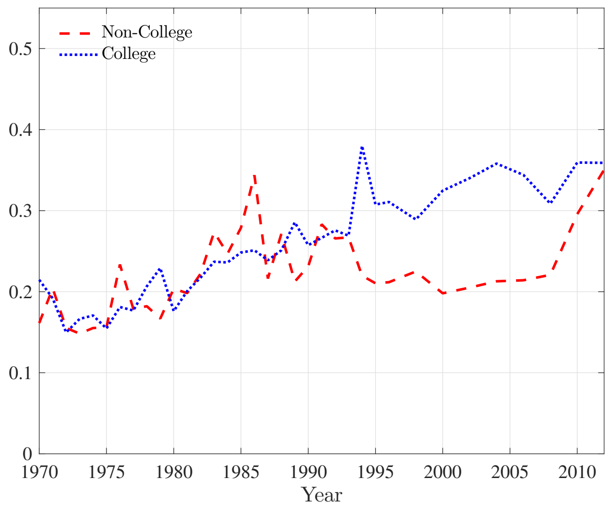


Figure 4: Variance of Log Wage Residuals by Education

period, we slightly modify the 2SLS approach based on equation (6) to estimate two-year growth rates, $\Delta_2\mu_t/\mu_{t-2}$, based on the following:

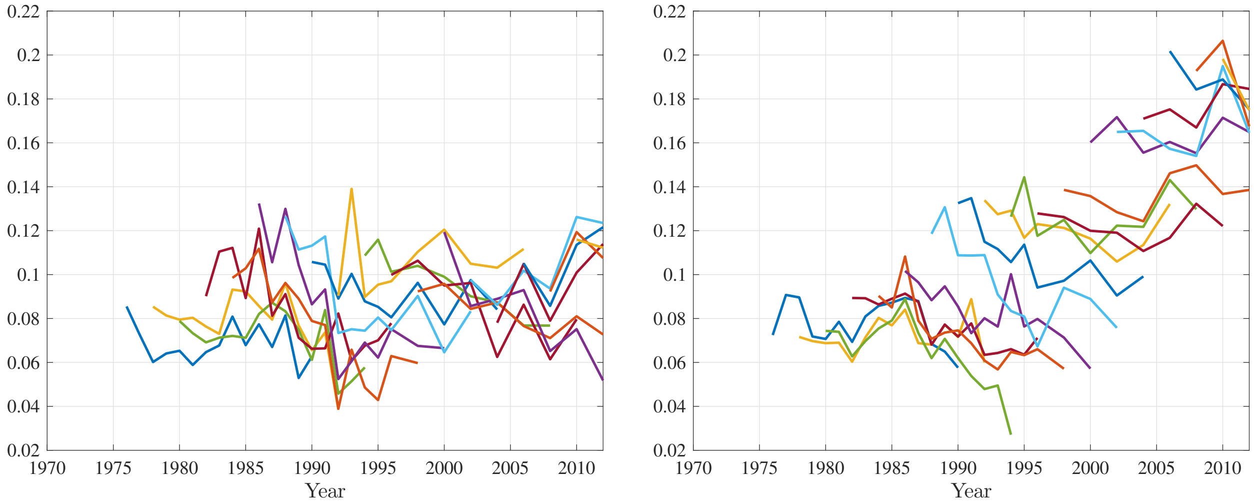
$$\Delta_2 w_{i,t} = \left(\frac{\Delta_2\mu_t}{\mu_{t-2}}\right) w_{i,t-2} + \left[\mu_t(\Delta\theta_{i,t-1} + \Delta\theta_{i,t}) + \varepsilon_{i,t} - \frac{\mu_t}{\mu_{t-2}}\varepsilon_{i,t-2}\right]. \tag{10}$$

Under Assumption 1, we can obtain consistent estimates of $\Delta_2\mu_t/\mu_{t-2}$ by estimating equation (10) via 2SLS using lags $w_{i,t'}$ for $t' \leq t - k - 2$ as instrumental variables.

Table 2 reports 2SLS estimates of skill return growth rates using equation (10) 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). Assuming $k = 6$, we use $(w_{i,t-8}, w_{i,t-9})$ as instruments. Table 3 reports 2SLS estimates for the later years of the PSID (t covering 1996–2012) when observations become biennial.[29] In all specifications, the instruments are 'strong' with very large first-stage F-statistics.

Panel A of Tables 2 and 3 reports estimates for the full sample of men in the PSID, while panels B and C report separate estimates for non-college and college men. Consistent with the autocovariances reported earlier, nearly all of these estimates are negative, with several statistically significant. Appendix Tables E-1 and E-2 report analogous results for the subsample of men with 21–40 years of experience (in year t) for whom we expect systematic heterogeneity in skill growth to be negligible. Figure 6 combines these estimates to trace out the implied paths for μ_t from 1979–2012, normalizing $\mu_{1985} = 1$. Altogether, these results suggest that the returns to unobserved skill *declined* by roughly half since the mid-1980s, mirroring the substantial decline in returns to cognitive skills (between the NLSY79 and NLSY97 cohorts)

[29] Estimates in Table 3 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.



(a) Non-College (b) College

Figure 5: Autocovariances for Log Wage Residuals by Education, All Experience levels

estimated by Castex and Dechter (2014). In Section 5, we consider the interpretation of these estimated return series when there are multiple unobserved skills whose returns may evolve differently over time. We also estimate similar return patterns for men working in different occupation types.

Appendix E.3 shows that analogous GMM estimates to those in Tables 2 and 3 are very similar.[30] More importantly, we test the validity of our lagged instruments (using Hansen J-tests), since the model is overidentified when using multiple instruments. We cannot reject exogeneity of our instruments at conventional levels in any year, suggesting that Assumption 1 cannot be rejected (for $k = 6$). By contrast, we show that future residuals are invalid instruments (during most time periods), highlighting the importance of accounting for idiosyncratic variation in lifecycle skill growth.

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. Rather than report several sets of 2SLS estimates with different instrument sets, we next employ minimum distance (MD) estimation to take advantage of all long autocovariances available in the data.

[30]The GMM estimates exploit the same moments but use the optimal weighting matrix (allowing for heteroskedasticity and serial correlation within individuals). As with those reported in Tables 2 and 3, these estimates require no assumptions about the variance of individual skill innovations $\Delta\theta_{i,t}$ (or non-skill shocks, $\varepsilon_{i,t}$) over time, across cohorts, or across experience groups.

Table 2: 2SLS estimates of $\Delta_2\mu_t/\mu_{t-2}$ for two- or three-year periods, 1979–1995

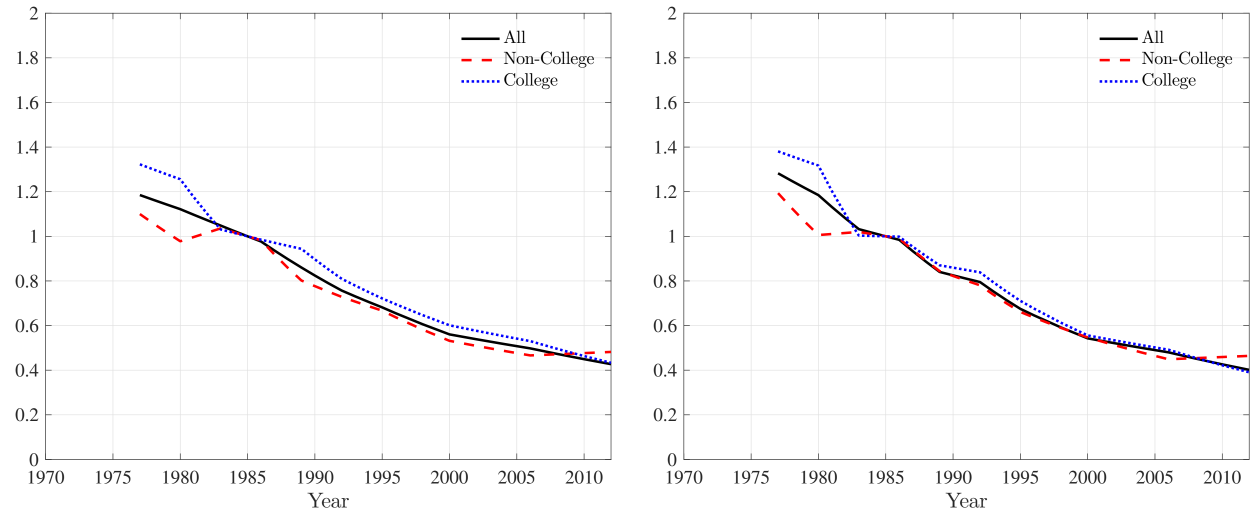
	1979–1980	1981–1983	1984–1986	1987–1989	1990–1992	1993–1995
			A. All men			
$\Delta_2\mu_t/\mu_{t-2}$	-0.036	-0.044	-0.046	-0.081*	-0.082*	-0.067
	(0.045)	(0.038)	(0.038)	(0.034)	(0.035)	(0.035)
Observations	1,349	2,077	2,188	2,245	2,189	2,095
1st stage F-Statistic	163.09	191.61	114.85	209.42	227.13	286.96
			B. Non-college men			
$\Delta_2\mu_t/\mu_{t-2}$	-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
			C. College men			
$\Delta_2\mu_t/\mu_{t-2}$	-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 $\Delta_2 w_{i,t}$ on $w_{i,t-2}$ using instruments $(w_{i,t-8}, w_{i,t-9})$.
* denotes significance at 0.05 level.

Table 3: 2SLS estimates of $\Delta_2\mu_t/\mu_{t-2}$ for five-year periods, 1996–2012

	1996–2000	2002–2006	2008–2012
	A. All men		
$\Delta_2\mu_t/\mu_{t-2}$	-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. Non-college men		
$\Delta_2\mu_t/\mu_{t-2}$	-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
	C. College men		
$\Delta_2\mu_t/\mu_{t-2}$	-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 $\Delta_2 w_{i,t}$ 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.



(a) All Experience Levels

(b) 21–40 Years of Experience in Year t

Figure 6: μ_t Implied by 2SLS Estimates ($\mu_{1985} = 1$)

3.4 Minimum Distance Estimation of Skill Returns using Long Autocovariances

We now explicitly incorporate cohorts, c, into our analysis. Assuming all conditions in Assumption 1 hold for each cohort, μ_t and $\Omega_{c,t'}$ are identified from long autocovariances as shown in equation (5). Separately for non-college and college men, we estimate μ_t and $\Omega_{C,t'}$ for all (t, t') satisfying $t - t' \geq 6$ for men with 21–40 years of experience in year t. Due to small sample sizes of single-year cohorts, we consider 4 broad cohort groups denoted by C, where each cohort group consists of 10-year labor market entry cohorts.[31] Table 4 describes these cohort groups, parameters estimated, and autocovariances used in estimation. Altogether, we exploit 157 covariances and use equally weighted MD to estimate 63 parameters (normalizing $\mu_{1985} = 1$) separately for non-college and college men. See Appendix D for further details on our MD estimation.

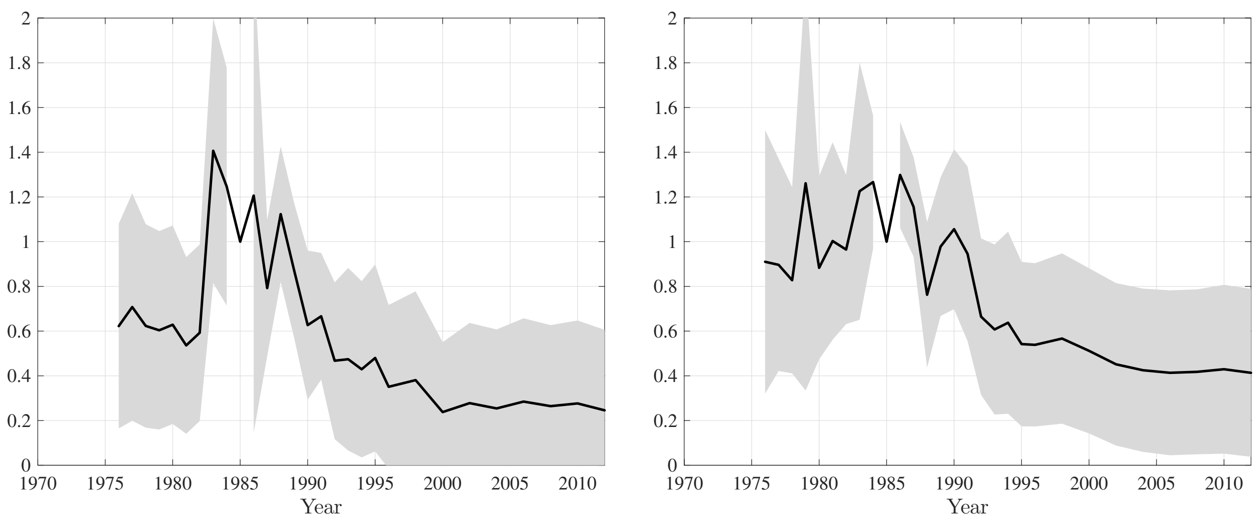
Figure 7 reports MD estimates of μ_t, while Figure 8 reports estimated $\Omega_{C,t'}$, both separately for non-college and college men. (Shaded areas in these figures reflect 95% confidence intervals.) Like their 2SLS counterparts, MD estimates of μ_t indicate substantial declines (roughly 50–70%) in the returns to skill over the late-1980s and 1990s. This contrasts sharply with the estimated rise in returns during the late-1970s and early-1980s. While the additional autocovariances used in MD estimation (compared to 2SLS) improve precision, confidence intervals in Figure 7 still admit the possibility that skill returns were

[31] We estimate $\Omega_{C,t'} \equiv \mu_{t'} \operatorname{Cov}(\theta_{t'+k-1}, \theta_{t'} | c \in C) + \operatorname{Cov}(\theta_{t'+k-1}, \varepsilon_{t'} | c \in C)$ and make no effort to separately identify $\Omega_{c,t'}$ for each annual entry cohort. Given Assumption 1, this does not impose any assumptions on variation in $\Omega_{c,t'}$ across annual cohorts even for cohorts c within broader cohort groups C. Requiring that all single-year cohorts in each cohort group have 21–40 years of experience in each year of t, we exclude older (1936–1941) and younger (1982–1991) cohorts due to limited variation in t.

Table 4: Cohort grouping

Cohort Group C	Range						Number		
	Cohort c		Year t'		Year t		$\Omega_{C,t'}$	μ_t	$\text{Cov}(w_t, w_{t'} \mid C)$
1	1942	1951	1970	1976	1976	1982	7	7	28
2	1952	1961	1976	1986	1982	1992	11	11	66
3	1962	1971	1986	1996	1992	2002	11	8	42
4	1972	1981	1996	2006	2002	2012	6	6	21
Total							35	29	157

Notes: Since μ_t is not cohort-specific, the total number of μ_t parameters does not equal the sum for each cohort due to overlap in years across cohorts.



(a) Non-College (b) College

Figure 7: μ_t implied by MD estimates using long autocovariances, 21–40 years of experience

relatively stable prior to 1985. They also suggest that returns fell by at least 40% for non-college men and 20% for college men. Estimated $\Omega_{C,t'}$ profiles in Figure 8 show a strong upward trend beginning in the 1980s. Under (mild) additional assumptions, we show in Section 4 that the $\Omega_{C,t'}$ trends indicate substantial growth in the variance of unobserved skills over time for the two most recent cohort groups.

3.5 Relaxing our Assumption on Skill Growth

An important assumption, thus far, is that skill growth is uncorrelated with sufficiently lagged skill levels. In this subsection, we consider two alternative specifications for skill dynamics that violate condition (i) of Assumption 1. To simplify the discussion, it is useful to slightly strengthen conditions (ii)–(iii) of

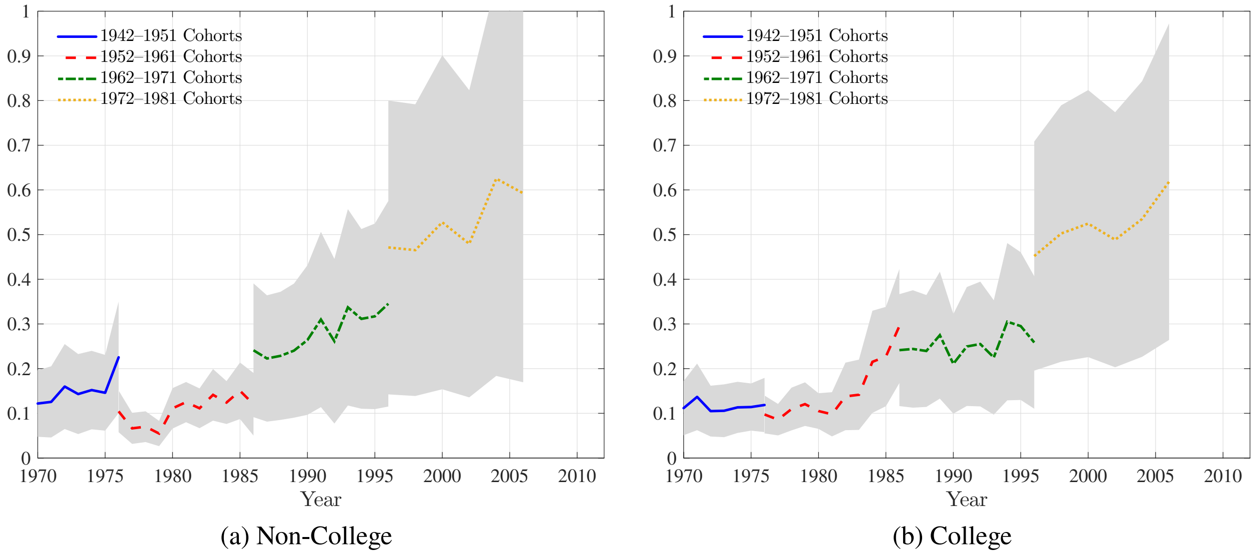


Figure 8: $\Omega_{C,t'}$ implied by MD estimates using long autocovariances, 21–40 years of experience

Assumption 1 to $\mathrm{Cov}(\theta_t, \varepsilon_{t'}) = 0$ for all t, t', while maintaining condition (iv) (i.e., limited persistence of non-skill shocks). In this case, our IV estimator converges to

$$\gamma_{t,t'} \equiv \frac{\mathrm{Cov}(\Delta w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\Delta \mu_t}{\mu_{t-1}} + \frac{\mu_t}{\mu_{t-1}} \frac{\mathrm{Cov}(\Delta \theta_t, \theta_{t'})}{\mathrm{Cov}(\theta_{t-1}, \theta_{t'})}, \quad \text{for } t' - t \geq k \text{ or } t - t' \geq k+1, \tag{11}$$

where $\mathrm{Cov}(\Delta\theta_t, \theta_{t'}) \neq 0$ would bias these estimates of skill return growth.

3.5.1 Heterogeneity in Lifecycle Skill Growth

We begin by exploring the possibility that unobserved skill growth innovations are correlated over time as in the heterogeneous income profile (HIP) models estimated in, e.g., Haider (2001), Baker and Solon (2003), Guvenen (2009), and Moffitt and Gottschalk (2012). We consider a more flexible process governing this skill growth heterogeneity, assuming

$$\Delta\theta_{i,t} = \lambda_t(c_i)\delta_i + \nu_{i,t}, \tag{12}$$

where δ_i is a mean zero individual-specific lifecycle growth rate factor, and the $\lambda_t(c) \geq 0$ terms allow for variation in systematic skill growth across time and cohorts/experience.[32] This skill process generally

[32] We assume $\mathrm{Var}(\delta|c) > 0$, allowing for the possibility that $\lambda_t(c) = 0$ for all t and c in the absence of heterogeneity in skill growth. Letting ψ_i reflect the initial skill for an individual entering the labor market, the level of unobserved skill for

violates condition (i) of Assumption 1 when $\lambda_t(c) > 0$, where the bias for IV estimator $\gamma_{t,t'}$ depends on

$$\mathrm{Cov}(\Delta\theta_t, \theta_{t'}) = \mathrm{E}\left[\mathrm{Cov}(\Delta\theta_t, \theta_{t'}|c)\right] = \mathrm{E}\left[\lambda_t(c)\,\mathrm{Cov}(\delta, \theta_{t'}|c) + \mathbb{1}(t' > t)\,\mathrm{Var}(\nu_t|c)\right],$$

the expectation is taken over cohorts, c, and $\mathbb{1}(\cdot)$ is the indicator function. This shows that $\gamma_{t,t'}$ estimates will be biased downward only when workers with higher skill growth rates, δ, have lower skill levels, $\theta_{t'}$. This is only likely to be a concern for young workers for whom initial skills may be negatively correlated with incentives to acquire new skills. Hence, our focus on experienced workers makes it unlikely that the estimated declines in skill returns over the late-1980s and 1990s are explained by systematic lifecycle skill growth heterogeneity.

This model also offers testable predictions related to the year, t', from which we take $w_{t'}$ as an instrument. In the absence of HIP (i.e., $\lambda_t = 0$ for all t), IV estimates should not vary with the year of lagged residuals (satisfying $t' \leq t - k - 1$) used as instruments nor with the year of future residuals (satisfying $t' \geq t+k$) used as instruments; although, estimates will be greater when using any future (rather than any past) residuals as instruments if $\mathrm{Var}(\nu_t) > 0$. By contrast, HIP (i.e., $\lambda_t > 0$ for all t) implies that IV estimates will generally vary with the year of lagged or future residuals used as instruments.

Tables 5 and 6 report GMM estimates of $\gamma_{t,t'}$ (using two-year differences) for all non-college and college men, respectively, using moments $\mathrm{E}[(\Delta_2 w_t - \gamma_{t,t'} w_{t-2})w_{t'}] = 0$ with different residual leads and lags, $w_{t'}$, as instruments. We highlight two patterns. First, we cannot reject equality of estimates when using only lags of $t-8$ and $t-12$ as instruments (specification 1) or when using only leads of $t+6$ and $t+10$ (specification 2) as instruments. Second, we reject equality of estimates when using lags ($t-8$) and leads ($t+6$) together as instruments (specification 3). Together, these results provide no indication of systematic heterogeneity in unobserved skill growth. Absent this heterogeneity, the larger $\gamma_{t,t'}$ estimates obtained when using leads as instruments imply an important role for idiosyncratic skill growth innovations (i.e., $\mathrm{Var}(\nu_t) > 0$).[33] Finally, we note that GMM estimates using only sufficiently lagged residuals as instruments (specification 1 of Tables 5 and 6) imply μ_t profiles that are very similar to those shown in Figure 6.

As discussed earlier, human capital theory (Becker, 1964; Ben-Porath, 1967) predicts that optimal skill

individual i from cohort c_i in year t can be written as

$$\theta_{i,t} = \psi_i + \Lambda_t(c_i)\delta_i + \sum_{j=0}^{t-c_i-1} \nu_{i,t-j},$$

where $\Lambda_t(c) \equiv \sum_{j=0}^{t-c-1} \lambda_{t-j}(c)$ reflects the accumulated influence of skill growth heterogeneity.

[33]Exogeneity tests reported in Section 3.3 and Appendix E.3.2 also suggest that (i) estimated growth in returns does not vary significantly with the year of (sufficiently) lagged wages and (ii) estimated skill return growth is significantly stronger when using future rather than lagged residuals as instruments.

Table 5: Multiple-Equation GMM Estimates: Non-College

	Instrument for Each Equation of $\Delta_2 w_{i,t}$: (1)		(2)		(3)	
	$w_{i,t-8}$	$w_{i,t-12}$	$w_{i,t+6}$	$w_{i,t+10}$	$w_{i,t-8}$	$w_{i,t+6}$
Coefficient on $w_{i,t-2}$ for years						
1972–1974			0.054	-0.013		
			(0.057)	(0.075)		
1975–1977			0.132	0.075		
			(0.095)	(0.078)		
1978–1980			0.081	0.107	-0.085	0.170
			(0.096)	(0.088)	(0.054)	(0.125)
1981–1983	0.017	-0.057	0.267	0.300*	0.084	0.143
	(0.082)	(0.096)	(0.137)	(0.126)	(0.082)	(0.085)
1984–1986	-0.074	0.001	0.114	0.137	0.032	0.089
	(0.059)	(0.065)	(0.103)	(0.093)	(0.109)	(0.092)
1987–1989	-0.199*	-0.161	0.050	0.026	-0.185	0.010
	(0.086)	(0.135)	(0.090)	(0.100)	(0.117)	(0.071)
1990–1992	-0.069	-0.096	0.029	-0.075	-0.151*	-0.045
	(0.059)	(0.080)	(0.084)	(0.079)	(0.075)	(0.078)
1993–1995	-0.076	-0.125	0.139	0.084	-0.057	0.047
	(0.063)	(0.085)	(0.086)	(0.128)	(0.062)	(0.089)
1996–2000	-0.079	-0.053	0.091	0.022	-0.056	0.052
	(0.051)	(0.056)	(0.062)	(0.072)	(0.048)	(0.051)
2002–2006	-0.043	-0.038	0.089	-0.052	-0.022	0.054
	(0.058)	(0.052)	(0.135)	(0.171)	(0.065)	(0.066)
2008–2012	-0.049	-0.038				
	(0.076)	(0.079)				
Observations	5,627		6,883		5,093	
Wald p-value	0.945		0.756		0.044	

Notes: * denotes significance at 0.05 level.

Table 6: Multiple-Equation GMM Estimates: College

	Instrument for Each Equation of $\Delta_2 w_{i,t}$:					
	(1)		(2)		(3)	
	$w_{i,t-8}$	$w_{i,t-12}$	$w_{i,t+6}$	$w_{i,t+10}$	$w_{i,t-8}$	$w_{i,t+6}$
Coefficient on $w_{i,t-2}$ for years						
1972–1974			0.068	0.030		
			(0.076)	(0.075)		
1975–1977			0.225*	0.065		
			(0.091)	(0.070)		
1978–1980			0.036	0.078	0.016	-0.004
			(0.077)	(0.075)	(0.083)	(0.069)
1981–1983	-0.125	0.002	0.156	0.195	-0.128*	0.120
	(0.080)	(0.096)	(0.088)	(0.117)	(0.061)	(0.086)
1984–1986	0.032	0.158	0.240*	0.422*	0.018	0.174*
	(0.066)	(0.084)	(0.071)	(0.112)	(0.079)	(0.074)
1987–1989	-0.004	-0.069	0.107*	0.106	-0.015	0.023
	(0.058)	(0.056)	(0.046)	(0.056)	(0.066)	(0.067)
1990–1992	-0.033	-0.119*	0.095	0.006	-0.095	0.069
	(0.056)	(0.058)	(0.061)	(0.069)	(0.054)	(0.055)
1993–1995	-0.030	-0.116	0.085	0.070	-0.071	0.069
	(0.053)	(0.073)	(0.051)	(0.060)	(0.048)	(0.051)
1996–2000	-0.080*	-0.044	0.109*	0.087*	-0.037	0.094*
	(0.035)	(0.049)	(0.046)	(0.040)	(0.034)	(0.039)
2002–2006	-0.016	0.030	0.155	0.165	0.024	0.048
	(0.034)	(0.042)	(0.084)	(0.097)	(0.037)	(0.040)
2008–2012	-0.069*	-0.030				
	(0.031)	(0.036)				
Observations	7,353		9,263		7,069	
Wald p-value	0.080		0.354		0.007	

Notes: * denotes significance at 0.05 level.

investment and accumulation become negligible as workers approach the end of their careers. Assuming no systematic *unobserved* heterogeneity in skill growth among experienced workers (i.e. $\lambda_t(c) = 0$ for all workers with at least 21 years of experience) and $\mathrm{Cov}(\theta_t, \varepsilon_{t'}) = 0$ for all t, t', baseline μ_t estimates (using only experienced workers) reported in Figure 7 can be used to scale log wage residuals to estimate

$$\mathrm{Cov}\left(\Delta\left(\frac{w_t}{\mu_t}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)\Big|c\right) = \lambda_t(c)\lambda_{t'}(c)\,\mathrm{Var}(\delta|c), \quad \text{for } t - t' \geq k+1, \tag{13}$$

for less-experienced workers. Systematic heterogeneity in skill growth at younger ages should be reflected in systematically positive covariances in scaled-residual growth. Yet, Figure 9 shows that for cohort groups $C \in \{3, 4\}$, the covariances in equation (13) fluctuate around zero for all ages. Figure 10 further shows that the distribution of all covariances (for workers with 1–20 years of experience in t) is centered around zero. These covariances strongly suggest that systematic skill growth heterogeneity is negligible early in the lifecycle for these cohorts. Related results for $\mathrm{Cov}(\Delta(w_t/\mu_t), w_{t'}|c) = \lambda_t(c)\mu_{t'}\,\mathrm{Cov}(\delta, \theta_{t'}|c)$ reported in Appendix E.4 further support this conclusion.

Altogether, Tables 5 and 6 and Figures 9 and 10 support condition (i) of Assumption 1: skill growth is uncorrelated with lagged skill levels throughout the careers of men in our sample.

3.5.2 AR(1) skill dynamics

We next consider an alternative model of skill dynamics characterized by a fixed effect, ψ_i, and an AR(1) component, $\phi_{i,t}$:

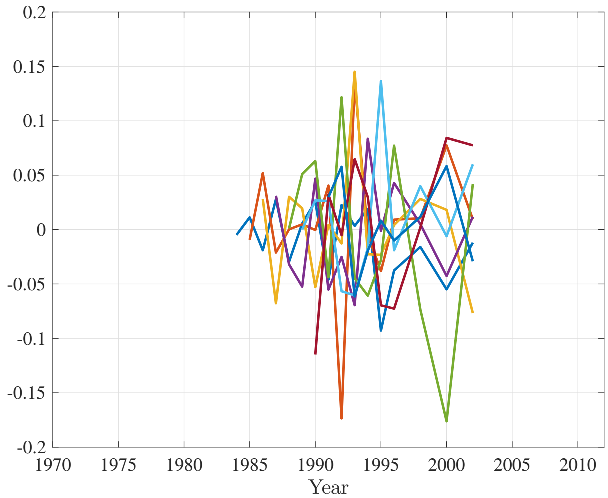
$$\theta_{i,t} = \psi_i + \phi_{i,t}, \qquad \text{where} \quad \phi_{i,t} = \rho_t \phi_{i,t-1} + \nu_{i,t}. \tag{14}$$

For $\rho_t < 1$, this specification is consistent with heterogeneous depreciation of skills acquired in the labor market generating mean-reversion to individual-specific baseline skill levels determined by ψ_i. Our baseline specification implicitly assumes $\rho_t = 1$ for all t.

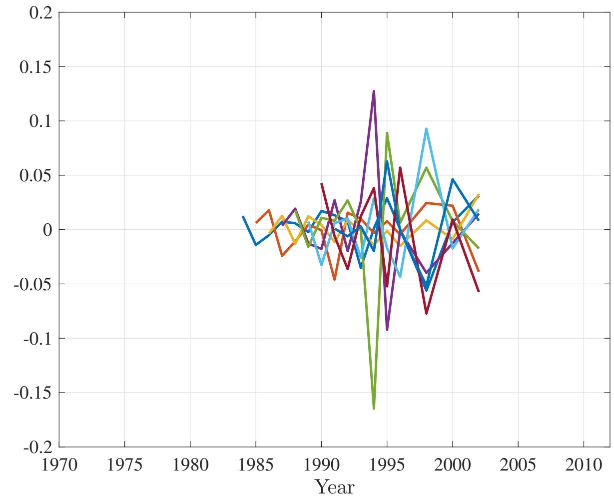
When $\rho_t \neq 1$, the AR(1) skill component violates Assumption 1(i), since skill growth will be correlated with all past skill levels. With this more general specification for skills, we assume that all skill components are uncorrelated with non-skill shocks, while maintaining our assumption of limited persistence in non-skill shocks.

Assumption 2. *For all cohorts, c: (i)* $\mathrm{Cov}(\psi, \phi_t|c) = 0$ *for all t; (ii)* $\mathrm{Cov}(\psi, \varepsilon_{t'}|c) = \mathrm{Cov}(\phi_t, \varepsilon_{t'}|c) = 0$ *for all* t, t'*; (iii)* $\mathrm{Cov}(\phi_{t'}, \nu_t, |c) = \mathrm{Cov}(\nu_{t'}, \nu_t, |c) = 0$ *for all* $t - t' \geq 1$*; (iv) for known* $k \geq 1$, $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = 0$ *for all* $t - t' \geq k$.

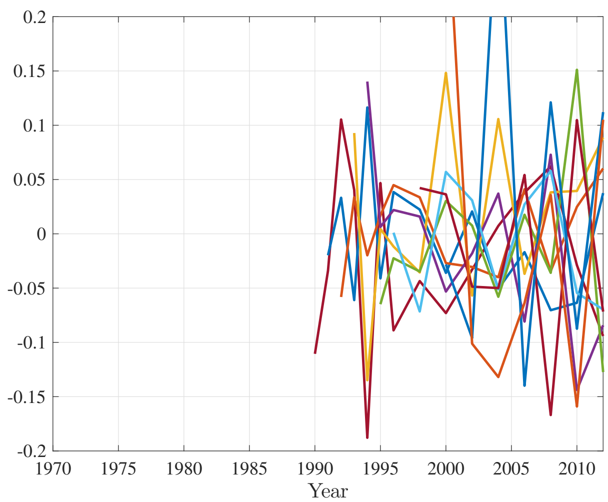
We interpret $\phi_{i,t}$ as skill, regardless of its persistence; however, it is possible to rewrite the problem such that skills are time-invariant and non-skill shocks include an autoregressive component (along with



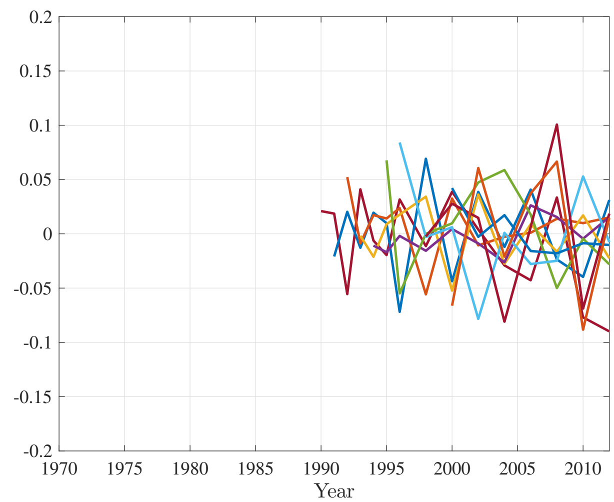
(a) Cohort Group 3, Non-College



(b) Cohort Group 3, College



(c) Cohort Group 4, Non-College



(d) Cohort Group 4, College

Figure 9: $\mathrm{Cov}(\Delta(w_t/\mu_t), \Delta(w_{t'}/\mu_{t'}))$ for Men by Cohort Group

Notes: Figure reports covariances for cohort groups $C \in \{3, 4\}$ where each line holds t' fixed and varies $t \geq t' + 7$.

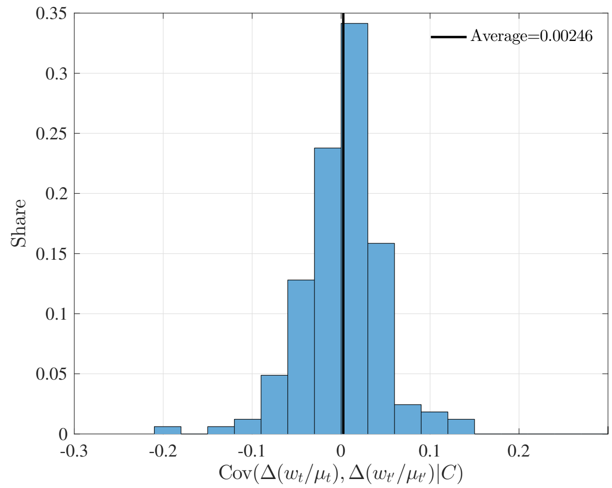


Figure 10: Distribution of $\text{Cov}(\Delta(w_t/\mu_t), \Delta(w_{t'}/\mu_{t'}))$ for all (t, t', C) for Low-Experience Men

Notes: Figure reports distribution of covariances based on (t, t') satisfying $t \geq t' + 7$ for cohort groups $C \in \{3, 4\}$ when some individuals in these cohort groups had less than 21 years of experience in year t (i.e., $t \leq 1991$ for $C = 3$ or $t \leq 2001$ for $C = 4$).

transitory shocks, ε_t):

$$w_{i,t} = \mu_t \psi_i + \tilde{\phi}_{i,t} + \varepsilon_{i,t}, \qquad \text{where} \quad \tilde{\phi}_{i,t} = \tilde{\rho}_t \tilde{\phi}_{i,t-1} + \tilde{\nu}_{i,t},$$

letting $\tilde{\phi}_{i,t} \equiv \mu_t \phi_{i,t}$, $\tilde{\rho}_t \equiv \rho_t \mu_t / \mu_{t-1}$, and $\tilde{\nu}_{i,t} \equiv \mu_t \nu_{i,t}$. This shows that the distinction between skill vs. non-skill persistent shocks is not important from a statistical point of view nor for identification of μ_t.

Due to the correlation between skill growth and past skill levels, our IV estimator will generally produce biased estimates for growth rates in skill returns when $\rho_t \neq 1$.[34] However, we show in Appendix C.3 that if $\text{Var}(\psi) > 0$ the evolution of both ρ_t and μ_t over time can still be identified under Assumption 2 and other plausible conditions.

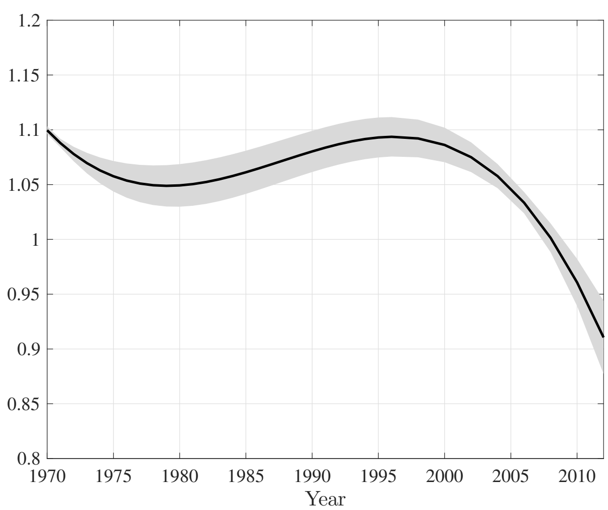
Identification breaks down when $\text{Var}(\psi) = 0$. In this case, our IV estimator (using past residuals as instruments) identifies $(\rho_t \mu_t - \mu_{t-1})/\mu_{t-1}$, and it is not generally possible to separate growth in skill returns from skill convergence without strong assumptions. This raises concerns that estimated declines in skill returns over the late-1980s and 1990s (see Figure 7) could instead reflect particularly strong skill convergence (i.e., $\rho_t < 1$) over those years.

Our examination of skill growth using test scores in Section 2.3 suggests that $\rho_t \approx 1$ for experienced workers in the HRS, covering the late-1990s and 2000s. We now use equally weighted MD estimation to

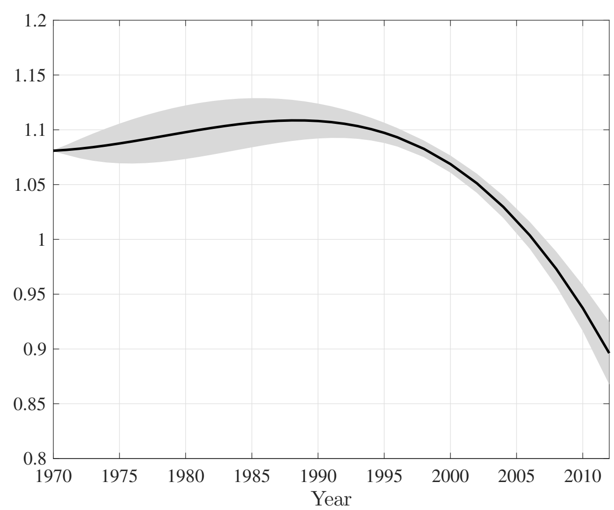
[34]The IV estimator converges to $\gamma_{t,t'}$ in (11), where $\text{Cov}(\Delta\theta_t, \theta_{t'}) = (\rho_t - 1)\,\text{Cov}(\phi_t, \phi_{t'})$. The term $\text{Cov}(\phi_t, \phi_{t'})$ generally depends on both t and t'. For example, when $t' \leq t - k - 1$, $\text{Cov}(\phi_t, \phi_{t'}) = \prod_{j=t'+1}^{t-1} \rho_j \, \text{Var}(\phi_{t'})$. The GMM estimates of Tables 5 and 6 do not support this dependence of $\gamma_{t,t'}$ on t'.

estimate the model with AR(1) skills as defined in equation (14) to account for the possibility that $\rho_t < 1$ over the longer time period we examine in the PSID. We begin by assuming that $\rho_t = \rho$ is time-invariant, but also consider the case in which ρ_t follows a cubic polynomial in time. We estimate ρ_t, μ_t (normalizing $\mu_{1985} = 1$), $\text{Var}(\nu_t|c)$, and $\text{Var}(\psi|c)$ separately for non-college and college men. To improve precision and facilitate estimation, we assume that $\text{Var}(\psi|c)$ is a cubic polynomial in entry cohort c and that $\text{Var}(\nu_t|c)$ is a cubic time trend multiplied by a quadratic experience trend. Long autocovariances for workers with at least 21 years of experience are targeted.[35]

Our estimates suggest that unobserved skills (for experienced men) are not mean-reverting, at least over most of the time period we examine. When assuming time-invariant $\rho_t = \rho$, its estimated value is 1.071 (0.001) for non-college men and 1.064 (0.001) for college men. Based on the more general time-varying ρ_t case, Figure 11 shows a modest increase in ρ_t over the 1980s and early-1990s, falling thereafter. There is no indication that ρ_t drops over the late-1980s and 1990s, which might explain our sharply falling IV estimated returns to skills over those years. Indeed, Figure 12 shows that the estimated μ_t series (for both fixed and time-varying ρ_t) are very similar to baseline estimates reported in Figure 7. Appendix Figure E-6 displays estimated $\text{Var}(\psi|c)$ for the time-varying ρ_t case. These estimates are generally consistent with $\text{Var}(\psi|c) > 0$, with estimates significantly greater than zero over labor-market-entry cohorts at the heart of our sample.



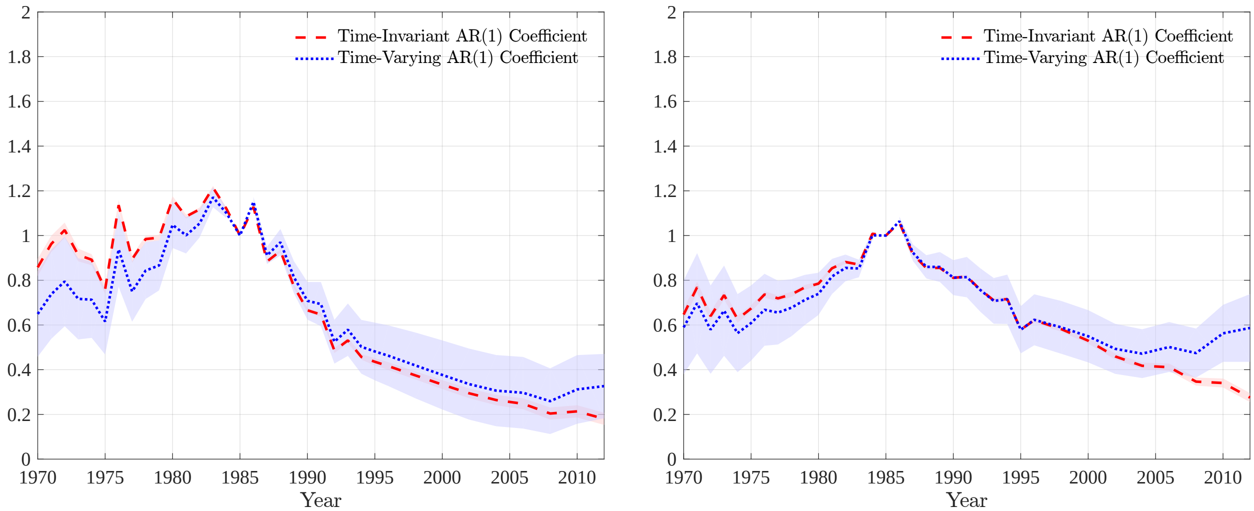
(a) Non-College Men



(b) College Men

Figure 11: ρ_t implied by MD estimates allowing for time-varying AR(1) skill shocks, 21–40 years of experience

[35] Specifically, we target $\widehat{\text{Cov}}(w_t, w_{t'}|E_j)$ for all $t - t' \geq 6$ and ten-year experience groups, E_j (21–30 and 31–40 years of experience in year t). There are 729 targeted autocovariances each for non-college and college men. See Appendix D for additional details.



(a) Non-College Men

(b) College Men

Figure 12: μ_t implied by MD estimates allowing for time-varying vs. time-invariant AR(1) skill shocks, 21–40 years of experience

3.6 Persistent Non-Skill Shocks

Results, thus far, rely on limited persistence in non-skill shocks, ε_t. Condition (iv) of Assumption 1 is convenient but not critical to our approach using wage dynamics to understand the evolution of skill returns. In this subsection, we consider the case in which ε_t contains an autoregressive process. Specifically, we assume that

$$\varepsilon_{i,t} = \varphi_{i,t} + \tilde{\varepsilon}_{i,t}, \qquad \text{where} \quad \varphi_{i,t} = \rho\varphi_{i,t-1} + \nu_{i,t}, \tag{15}$$

and $\tilde{\varepsilon}_{i,t}$ has limited persistence.

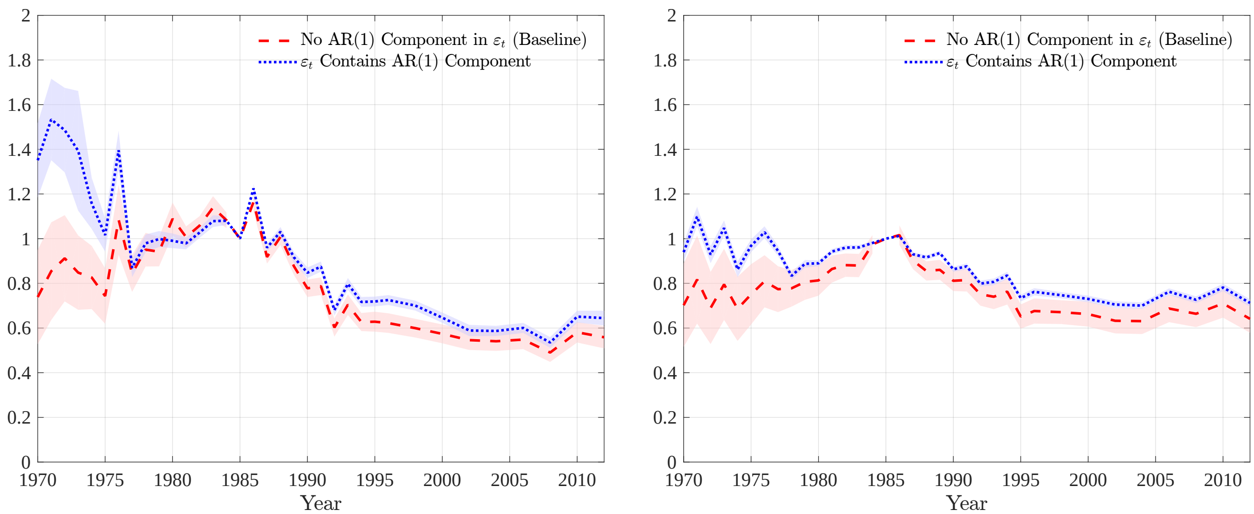
With this more general process for non-skill shocks, we strengthen conditions (i)–(iii) of Assumption 1 slightly and maintain our baseline assumption on the limited persistence of transitory non-skill shocks.

Assumption 3. *For any cohort c, (i)* $\mathrm{Cov}(\Delta\theta_t, \theta_{t'}|c) = 0$ *for all $t - t' \geq 1$; (ii)* $\mathrm{Cov}(\theta_t, \varphi_{t'}|c) = \mathrm{Cov}(\theta_t, \nu_{t'}|c) = \mathrm{Cov}(\theta_t, \tilde{\varepsilon}_{t'}|c) = 0$ *for all t, t'; (iii)* $\mathrm{Cov}(\varphi_{t'}, \nu_t|c) = \mathrm{Cov}(\nu_{t'}, \nu_t|c) = 0$ *for all $t - t' \geq 1$,* $\mathrm{Cov}(\varphi_t, \tilde{\varepsilon}_{t'}|c) = \mathrm{Cov}(\nu_t, \tilde{\varepsilon}_{t'}|c) = 0$ *for all t, t'; and (iv) for known $k \geq 1$,* $\mathrm{Cov}(\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t'}|c) = 0$ *for all $t - t' \geq k$.*

This assumes that skill growth is uncorrelated with past skills and that the evolution of skills is unrelated to the process for non-skill shocks; although, the influence of non-skill shocks on wages never fully disappears. Appendix C.4 establishes identification of skill returns, as well as ρ and the variance of initial skills and skill growth, under Assumption 3.[36]

[36] Indeed, we establish identification for $\varphi_t \sim \mathrm{ARMA}(1, q)$ as sometimes used in the literature on earnings dynamics.

As in the previous subsection, we turn to equally weighted MD using long residual autocovariances to estimate this very general specification for wage residuals separately for non-college and college men.[37] For $k = 6$, Figure 13 compares the estimated μ_t sequence for our baseline specification ($\varepsilon_t = \tilde{\varepsilon}_t$) vs. the specification that includes AR(1) non-skill shocks in equation (15). In both cases, the estimated paths for μ_t present the familiar pattern of rising returns in the early-1980s, followed by significant declines over the late-1980s and 1990s. Importantly, the evolution of skill returns is largely unaffected by the introduction of an autoregressive non-skill component.[38]



(a) Non-College

(b) College

Figure 13: Estimated μ_t with and without an AR(1) non-skill component

The literature on earnings dynamics estimates a similar structure for log earnings residuals to that of this subsection; although, this literature has focused primarily on the relative importance of permanent vs. transitory shocks while ignoring changes in the returns to unobserved skills.[39] Haider (2001) and Moffitt and Gottschalk (2012) are notable exceptions in that they also estimate the evolution of returns to unobserved skills using the PSID.[40] We show in Appendix E.6 that by restricting the autocovariance

[37]Specifically, we estimate μ_t (normalizing $\mu_{1985} = 1$), $\text{Var}(\Delta\theta_t|c)$ (assuming a cubic time trend multiplied by a quadratic experience trend), ρ, $\text{Var}(\nu_t|c)$ (assuming a year-specific constant multiplied by a quadratic experience trend), and $\text{Var}(\psi|c)$ (assuming a cubic polynomial in entry cohort c). We target $\widehat{\text{Cov}}(w_t, w_{t'}|E_j)$ for all $t - t' \geq 6$ and ten-year experience groups, E_j (1–10,..., 31–40 years of experience in year t). There are 855 targeted autocovariances each for non-college and college men. See Appendix D for additional details.

[38]For both non-college and college men, we estimate $\rho \approx 0.8$.

[39]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.

[40]Other studies exploit different panel data sets on earnings to estimate very 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.

structure for non-skill shocks and the distributions of skills (over time and across cohorts), these studies estimate upward biased growth in skill returns, μ_t, over time.

4 Skill Distributions

We now consider identification and estimation of the variance of skills and non-skill shocks over time. We further decompose the variance of skills into contributions from heterogeneity in initial skills and variation due to idiosyncratic lifecycle skill growth.

To facilitate this analysis, we slightly strengthen conditions (i)–(iii) of Assumption 1 but maintain the limited persistence of non-skill shocks (Assumption 1(iv)).

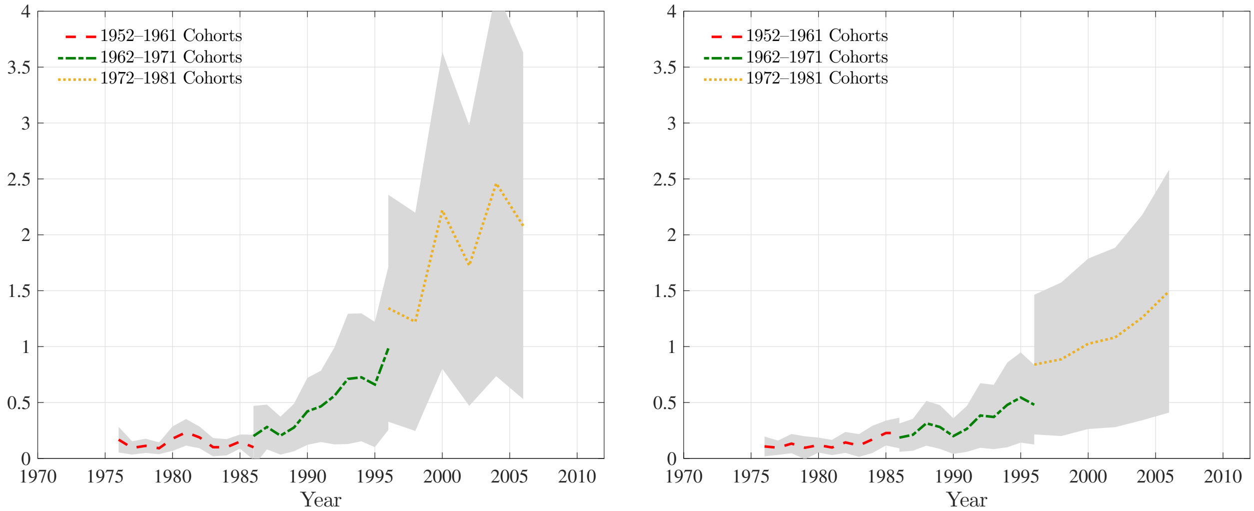
Assumption 4. *(i)* $\text{Cov}(\Delta\theta_t, \theta_{t'}) = 0$ *for all* $t - t' \geq 1$*; (ii)* $\text{Cov}(\theta_t, \varepsilon_{t'}) = 0$ *for all* t, t'*; and (iii) for known* $k \geq 1$*,* $\text{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$ *for all* $t - t' \geq k$*.*

The first two conditions imply that skills follow a random walk and are uncorrelated with non-skill shocks. This attributes all transitory wage innovations to the non-skill component. All three conditions imply that $\text{Cov}(w_t, w_{t'}) = \mu_t \mu_{t'} \text{Var}(\theta_{t'})$ for $t - t' \geq k$. As in Proposition 1, the IV estimator of equation (4) can be used to identify $\mu_{\underline{t}+k}, ..., \mu_{\bar{t}}$ (with one normalization). Given this and panel length $\bar{t} - \underline{t} \geq 2k$, the variance of unobserved skills, $\text{Var}(\theta_{t'}) = \text{Cov}(w_t, w_{t'})/\mu_t \mu_{t'}$, can be identified for all but the first and last k periods. This variance is not identified for earlier periods (without additional assumptions), because it cannot be separated from skill returns — only $\mu_t \text{Var}(\theta_t)$ can be identified for the first k periods. The unobserved skill variance cannot be identified for later periods, because it is impossible to distinguish between the roles of unobserved skills and transitory non-skill shocks without observing (distant) future wages.

Having identified the variance of unobserved skills over time, it is straightforward to then identify variation in skill growth, $\text{Var}(\Delta\theta_t) = \text{Var}(\theta_t) - \text{Var}(\theta_{t-1})$, for $t = \underline{t} + k + 1, ..., \bar{t} - k$, and the variance of non-skill shocks, $\text{Var}(\varepsilon_t) = \text{Var}(w_t) - \mu_t^2 \text{Var}(\theta_t)$, for all but the first and last k periods. Proposition 3 in Appendix C.1 extends these identification results to the variances of unobserved skills, skill growth innovations, and non-skill transitory shocks when these all vary by cohort.

Using future residuals as instruments. As discussed in Section 2.2, future wage residuals are not generally valid instruments in equations (4) or (6) when skills vary over time. Under Assumption 4, it is straightforward to show that IV regression using future wage residuals as instruments identifies:

$$\frac{\text{Cov}(\Delta w_t, w_{t''})}{\text{Cov}(w_{t-1}, w_{t''})} = \frac{\Delta\mu_t}{\mu_{t-1}} + \frac{\mu_t}{\mu_{t-1}} \frac{\text{Var}(\Delta\theta_t)}{\text{Var}(\theta_{t-1})} \quad \text{for } t'' \geq t + k. \tag{16}$$



(a) Non-College

(b) College

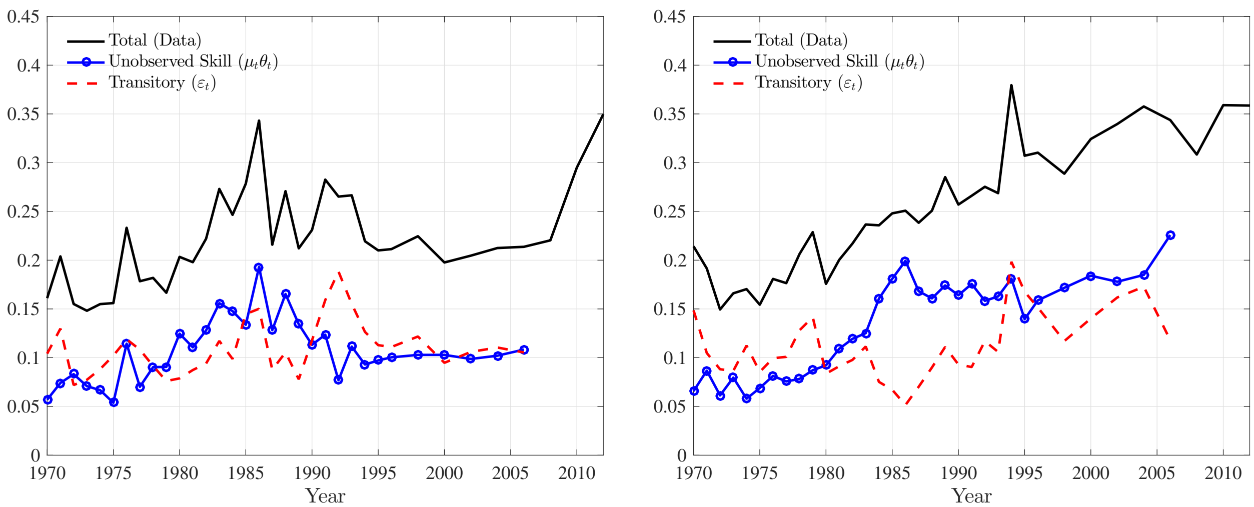
Figure 14: $\text{Var}(\theta_{t'}|C) = \Omega_{C,t'}/\mu_{t'}$ implied by MD estimates using long autocovariances, 21–40 years of experience

Since IV estimates using past residuals consistently estimate $\Delta\mu_t/\mu_{t-1}$, the difference between IV estimates obtained using future vs. past residuals as instruments can be used to identify the importance of skill growth innovations (relative to variation in lagged skill levels).

IV estimates presented in Tables 5 and 6, as well as Appendix E.3, empirically show that using future rather than lagged residuals as instruments nearly always produces higher estimated returns. Comparing estimates using future vs. past residuals as instruments, we show in Appendix E.3 that the variance of two-year skill growth relative to prior skill levels, $\frac{\text{Var}(\Delta\theta_{t-1}+\Delta\theta_t)}{\text{Var}(\theta_{t-2})}$, ranges from 0.16 to 0.29 over our sample period. Skills are not fixed and unchanging over the lifecycle.

Evolution of skill variation by cohort. Figure 14 reports $\text{Var}(\theta_{t'}|C) = \Omega_{C,t'}/\mu_{t'}$ obtained from estimates reported in Figures 7 and 8. This figure indicates that unobserved skill heterogeneity for the 1952–1961 birth cohorts was largely stable over the late-1970s and early-1980s. However, beginning in the early-1990s, the variance of unobserved skills grew sharply for both non-college and college men from the 1962–1971 and 1972–1981 birth cohorts.

Decomposing residual variation. We next explore the extent to which the long-term increase in residual variation reported in Figure 4 is driven by increasing variability of non-skill wage shocks, $\text{Var}(\varepsilon_t)$, vs. growing dispersion in skills and their returns, $\text{Var}(\mu_t\theta_t)$. Given our interest in understanding these trends for all workers, we focus on our baseline model under Assumption 4, estimated separately by education (for all ages) using MD estimation. We have already estimated this model in Section 3.6, imposing mild cohort- and time-based smoothness assumptions on the variance of initial skills and skill growth innovations. Estimated μ_t profiles are shown as red dashed lines in Figure 13. Figure 15 decomposes $\mathrm{Var}(w_t)$ into its skill and non-skill components over time. The main trends in residual inequality are driven by inequality in skills (multiplied by their returns) for both education groups; however, growth in the variance of non-skill shocks contributes to rising residual inequality in the late-1980s and 1990s for college men.

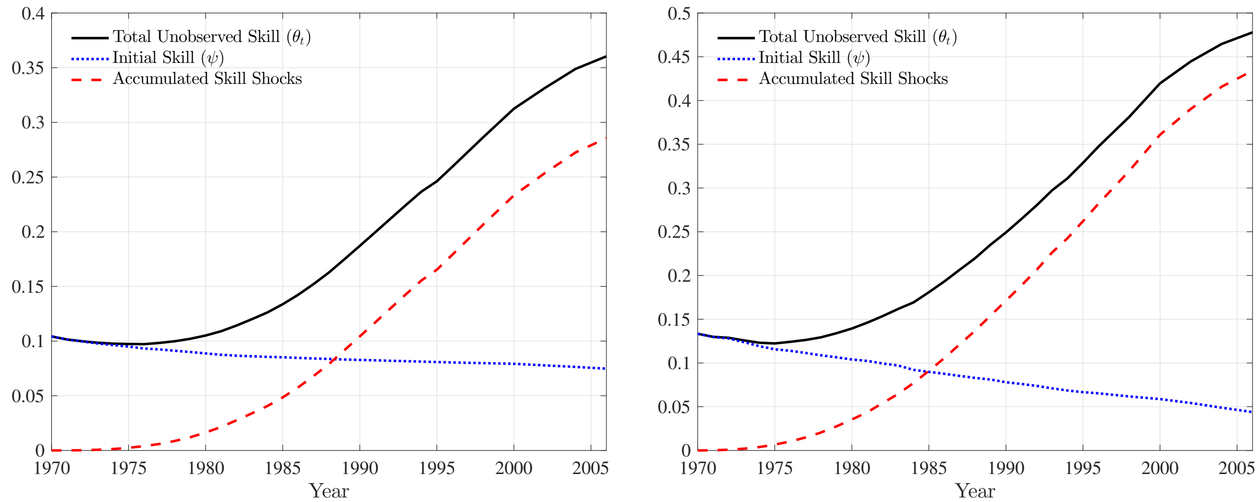


(a) Non-College (b) College

Figure 15: Log wage residual variance decomposition

Decomposing skill variation. Given the importance of rising skill inequality since the mid-1980s, we examine the extent to which changes in the distribution of initial skills vs. the distribution of skill growth innovations contribute to this trend. Figure 16 decomposes the annual variance of skills into the variance of initial skills, $\mathrm{Var}(\psi)$, and the variance of skills accumulated since labor market entry, $\mathrm{Var}(\theta_t - \psi)$. This figure reveals modest declines in initial skill inequality due to long-term secular declines in $\mathrm{Var}(\psi|c)$ across cohorts. The large rise in skill inequality is, therefore, driven by a strong increase in the variance of skill growth innovations since the mid-1980s.

Summary. Altogether, our analysis suggests that rising residual inequality in the late-1970s and early-1980s was driven primarily by increasing returns to unobserved skill for both non-college- and college-educated men. Residual inequality among non-college men declined slightly in the late-1980s, balancing two strong opposing forces: a sharp decline in skill returns, partially offset by a strong increase in skill



(a) Non-College

(b) College

Figure 16: Skill variance decomposition

dispersion. Meanwhile, college men continued to experience rising residual inequality throughout our sample period. As with non-college men, a fall in their return to skill was offset by an increase in skill inequality; however, this was accompanied by rising variability of non-skill shocks over the late-1980s and early-1990s. For both education groups, the strong secular increase in skill dispersion beginning in the 1980s was driven exclusively by increased volatility in skills rather than growing dispersion in skill levels at labor market entry.

5 Occupations and Multiple Unobserved Skills

Thus far, we have focused on a single dimension of skill with a specification for log wage residuals, equation (2) that is broadly consistent with traditional assignment models of the labor market with a continuum of skill levels and jobs (Tinbergen, 1956; Sattinger, 1993; Costinot and Vogel, 2010).[41] In this framework, each worker is assigned to a different job based on a single-dimensional ranking of worker productivity, i.e., skill. This section considers more general wage specifications in which skills are rewarded differently by occupations. We also consider the interpretation of our IV estimator when there are multiple skills earning different returns in the market. Throughout this analysis, we continue to account for the fact that skills vary over time for individual workers. (See Sanders and Taber (2012) for a survey of the literature on lifecycle wage dynamics in models with multiple skills and occupations.)

[41] See Lochner, Park, and Shin (2018) for a specification of production technology and skill and job productivity distributions in a traditional assignment model that yields equation (2) as the equilibrium log wage function.

5.1 Occupations

Motivated by task-based models of the labor market in which workers are assigned to a limited set of different tasks or jobs (Autor, Levy, and Murnane, 2003; Acemoglu and Autor, 2011; Cortes, 2016; Acemoglu and Restrepo, 2022; Acemoglu and Loebbing, 2022), we extend our analysis to consider occupation-specific wage functions with

$$w_{i,t} = \gamma_t^{o_{i,t}} + \mu_t^{o_{i,t}} \theta_{i,t} + \varepsilon_{i,t}, \tag{17}$$

where $o_{i,t}$ denotes the occupation for worker i in year t, and we normalize $\mu_t^o = 1$ and $\gamma_t^o = 0$ for a single occupation-year pair. Average wages may differ across occupations due to differences in wage functions (i.e., (γ_t^o, μ_t^o)) and in the average skill levels of workers in those occupations, with $\mathrm{E}[w_t|o_t = o] = \gamma_t^o + \mu_t^o \, \mathrm{E}[\theta_t|o_t = o]$. For example, management occupations might provide a higher return to skill and be filled by more-skilled workers relative to low-level clerical occupations. Although assignment and task-based models generally assign all workers of a given skill level to a single task/job, workers with identical skills may choose to work in different occupations due to search and information frictions or heterogeneous preferences for job attributes (e.g., Papageorgiou, 2014; Taber and Vejlin, 2020; Guvenen et al., 2020; Lise and Postel-Vinay, 2020; Roys and Taber, 2022; Adda and Dustmann, 2023).

Cortes (2016) considers a special case of equation (17), assuming time-invariant skill returns and skill levels (i.e., $\mu_t^o = \mu^o$ and $\theta_{i,t} = \theta_i$). We further note that equation (17) generalizes related specifications commonly employed in studies of sectoral or firm differences in pay, which typically assume that skills are fixed over time.[42]

Occupation-specific skill returns can be identified by strengthening Assumption 1 to condition on recent occupation histories.

Assumption 5. *For known $k \geq 1$ and for all $t - t' \geq k + 1$: (i)* $\mathrm{Cov}(\Delta\theta_t, \theta_{t'}|o_t, o_{t-1}, o_{t'}) = 0$*; (ii)* $\mathrm{Cov}(\Delta\theta_t, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'}) = 0$*; (iii)* $\mathrm{Cov}(\varepsilon_t, \theta_{t'}|o_t, o_{t-1}, o_{t'}) = \mathrm{Cov}(\varepsilon_{t-1}, \theta_{t'}|o_t, o_{t-1}, o_{t'}) = 0$*; and (iv)* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'}) = \mathrm{Cov}(\varepsilon_{t-1}, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'}) = 0$.

Assumption 5 requires that skill dynamics not depend on or influence occupation choices, much as the literature on firm-specific returns assumes that job changes are exogenous (Abowd, Kramarz, and

[42]This specification is broadly consistent with the multi-sector assignment model of Gola (2021), where wage functions vary with sector (e.g., manufacturing, services) rather than occupation. It is also related to that of Bonhomme, Lamadon, and Manresa (2019), who allow γ_t^o and μ_t^o to differ across firms instead of occupations, assuming time-invariant worker skills (i.e., $\theta_{i,t} = \theta_i$). Following the canonical Abowd, Kramarz, and Margolis (1999), the literature on pay differentials across firms typically focuses on estimating differences in the intercept term across firms (i.e., $\gamma_t^o = \gamma^o$), assuming no variation in returns to time-invariant individual skills, ($\mu_t^o = \mu$ and $\theta_{i,t} = \theta_i$). However, several more recent studies estimate time-varying premia. For example, Card, Heining, and Kline (2013) estimate firm premia separately by subperiods (i.e., rolling-window estimation), while Lachowska et al. (2023) and Engbom, Moser, and Sauermann (2023) allow firm premia to vary freely over time.

Margolis, 1999). This assumption is likely too strong for young workers simultaneously making early skill investment and career decisions; however, it is more plausible for older workers for whom skill variation is likely to be idiosyncratic and who face weaker incentives to search for a new occupation in response to skill or non-skill wage innovations given their shorter career horizon, greater skill specialization, and stronger (revealed) preferences for current job/occupation amenities (Cavounidis and Lang, 2020).[43]

As shown in Appendix C.6, μ_t^o can be identified for all occupations (in all but the first k years) under Assumption 5 using the following IV estimator (for $t - t' \geq k + 1$):

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'}|o_t, o_{t-1}, o_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'}|o_t, o_{t-1}, o_{t'})} = \frac{\mu_t^{o_t} - \mu_{t-1}^{o_{t-1}}}{\mu_{t-1}^{o_{t-1}}}. \tag{18}$$

Since occupational mobility is low, especially among older workers, we highlight that occupation-specific growth in skill returns can be identified from occupation stayers (from $t-1$ to t) alone. This is noteworthy, because Assumption 5 is most credible for occupation-stayers who are largely representative of the population given low rates of occupational switching (especially among experienced workers).[44]

While occupation-specific growth in skill returns can be identified from occupation stayers (from $t - 1$ to t), occupation switchers must be incorporated to identify the relative returns to skill across occupations, $\mu_t^o / \mu_t^{o'}$, and the sequence of occupation-specific wage levels, γ_t^o. In addition to Assumption 5, identification of γ_t^o requires $\mathrm{E}[\Delta\theta_t|o_t, o_{t-1}] = 0$ and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}] = \mathrm{E}[\varepsilon_{t-1}|o_t, o_{t-1}] = 0$, which imply

$$\mathrm{E}[\Delta w_t|o_t, o_{t-1}] - \left(\frac{\mu_t^{o_t} - \mu_{t-1}^{o_{t-1}}}{\mu_{t-1}^{o_{t-1}}}\right)\mathrm{E}[w_{t-1}|o_t, o_{t-1}] = \gamma_t^{o_t} - \frac{\mu_t^{o_t}}{\mu_{t-1}^{o_{t-1}}}\gamma_{t-1}^{o_{t-1}}.$$

Given small sample sizes for many occupation sequences $(o_t, o_{t-1}, o_{t'})$, we rely on the stronger assumptions $\mathrm{E}[\Delta\theta_t|o_t, o_{t-1}, o_{t'}] = 0$ and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_{t-1}|o_t, o_{t-1}, o_{t'}] = 0$, which allows us to condition only on (o_t, o_{t-1}) in estimation of μ_t^o and γ_t^o. See Appendix C.6.

5.1.1 2SLS Estimation of Skill Return Growth for Occupation-Stayers

We use the PSID to estimate growth in skill returns for occupation-stayers in two broad and exclusive occupation groups (cognitive and routine occupations) considered by Cortes (2016).[45] We also estimate

[43] Gathmann and Schönberg (2010) show that older workers make fewer occupational changes and that those changes entail smaller changes in occupational task content. Gervais et al. (2016) also document declining occupational mobility over the lifecycle.

[44] Assumption 5 implies that estimated return growth for stayers in occupation $o_t = o_{t-1} = o$ should not depend on earlier occupation $(o_{t'})$. Results reported in Appendix E.7 confirm this prediction.

[45] Cortes (2016) also considers manual occupations, but our sample contains too few observations to obtain precise results for its associated parameters. Appendix E.7 provides details on occupation classifications in the PSID.

skill returns for those who remain in occupations with high social skill requirements, based on the measure of social skill intensity considered by Deming (2017).[46] As Deming (2017) notes, there is considerable overlap between cognitive occupations and social occupations – in our sample, 59% of worker-year observations in cognitive occupations are also in social occupations and 76% of observations in social occupations are also in cognitive occupations.

As with equation (10) earlier, we use 2SLS (with lagged residuals as instruments) to estimate two-year growth rates in occupation-specific skill returns based on

$$\Delta_2 w_{i,t} = \left(\gamma_t^{o_{i,t}} - \left[\frac{\mu_t^{o_{i,t}}}{\mu_{t-1}^{o_{i,t-2}}}\right]\gamma_{t-2}^{o_{i,t-2}}\right) + \left[\frac{\mu_t^{o_{i,t}} - \mu_{t-2}^{o_{i,t-2}}}{\mu_{t-1}^{o_{i,t-2}}}\right] w_{i,t-2} + \underbrace{\left(\varepsilon_{i,t} - \left[\frac{\mu_t^{o_{i,t}}}{\mu_{t-2}^{o_{i,t-2}}}\right]\varepsilon_{i,t-2} + \mu_t^{o_{i,t}}\Delta_2\theta_{i,t}\right)}_{\equiv \xi_{i,t}}, \quad (19)$$

estimated separately for stayers with $o_{i,t} = o_{i,t-2} = o$ for cognitive, routine, or social occupation groups.[47]

Figure 17 reports implied skill returns (relative to μ_{1985}^o) for all men who remain in cognitive, routine, or social occupations. Panel (a) reports estimates based on workers of all experience levels, while panel (b) reports estimates based on those with 21–40 years of experience. In both panels, we obtain similar estimated return profiles for job stayers regardless of their occupation type, indicating strong declines in the returns to skill in cognitive, routine, and social occupations. The estimated return profiles also accord well with those estimated earlier for the full sample.

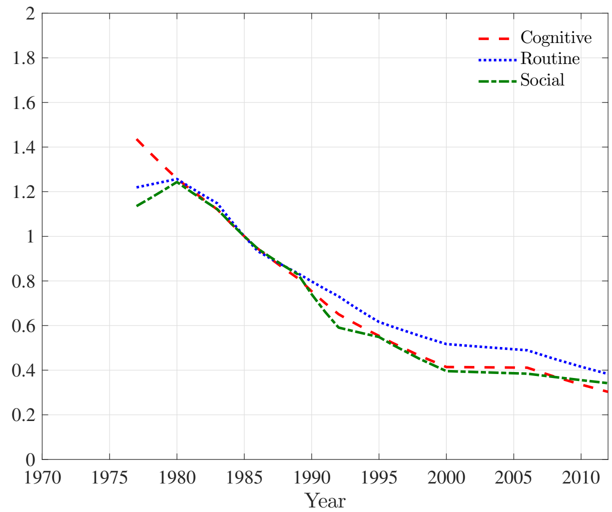
5.1.2 GMM Estimation of γ_t^o and μ_t^o for Cognitive and Routine Occupations

In order to estimate differences in the levels of skill returns across occupations and occupation-specific average wage differences, we must also exploit occupational switchers. We use GMM to simultaneously estimate μ_t^o and γ_t^o, now including all occupation stayers and switchers in our sample. Based on equation (19), we exploit the following moments in the PSID: $E[\xi_t | o_t, o_{t-2}] = 0$ and $E[w_{t'}\,\xi_t | o_t, o_{t-2}] = 0$, where we use lagged residuals $w_{t'}$ from periods $t-8$ and $t-9$ (or $t-10$ in later years) as instruments. See Appendix E.7 for details.

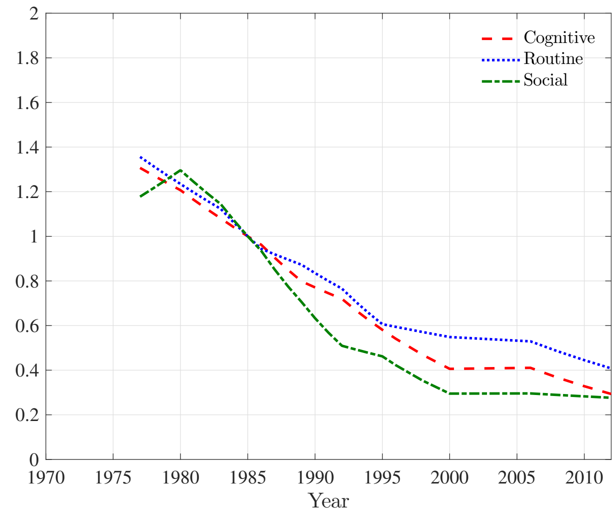
Given the substantial overlap between cognitive and social occupations (and similar skill return profiles in Figure 17), we focus this analysis on the two mutually exclusive categories from Cortes (2016): cognitive and routine occupations. Here, we normalize $\mu_t^o = 1$ and $\gamma_t^o = 1$ for routine occupations in 1985; however, no normalizations are needed for cognitive occupations. Figure 18(a) shows that

[46] We define social occupations as those that fall in the top third of the social skill intensity distribution in the pooled sample of worker-year observations. See Appendix E.7.

[47] We use (w_{t-8}, w_{t-9}) as instruments when both are available (in early survey years); otherwise, we use (w_{t-8}, w_{t-10}) as instruments. Our use of two-year differences relies on the natural modification of all assumptions to condition on $(o_t, o_{t-2}, o_{t'})$.



(a) All Experience Levels



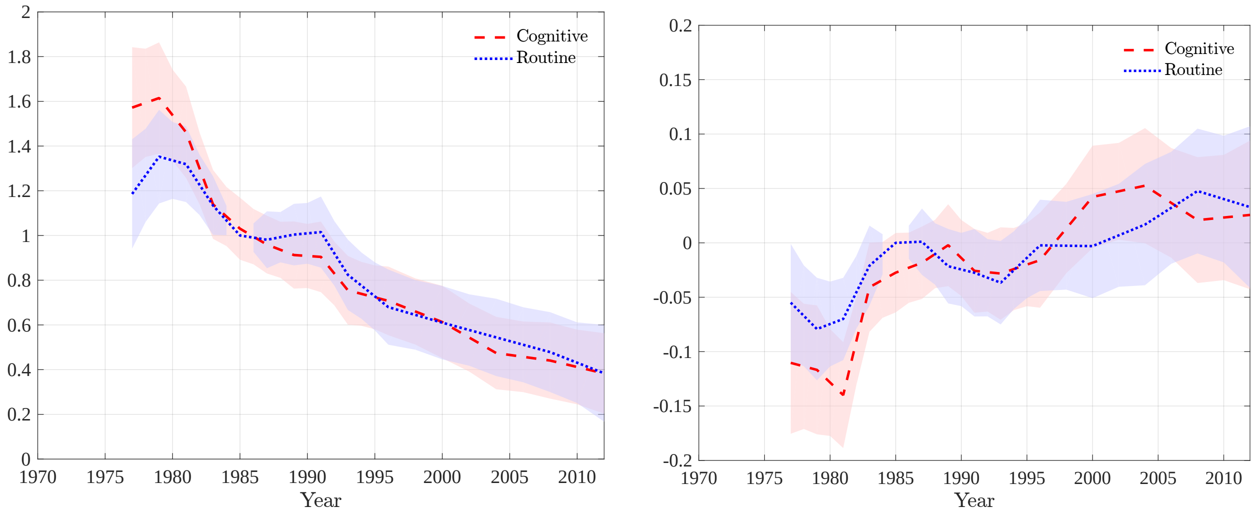
(b) 21–40 Years of Experiences in Year t

Figure 17: μ_t^o/μ_{1985}^o implied by 2SLS estimates for cognitive, routine, and social occupation-stayers between $t-2$ and t

estimated μ_t^o series both exhibit substantial declines over time, similar to the 2SLS estimates in Figure 17 and earlier estimates based on the full sample. We cannot reject that the two skill return series are equal using a standard J-test (p-value = 0.13). Despite sharp drops in the returns to skill in both occupations, Figure 18(b) shows sizeable increases in γ_t^o – nearly 0.20 in cognitive occupations and about 0.12 in routine occupations. Appendix Figure E-10 shows similar time patterns for μ_t^o and γ_t^o when using only workers with 21–40 years of experience in year t.

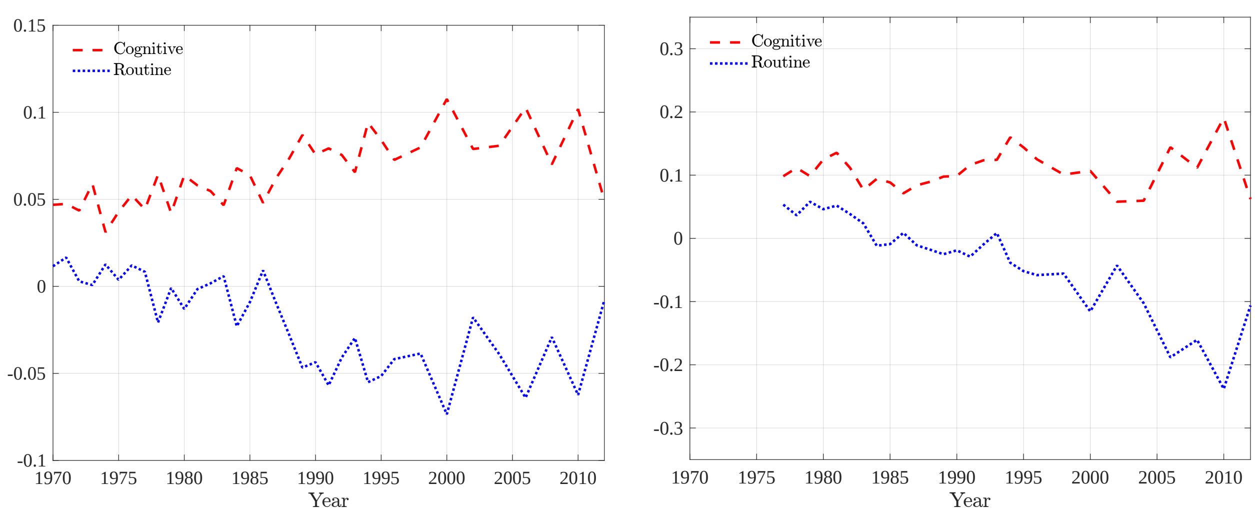
Using the estimates reported in Figure 18 and average log wage residuals by occupation, we can estimate the evolution of average skills by occupation over time from $\mathrm{E}[\theta_t|o_t] = \left(\mathrm{E}[w_t|o_t] - \gamma_t^{o_t}\right)/\mu_t^{o_t}$. Figure 19 shows the evolution of average log wage residuals and average skills for cognitive and routine workers. During our sample period, average log wage residuals rose by about 0.05 for workers in cognitive occupations, while they fell by a similar amount in routine occupations. Together with estimated μ_t^o and γ_t^o, these imply little long-term change in the average unobserved skills of workers in cognitive occupations but roughly 20 log point declines in the average unobserved skills of workers in routine jobs. Notably, these represent skill changes conditional on worker education and experience levels.

Appendix Figure E-9 shows that failing to account for changes in the returns to skill over time (i.e., assuming $\mu_t^o = \mu^o$ for all t as in Cortes (2016)) yields estimates that exhibit little long-term change in γ_t^o or average skills, $\mathrm{E}[\theta_t|o_t]$, in both cognitive and routine occupations. Thus, accounting for the estimated declines in skill returns has important implications for trends in occupation-specific skill levels.



(a) μ_t^o (b) γ_t^o

Figure 18: GMM estimates of μ_t^o and γ_t^o for cognitive and routine occupations



(a) $\mathrm{E}\left[w_t|o_t\right]$ (b) $\mathrm{E}\left[\theta_t|o_t\right]$

Figure 19: Average log wage residual and skill for cognitive and routine occupations

5.2 Multiple Unobserved Skills

A growing literature emphasizes the multi-dimensional nature of skills, suggesting that the returns to some types of skills have risen while returns to others have fallen (Castex and Dechter, 2014; Deming, 2017; Edin et al., 2022). Motivated by these studies, we now consider wage functions that depend on multiple unobserved skills, denoted by $\theta_{i,j,t}$ for $j = 1, ..., J$:

$$w_{i,t} = \sum_{j=1}^{J} \mu_{j,t}\theta_{i,j,t} + \varepsilon_{i,t}, \tag{20}$$

where $\mu_{j,t}$ reflects the market-level value of skill j in year t.[48]

We now make use of a multi-dimensional version of Assumption 4 to show that our IV estimator identifies a weighted-average growth rate across all skill returns. Specifically, we assume that growth in *each* type of skill is uncorrelated with *all* past skill levels.[49]

Assumption 6. *(i)* $\mathrm{Cov}(\Delta\theta_{j,t}, \theta_{j',t'}) = 0$ *for all* j, j'*, and* $t - t' \geq 1$*; (ii)* $\mathrm{Cov}(\theta_{j,t}, \varepsilon_{t'}) = 0$ *for all* j, t, t'*; and (iii)* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$ *for all* $t - t' \geq k$*.*

This assumption implies:

$$\mathrm{Cov}(\Delta w_t, w_{t'}) = \sum_{j=1}^{J}\sum_{j'=1}^{J} \Delta\mu_{j,t}\, \mu_{j',t'}\, \mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}), \quad \text{for } t - t' \geq k + 1, \tag{21}$$

where we highlight that the $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'})$ are within-period covariances across skills. Equation (21) shows that when all (within-period) correlations between skills are non-negative, a positive (negative) $\mathrm{Cov}(\Delta w_t, w_{t'})$ for $t - t' \geq k + 1$ implies that total returns $\mu_{j,t}$ are increasing (decreasing) for at least one skill. Thus, Figures 3 and 5 suggest that the returns to at least one skill declined sharply over the late-1980s and 1990s. Consistent with this conclusion, Castex and Dechter (2014) and Deming (2017) estimate strong declines in returns to cognitive skill over this period.

As the next result shows, our IV estimator provides a useful summary measure of skill return growth when there are many skills whose returns grow at different rates. For this result, it is useful to define

[48] Multi-dimensional assignment and search/matching models of the labor market can give rise to equilibrium log wage functions of the form in equation (20) (e.g. Lindenlaub, 2017; Lise and Postel-Vinay, 2020; Lindenlaub and Postel-Vinay, 2023). These models can also yield more general log wage functions of the entire skill vector, in which case equation (20) can be thought of as a linear approximation. Equation (20) is reminiscent of wage (rather than log wage) functions in Heckman and Scheinkman (1987) when worker characteristics can be "unbundled".

[49] A weaker assumption analogous to Assumption 1 (generalized to account for multiple skills) will also ensure that the IV estimator identifies a weighted-average growth in returns. We impose the stronger conditions based on Assumption 4 here to facilitate interpretation of the weights.

$\overline{\theta}_{i,t} \equiv \sum_{j=1}^{J} \mu_{j,t}\theta_{i,j,t}$, the total value of a worker's skill vector in period t.

Proposition 2. *If Assumption 6 holds, then for all $t - t' \geq k + 1$ the IV estimator identifies a weighted-average growth rate across all skill returns:*

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \sum_{j=1}^{J} \omega_{j,t',t-1} \left(\frac{\Delta\mu_{j,t}}{\mu_{j,t-1}}\right), \tag{22}$$

with weights for $j = 1, ..., J$ given by $\omega_{j,t',t-1} = \mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})\mu_{j,t-1} / \sum_{j'=1}^{J} \mathrm{Cov}(\theta_{j',t'}, \overline{\theta}_{t'})\mu_{j',t-1}$. If $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$, then the weights $\omega_{j,t',t-1} \in [0, 1]$ for all j.

When there are multiple skills, our IV estimator identifies the weighted-average growth rate across all skill returns, where the weights, $\omega_{j,t',t-1}$, are larger for skills that are strongly related to wages (in t') and which have a high return (in $t-1$).[50]

The multi-skill problem effectively reduces to the single-skill problem when the relative productivity of different skills is time-invariant: $\mu_{j,t}/\mu_{1,t} = \overline{\mu}_j$ for all j and t. As such, the IV estimator identifies growth rates for all skill returns during periods with fixed relative skill valuations.

Occupations as bundles of skills. A simple view of occupations, consistent with multi-dimensional assignment models (e.g., Lindenlaub, 2017; Lindenlaub and Postel-Vinay, 2023), is that they represent different combinations of skill-intensities, $\alpha^{o}_{j,t}$, leading to different wages by occupation:

$$w_{i.t} = \sum_{j=1}^{J} \mu_{j,t}\alpha^{o_{i,t}}_{j,t}\theta_{i,j,t} + \varepsilon_{i,t}. \tag{23}$$

The returns to skill j in occupation o in year t, $\tilde{\mu}^{o}_{j,t} = \mu_{j,t}\alpha^{o}_{j,t}$, depend on the market-level value for that skill, $\mu_{j,t}$, and the occupation-specific skill intensity factor, $\alpha^{o}_{j,t}$.

Conditioning all covariances in Assumption 6 on occupation sequence $(o_t = o_{t-1}, o_{t'})$, the IV estimator applied to stayers in occupation o (from $t-1$ to t) recovers a weighted average of skill-specific return growth, $\Delta\tilde{\mu}^{o}_{j,t}/\tilde{\mu}^{o}_{j,t-1} = (\Delta\mu_{j,t}\alpha_{j,t} + \mu_{j,t-1}\Delta\alpha^{o}_{j,t})/\tilde{\mu}^{o}_{j,t-1}$, in occupation o, where the returns to skills that are more important for wages in occupation o receive more weight.[51] Notice that stability of occupation skill intensities (i.e., $\alpha^{o}_{j,t} = \alpha^{o}_{j}$), as assumed by much of the literature (e.g., Autor and Dorn, 2013;

[50] Appendix C.7 further shows that the weights are proportional to the extent to which total productivity in period t' predicts the rewards from skill j in period $t-1$. Proposition 4 in Appendix C.7 shows that the IV estimator also reflects growth in a weighted-average measure of skill returns.

[51] See Appendix C.7 for details on all results in this subsection.

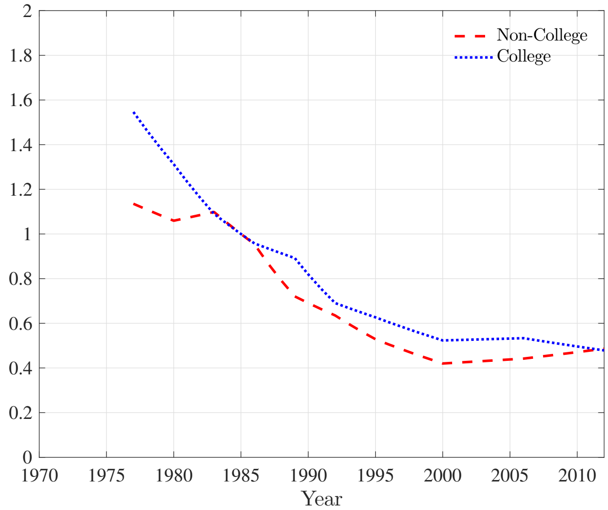


Figure 20: μ_t Implied by 2SLS Estimates for 3-Digit Occupation Stayers Between $t-2$ and t

Acemoglu and Autor, 2011; Böhm, 2020; Böhm, von Gaudecker, and Schran, 2024), would imply that IV estimates using stayers in occupation o identify weighted averages of $\Delta\mu_{j,t}/\mu_{j,t-1}$, where the weights continue to depend on occupation o (e.g., IV estimates based on stayers in sales- or communication-based occupations will largely reflect growth in the returns to social skills, while IV estimates based on stayers in manufacturing jobs will primarily reflect growth in returns to manual skills). Altogether, IV estimators applied to a diverse set of occupations will yield different skill return trends if either (i) relative skill intensities evolve differently across occupations or (ii) returns to various skills evolve differently over time. The similarity of IV estimated return series across occupation groups reported in Figure 17, therefore, suggests relatively stable occupation skill intensities and similar declines in the returns to a broad range of skills.[52]

Finally, we explore estimated returns for college vs. non-college men using a sample of all occupation-stayers (from $t-1$ to t), regardless of occupation. For stable within-occupation skill intensities, this identifies a weighted average of $\Delta\mu_{j,t}/\mu_{j,t-1}$ where the weights depend on the share of stayers in each occupation o. Figure 20 displays the implied skill return profiles (by education) for all 3-digit occupation stayers. Both estimated return series are very similar to our baseline estimates for the full sample reported in Figure 6.

[52]A few recent studies document within-occupation changes in the skill/task content/requirements of jobs (Spitz-Oener, 2006; Cavounidis et al., 2021; Cortes, Jaimovich, and Siu, 2023). Given the inherent challenges of such efforts, some of these documented changes may reflect changes in the skill levels of workers within occupations over time rather than changes in the actual tasks performed by workers. A separate challenge is that workers may perform different mixes of tasks within the same occupation (Autor and Handel, 2013; Spitz-Oener, 2006). In our analysis, any such differences would be interpreted as variation in worker skill bundles.

5.3 Occupation-Specific Wage Functions with Multiple Skills

Several studies consider a more substantial role for occupations in multi-skill models of the labor market (see, e.g., Gathmann and Schönberg, 2010; Yamaguchi, 2012, 2018; Böhm, 2020; Guvenen et al., 2020; Roys and Taber, 2022; Böhm, von Gaudecker, and Schran, 2024).[53] Motivated by this literature, we interpret our IV estimator when wages depend on multiple unobserved skills that are rewarded differently across occupations:

$$w_{i,t} = \gamma_t^{o_{i,t}} + \sum_{j=1}^{J} \mu_t^{o_{i,t}} \alpha_{j,t}^{o_{i,t}} \theta_{i,j,t} + \varepsilon_{i,t}. \tag{24}$$

In this case, occupation- and skill-specific returns, $\tilde{\mu}_{j,t}^{o} = \mu_t^o \alpha_{j,t}^o$, arise when occupations reward a worker's total productivity differently, where total productivity depends on the intensity of each skill used in that occupation.[54] We focus on estimating growth in occupation-specific returns to skills, μ_t^o; however, wage levels may also differ across occupations and over time, γ_t^o, due to, for example, compensating differences or market frictions. Estimating changes in γ_t^o is challenging with multiple skills, so we leave that to future work.

Conditioning covariances in Assumption 6 on occupation sequence $(o_t = o_{t-1}, o_{t'})$, it is straightforward to show that for occupation-stayers, our IV estimator identifies:

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'} | o_t = o_{t-1} = o, o_{t'} = o')}{\mathrm{Cov}(w_{t-1}, w_{t'} | o_t = o_{t-1} = o, o_{t'} = o')} = \frac{\Delta \mu_t^o}{\mu_{t-1}^o} + \frac{\mu_t^o}{\mu_{t-1}^o} \sum_{j=1}^{J} \tilde{\omega}_{j,t',t-1}^{o,o'} \left(\frac{\Delta \alpha_{j,t}^o}{\alpha_{j,t-1}^o} \right), \tag{25}$$

where the weights on skill intensity growth, $\tilde{\omega}_{j,t',t-1}^{o,o'}$, sum to one and are non-negative if all skills are non-negatively correlated conditional on occupations (o, o'). See Appendix C.8 for details.

If skill intensities do not vary over time within occupations (i.e., $\alpha_{j,t}^o = \alpha_j^o$), then our IV estimator for stayers identifies occupation-specific skill return growth, $\Delta\mu_t^o/\mu_{t-1}^o$, as in Section 5.1.[55] More generally, the IV estimator for stayers in occupation o also reflects any growth in skill intensities within that occupation. If log wage residuals are characterized by equation (24), the results in Figure 17 are consistent with similar growth in all skill intensities within manual, routine, and social occupations, coupled with similar declines in the returns to these skills within these occupations.

[53] The canonical Roy model (Roy, 1951) is a special case in which there are an equal number of occupations and skills with each skill rewarded only in its "own" sector. See Heckman and Sedlacek (1985), Keane and Wolpin (1997), and Kambourov and Manovskii (2009) for important empirical applications of this framework to occupational choice and wages.

[54] Equation (24) is analogous to wage functions in the skill-weights model of Lazear (2009); although, w_t reflects *log* wage (residuals) here rather than wages as in Lazear (2009).

[55] Appendix C.8 discusses our IV estimator applied to the sample of stayers in occupation $o_t = o_{t-1} = o$, regardless of past occupation $o_{t'}$, as well as for all stayers in any occupation $o_t = o_{t-1}$.

6 Returns Estimated from Administrative Earnings Data

Previous studies have documented different trends in income volatility when using administrative data rather than the PSID (see, e.g., Sabelhaus and Song, 2010; DeBacker et al., 2013).[56] We show in this section that estimated patterns for skill returns are similar to those already presented when using earnings records from IRS W-2 Forms (maintained by the Social Security Administration, SSA) linked with survey data from the SIPP. These data include the full SSA history of wage and salary measures for all linked respondents from 1951 to 2011.

Our analysis begins with a sample of US-born white men ages 16–64 who could be linked to any of nine SIPP panels (i.e., panels from 1984–2008). We work with log wage residuals constructed as with the PSID and restrict observations to years when individuals were no longer enrolled in school. We focus mainly on results using Detailed Earnings Records (DER), which are uncapped and available from 1978 onward; however, we also take advantage of Summary Earnings Records (SER) available since 1951, which report earnings capped at the FICA taxable maximum. See Appendix G for a detailed discussion of these data and our sample. We highlight Appendix Figures G-2 and G-3, which show very similar patterns to Figures 2 and 5 regarding convergence in predicted wage residuals given base-year residual quartiles and sharp declines in residual autocovariances $\mathrm{Cov}(w_t, w_b)$ over years $t \geq b + 6$ for fixed base year b. Together, these indicate declines in the return to skills over the late-1980s and 1990s, consistent with our PSID results.

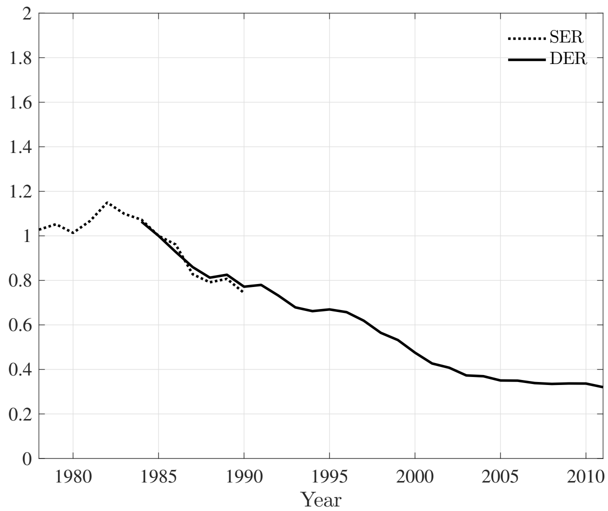
We use our IV estimator in equation (4) to estimate growth rates for skill returns using w_{t-7} as an instrument (consistent with $k = 6$). Since sample sizes are much larger than in the PSID, we limit our sample to men with 32–40 years of experience to focus on years when wage growth is generally negligible, yet before most men begin retiring.[57] Figure 21 reports the implied 1984–2011 time series for μ_t (normalizing $\mu_{1985} = 1$) when using only DER-based residual earnings. To identify μ_t over earlier years, we combine DER- and SER-based residuals, using the latter only as lagged instruments. Both sets of estimates are very similar to analogous PSID-based estimates reported in Figure 6.[58]

Occupational stayers. We next explore growth in skill returns for occupation-stayers as in Section 5. Here, we must limit our sample to those observed in one of the SIPP panels during years $t - 1$ and t, since only the survey data contains occupation information. We estimate skill return growth for (i) those remaining in the same occupation, (ii) those remaining in a cognitive occupation, and (iii) those remaining

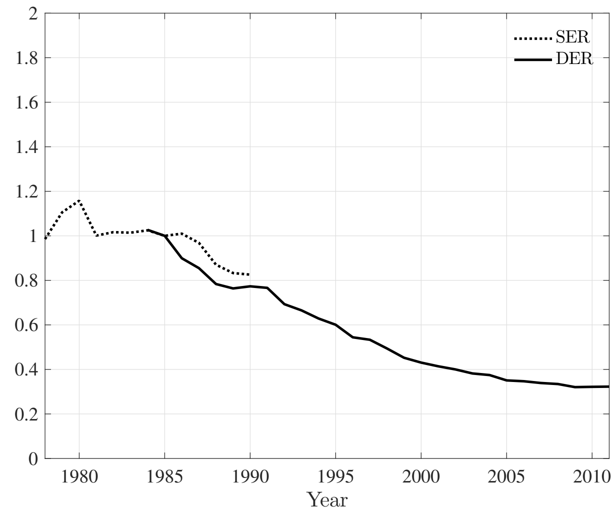
[56] See Moffitt et al. (2022) for a useful effort to reconcile disparate findings across data sources.

[57] Preliminary results were similar for broader experience ranges like those used in the PSID.

[58] See Appendix Tables G-2 and G-3 for the estimates, standard errors, and sample sizes when using the SER and DER earnings residuals as instruments.



(a) Non-College



(b) College

Figure 21: μ_t Implied by IV Estimates (instrument: w_{t-7}), Experience 32–40 in t (SIPP/W-2)

in a routine occupation during years $t-1$ and t.[59] Given the timing of SIPP panels and sample sizes, we estimate annual skill return growth rates for two separate periods: 1991–1999 and 2002–2011. Table 7 reports these IV results using w_{t-7} as instruments, again focusing on men with 32–40 years of experience. The first two columns suggest that skill returns fell by about 2% per year over the 1990s and 2000s, consistent with earlier estimates. The remaining columns suggest that skill returns were fairly stable for cognitive occupations but declined by 3.7–4.9% per year for routine occupations. While we cannot reject equality of skill return growth rates (within periods) across the two occupation groups,[60] stronger declines in routine occupations could be driven by forces related to routine-biased technical change (Autor and Dorn, 2013).

7 Conclusions

Economists have devoted considerable effort to understand the underlying causes of rising residual wage inequality over the past few decades. Most studies have relied on repeated cross-sectional data on wages with a few recent studies incorporating additional measures of worker skills or job tasks. While these efforts have yielded important insights and motivated robust theoretical literatures, they typically assume that distributions of skills or early-career skill growth have remained stable across cohorts born decades

[59] Occupations are based on 24 categories created by the Census Bureau. As with the PSID, we consider routine and cognitive occupation groupings. See Appendix G for details.

[60] The difference in return growth across occupation groups is 0.036 (SE=0.031) for 1991–1999 and 0.046 (SE=0.024) for 2002–2011.

Table 7: 2SLS estimates of $\Delta\mu_t/\mu_{t-1}$ for occupational stayers with Experience 32–40 in t (SIPP/W-2)

	Same occupation		Cognitive occupations		Routine occupations	
	1991–1999	2002–2011	1991–1999	2002–2011	1991–1999	2002–2011
$\Delta\mu_t/\mu_{t-1}$	-0.021	-0.017	-0.013	0.009	-0.049*	-0.037*
	(0.013)	(0.011)	(0.021)	(0.016)	(0.023)	(0.018)
Observations	8,400	11,000	2,900	4,400	5,200	6,100

Notes: Reports coefficient estimates from 2SLS regression of Δw_t on w_{t-1} using w_{t-7} as an instrument. * denotes significance at 0.05 level. The number of observations is rounded to the nearest 100 due to confidentiality requirements.

apart.

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 IV strategy (that exploits panel date on log wage residuals) can be used to estimate the returns to unobserved skill over time. We test and cannot reject key assumptions, further showing that our main conclusions are robust to relaxing most assumptions. 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. Importantly, none of this requires measures of the tasks workers perform nor direct measures of worker skill levels; although, future work could incorporate such measures (when available) within our framework to relax various assumptions, improve the precision of estimates, and/or identify the full complement of task- or skill-specific returns.

Using survey data on wages from the PSID and administrative earnings records from W2 forms, we show that skill returns for American men were fairly stable or increasing in the 1970s and early-1980s, but then fell sharply over the late-1980s and 1990s (especially among non-college men) before stabilizing again. The decline in returns was offset by a strong increase in the variance of unobserved skills beginning in the early-1980s, driven by rising skill volatility (rather than changes in the dispersion of skills at labor market entry). We also estimate a moderate increase in the variance of transitory non-skill wage innovations during the late-1980s and 1990s for college-educated men, contributing to growth in their residual inequality over that period. These conclusions stand in stark contrast to prevailing views, which attribute rising residual inequality primarily to rising skill returns, despite recent evidence by Castex and Dechter (2014) suggesting that the returns to cognitive skill fell by half between the late-1980s and 2010 (consistent with our estimated declines in skill returns).

Given growing interest in the importance of tasks, occupations, and the multiplicity of skills for recent trends in wage inequality, we extend our analysis to account for heterogeneous pricing of multiple unobserved skills across occupations. Our analysis of PSID data indicates that skill returns fell similarly for men working in routine, cognitive, and social occupations. This finding is consistent with similar changes (or stability) in the skill-intensities of these occupation types, coupled with similar declines in the returns to heterogeneous skills used within those occupations. We find that the substantial decline in log wage residuals among workers in routine relative to cognitive occupations can be attributed to (i) weaker growth in wages paid to similarly skilled workers in routine relative to cognitive occupations, and (ii) substantial (unobserved) skill deterioration among workers in routine relative to cognitive jobs. Our estimates based on administrative W2 earnings records suggest that the returns to skill may have fallen more for workers in routine relative to cognitive occupations; however, we cannot reject that the declines in returns were equal (as estimated in the PSID). Whether skill and wage levels (as suggested by the PSID) or the returns to skill (as suggested by W2 records) fell more strongly within routine occupations, our findings are broadly consistent with some form of routine-biased technical change (Autor and Dorn, 2013).

At the most basic level, our conclusion that skill returns declined in the late-1980s and 1990s is a reflection of the sharp drop in long-autocovariances for log wage residuals during that period (see Figures 3 and 5). These drops are broad-based, evident for young and old, non-college and college workers. They are equally pronounced for occupation-stayers, suggesting that they are not simply explained by shifts in the occupation structure or by an increase in occupational switching (Kambourov and Manovskii, 2008). We interpret these changes through the lens of the canonical wage function for unobserved skills introduced by Juhn, Murphy, and Pierce (1993); however, we make no attempt to explain why unobserved skill returns fell over a period when returns to education rose.[61] Just as economic theories developed to explain long-term growth in the returns to unobserved skills, motivated by earlier studies like Juhn, Murphy, and Pierce (1993), we hope that our findings spur new thinking on this issue.

Equally important, our results suggest that more attention be devoted to understanding the dramatic increase in unobserved skill volatility. This may simply reflect a different type of technological change – one characterized by the frequent introduction of new tasks that displace others (e.g., Andolfatto and Smith, 2001; Acemoglu and Restrepo, 2018). Defining workers' skill levels by the most productive task(s) they can perform, this type of technological change would generate growing volatility in skills over the lifecycle (or economic turbulence as in Ljungqvist and Sargent (1998)), which could, in turn, reduce skill returns (see, e.g., Lochner, Park, and Shin, 2018). Growing knowledge/task specialization in the workforce is likely to further exacerbate these forces. An alternative explanation may be that more able workers are simply more capable of learning and adapting to new tasks (Nelson and Phelps, 1966), which would imply

[61] Results in Juhn, Murphy, and Pierce (1993) raised an alternative challenge: why did the returns to unobserved skills rise well the returns to education fell in the late-1970s.

greater variation in lifecycle wage growth during periods of rapid innovation.[62] Finally, if firms possess imperfect information about workers' skills, our estimated "skill distributions" would instead reflect the distributions of beliefs about worker skills. Thus, our estimates may also reflect changes in firms' abilities to identify (and reward) workers' skill levels over their careers (e.g., see Lemieux, MacLeod, and Parent, 2009; Jovanovic, 2014).

[62] See Section 3.2 of Hornstein, Krusell, and Violante (2005) for a survey of theory and evidence on this view of technological change and skills.

References

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

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2):251–333.

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.

Acemoglu, Daron and Jonas Loebbing. 2022. "Automation and polarization." NBER Working Paper No. 30528.

Acemoglu, Daron and Pascual Restrepo. 2018. "The race between man and machine: Implications of technology for growth, factor shares, and employment." *American Economic Review* 108 (6):1488–1542.

———. 2022. "Tasks, automation, and the rise in US wage inequality." *Econometrica* 90 (5):1973–2016.

Adda, Jerome and Christian Dustmann. 2023. "Sources of wage growth." *Journal of Political Economy* 131 (2):456–503.

Altonji, Joseph G, Prashant Bharadwaj, and Fabian Lange. 2012. "Changes in the characteristics of American youth: Implications for adult outcomes." *Journal of Labor Economics* 30 (4):783–828.

Andolfatto, David and Eric Smith. 2001. "Distributional Dynamics Following a Technological Revolution." *Canadian Journal of Economics* 34 (3):739–759.

Ashworth, Jared, V Joseph Hotz, Arnaud Maurel, and Tyler Ransom. 2021. "Changes across cohorts in wage returns to schooling and early work experiences." *Journal of Labor Economics* 39 (4):931–964.

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):1553–1597.

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, 1976-1992: Evidence from Longitudinal Income Tax Records." *Journal of Labor Economics* 21 (2):289–321.

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

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

Benedetto, Gary, Jordan C. Stanley, and Evan Totty. 2018. "The Creation and Use of the SIPP Synthetic Beta v7.0." Working paper, U.S. Census Bureau.

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.

Böhm, Michael J. 2020. "The price of polarization: Estimating task prices under routine-biased technical change." *Quantitative Economics* 11 (2):761–799.

Böhm, Michael J., Hans-Martin von Gaudecker, and Felix Schran. 2024. "Occupation Growth, Skill Prices, and Wage Inequality." *Journal of Labor Economics* 42 (1):201–243.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A Distributional Framework for Matched Employer Employee Data." *Econometrica* 87 (3):699–739.

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.

Bound, John and George Johnson. 1992. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *The American Economic Review* 82 (3):371–392.

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

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, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics* 128 (3):967–1015.

Card, David and Thomas Lemieux. 1994. "Changing Wage Structure and Black-White Wage Differentials." *American Economic Review* 84 (2):29–33.

———. 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.

Cavounidis, Costas, Vittoria Dicandia, Kevin Lang, and Raghav Malhotra. 2021. "The evolution of skill use within and between jobs." NBER Working Paper No. 29302.

Cavounidis, Costas and Kevin Lang. 2020. "Ben-porath meets Lazear: Microfoundations for dynamic skill formation." *Journal of Political Economy* 128 (4):1405–1435.

Cawley, John, James Heckman, and Edward Vytlacil. 2001. "Three observations on wages and measured cognitive ability." *Labour Economics* 8 (4):419–442.

Chay, Kenneth and David S Lee. 2000. "Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black–White Wage Differentials." *Journal of Econometrics* 99 (1):1–38.

Cortes, Guido Matias. 2016. "Where Have the Middle-Wage Workers Gone? A Study of Polarization Using Panel Data." *Journal of Labor Economics* 34 (1):63–105.

Cortes, Guido Matias, Nir Jaimovich, and Henry E Siu. 2023. "The growing importance of social tasks in high-paying occupations: Implications for sorting." *Journal of Human Resources* 58 (5):1429–1451.

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 U.S. Tax Returns." *Brookings Papers on Economic Activity* 2013 (1):67–142.

Deming, David. 2017. “The Growing Importance of Social Skills in the Labor Market.” *Quarterly Journal of Economics* 132 (4):1593–1640.

Edin, Per-Anders, Peter Fredriksson, Martin Nybom, and Björn Öckert. 2022. “The rising return to noncognitive skill.” *American Economic Journal: Applied Economics* 14 (2):78–100.

Engbom, Niklas, Christian Moser, and Jan Sauermann. 2023. “Firm pay dynamics.” *Journal of Econometrics* 233 (2):396–423.

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

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.

Gathmann, Christina and Uta Schönberg. 2010. “How general is human capital? A task-based approach.” *Journal of Labor Economics* 28 (1):1–49.

Gervais, Martin, Nir Jaimovich, Henry E. Siu, and Yaniv Yedid-Levi. 2016. “What should I be when I grow up? Occupations and unemployment over the life cycle.” *Journal of Monetary Economics* 83:54–70.

Gola, Pawe l. 2021. “Supply and demand in a two-sector matching model.” *Journal of Political Economy* 129 (3):940–978.

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, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer. 2020. “Multidimensional skill mismatch.” *American Economic Journal: Macroeconomics* 12 (1):210–244.

Haider, Steven J. 2001. “Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991.” *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, 1967-2006.” *Review of Economic Dynamics* 13 (1):15–51.

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 and Jose Scheinkman. 1987. "The importance of bundling in a Gorman-Lancaster model of earnings." *Review of Economic Studies* 54 (2):243–255.

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.

Heckman, James J and Guilherme Sedlacek. 1985. "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market." *Journal of Political Economy* 93 (6):1077–1125.

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

Hornstein, Andreas, Per Krusell, and Giovanni L Violante. 2005. "The Effects of Technical Change on Labor Market Inequalities." In *Handbook of Economic Growth*, vol. 1B, edited by Philippe Aghion and Steven N Durlauf, chap. 20. Elsevier, 1275–1370.

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

Jovanovic, Boyan. 2014. "Misallocation and Growth." *American Economic Review* 104 (4):1149–1171.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101 (3):410–442.

Kambourov, Gueorgui and Iourii Manovskii. 2008. "Rising occupational and industry mobility in the United States: 1968–97." *International Economic Review* 49 (1):41–79.

———. 2009. "Occupational mobility and wage inequality." *Review of Economic Studies* 76 (2):731–759.

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.

Keane, Michael P and Kenneth I Wolpin. 1997. "The career decisions of young men." *Journal of Political Economy* 105 (3):473–522.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury. 2023. “Do firm effects drift? Evidence from Washington administrative data.” *Journal of Econometrics* 233 (2):375–395.

Lazear, Edward P. 2009. “Firm-specific human capital: A skill-weights approach.” *Journal of political economy* 117 (5):914–940.

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

Lemieux, Thomas, W Bentley MacLeod, and Daniel Parent. 2009. “Performance Pay and Wage Inequality.” *The Quarterly Journal of Economics* 124 (1):1–49.

Lillard, Lee A and Yoram Weiss. 1979. “Components of variation in panel earnings data: American scientists 1960-70.” *Econometrica* 47 (2):437–454.

Lindenlaub, Ilse. 2017. “Sorting multidimensional types: Theory and application.” *Review of Economic Studies* 84 (2):718–789.

Lindenlaub, Ilse and Fabien Postel-Vinay. 2023. “Multidimensional sorting under random search.” *Journal of Political Economy* 131 (12):3497–3539.

Lise, Jeremy and Fabien Postel-Vinay. 2020. “Multidimensional skills, sorting, and human capital accumulation.” *American Economic Review* 110 (8):2328–2376.

Ljungqvist, Lars and Thomas Sargent. 1998. “The European Unemployment Dilemma.” *Journal of Political Economy* 106 (3):514–550.

Lochner, Lance, Youngmin Park, and Youngki Shin. 2018. “Wage Dynamics and Returns to Unobserved Skill.” Working Paper 24220, National Bureau of Economic Research.

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.

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

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

Moffitt, Robert, John Abowd, Christopher Bollinger, Michael Carr, Charles Hokayem, Kevin McKinney, Emily Wiemers, Sisi Zhang, and James Ziliak. 2022. “Reconciling trends in US male earnings volatility: Results from survey and administrative data.” *Journal of Business & Economic Statistics* 41 (1):1–11.

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

Moffitt, Robert A and Peter Gottschalk. 2011. "Trends in the covariance structure of earnings in the US: 1969–1987." *Journal of Economic Inequality* 9 (3):439–459.

Moffitt, Robert A. and Peter Gottschalk. 2012. "Trends in the Transitory Variance of Male Earnings Methods and Evidence." *Journal of Human Resources* 47 (1):204–236.

Murnane, Richard, John Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 2 (77):251–266.

Nelson, Richard and Edmund Phelps. 1966. "Investment in Humans, Technological Diffusion, and Economic Growth." *American Economic Review* 56 (1/2):69–75.

Papageorgiou, Theodore. 2014. "Learning your comparative advantages." *Review of Economic Studies* 81 (3):1263–1295.

Reeder, Lori B., Jordan C. Stanley, and Lars Vilhuber. 2018. "Codebook for the SIPP Synthetic Beta 7.0 [Codebook file]." Cornell Institute for Social and Economic Research and Labor Dynamics Institute [distributor]. Cornell University, Ithaca, NY.

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford Economic Papers* 3 (2):135–146.

Roys, Nicolas A and Christopher R Taber. 2022. "Skill Prices, Occupations, and Changes in the Wage Structure for Low Skilled Men." Working Paper 26453, National Bureau of Economic Research.

Sabelhaus, John and Jae Song. 2010. "The great moderation in micro labor earnings." *Journal of Monetary Economics* 57 (4):391–403.

Sanders, Carl and Christopher Taber. 2012. "Life-cycle wage growth and heterogeneous human capital." *Annual Review of Economics* 4 (1):399–425.

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

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.

Taber, Christopher and Rune Vejlin. 2020. "Estimation of a Roy/search/compensating differential model of the labor market." *Econometrica* 88 (3):1031–1069.

Tinbergen, Jan. 1956. "On the Theory of Income Distribution." *Weltwirtschaftliches Archiv* 77 (2):155–173.

U.S. Census Bureau. 2018. "SIPP Synthetic Beta: Version 7.0 [Computer file]." Cornell University, Synthetic Data Server [distributor], Ithaca, NY.

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

Yamaguchi, Shintaro. 2012. "Tasks and heterogeneous human capital." *Journal of Labor Economics* 30 (1):1–53.

———. 2018. "Changes in returns to task-specific skills and gender wage gap." *Journal of Human Resources* 53 (1):32–70.

Appendix

A Prior Assumptions in the Literature

A.1 Juhn, Murphy, and Pierce (1993)

Let c reflect the year an individual enters the labor market and $e = t - c$ labor market experience. Then,

$$\mathrm{Var}(w_t|c) = \mu_t^2 \mathrm{Var}(\theta_t|c) + \mathrm{Var}(\varepsilon_t|c) = \mu_t^2 \left[\mathrm{Var}(\theta_c|c) + \sum_{\tau=c+1}^{t} \mathrm{Var}(\Delta\theta_\tau|c)\right] + \mathrm{Var}(\varepsilon_t|c).$$

where the second equality holds by assuming $\mathrm{Cov}(\Delta\theta_t, \theta_{t'}|c) = 0$ for $t \geq t' + 1$.

The period t to $t + \ell$ time difference for this variance for a given cohort can be written as follows:

$$\begin{aligned}\Delta_c(t, \ell) &\equiv \mathrm{Var}(w_{t+\ell}|c) - \mathrm{Var}(w_t|c) \\ &= \left(\mu_{t+\ell}^2 - \mu_t^2\right)\left[\mathrm{Var}(\theta_c|c) + \sum_{\tau=c+1}^{t} \mathrm{Var}(\Delta\theta_\tau|c)\right] + \mu_{t+\ell}^2 \sum_{\tau=t+1}^{t+\ell} \mathrm{Var}(\Delta\theta_\tau|c) + \left[\mathrm{Var}(\varepsilon_{t+\ell}|c) - \mathrm{Var}(\varepsilon_t|c)\right].\end{aligned}$$

Next, consider the time difference for the residual variance following an experience group over time (assuming $c + \ell < t$):

$$\begin{aligned}\Delta_e(t, \ell) \equiv{}& \mathrm{Var}(w_{t+\ell}|c + \ell) - \mathrm{Var}(w_t|c) \\ ={}& \left(\mu_{t+\ell}^2 - \mu_t^2\right)\left[\mathrm{Var}(\theta_c|c) + \sum_{\tau=c+1}^{t} \mathrm{Var}(\Delta\theta_\tau|c)\right] \\ &+ \mu_{t+\ell}^2\left[\mathrm{Var}(\theta_{c+\ell}|c + \ell) + \sum_{\tau=c+\ell+1}^{t+\ell} \mathrm{Var}(\Delta\theta_\tau|c + \ell) - \mathrm{Var}(\theta_c|c) - \sum_{\tau=c+1}^{t} \mathrm{Var}(\Delta\theta_\tau|c)\right] \\ &+ \left[\mathrm{Var}(\varepsilon_{t+\ell}|c + \ell) - \mathrm{Var}(\varepsilon_t|c)\right].\end{aligned}$$

To simplify the comparison between cohort- and experienced-based growth in residual inequality, assume that shocks only depend on time and not cohort/experience: $\mathrm{Var}(\varepsilon_t|c) = \mathrm{Var}(\varepsilon_t)$ and $\mathrm{Var}(\Delta\theta_t|c) = \mathrm{Var}(\Delta\theta_t)$ for all c, t. In this case, we have (for $c + \ell < t$):

$$\Delta_c(t, \ell) = \left(\mu_{t+\ell}^2 - \mu_t^2\right)\left[\mathrm{Var}(\theta_c|c) + \sum_{\tau=c+1}^{t} \mathrm{Var}(\Delta\theta_\tau)\right] + \mu_{t+\ell}^2 \sum_{\tau=t+1}^{t+\ell} \mathrm{Var}(\Delta\theta_\tau) + \left[\mathrm{Var}(\varepsilon_{t+\ell}) - \mathrm{Var}(\varepsilon_t)\right], \tag{26}$$

and

$$
\begin{aligned}
&\Delta_e(t,\ell) - \Delta_c(t,\ell) \\
=&\mu_{t+\ell}^2 \left[\mathrm{Var}(\theta_{c+\ell}|c+\ell) - \mathrm{Var}(\theta_c|c) + \sum_{\tau=c+\ell+1}^{t+\ell} \mathrm{Var}(\Delta\theta_\tau) - \sum_{\tau=c+1}^{t} \mathrm{Var}(\Delta\theta_\tau) - \sum_{\tau=t+1}^{t+\ell} \mathrm{Var}(\Delta\theta_\tau) \right] \\
=&\mu_{t+\ell}^2 \left[\mathrm{Var}(\theta_{c+\ell}|c+\ell) - \mathrm{Var}(\theta_c|c) - \sum_{\tau=c+1}^{c+\ell} \mathrm{Var}(\Delta\theta_\tau) \right] \\
=&\mu_{t+\ell}^2 \left[\mathrm{Var}(\theta_{c+\ell}|c+\ell) - \mathrm{Var}(\theta_{c+\ell}|c) \right]
\end{aligned}
$$

As discussed in Juhn, Murphy, and Pierce (1993), equation (26) shows that the change in variances over time for a given cohort incorporates both time effects and experience effects. The experience effects are reflected in the accumulation of permanent skill shocks from $t+1$ to $t+\ell$ (second term), while the time effects reflect changes in skill returns (first term) *and* in non-skill transitory shocks (third term). The evolution of variances over time for a given experience group includes the same three effects plus a fourth reflecting the difference the variance of skill levels between the cohorts as of the later time period. This is important, since it suggests that similar time patterns for residual variances obtained by following cohorts or experience groupings (i.e., $\Delta_e(t,\ell) \approx \Delta_c(t,\ell)$) implies that there is little variation across cohorts in early skill levels (i.e., $\mathrm{Var}(\theta_{c+\ell}|c+\ell) \approx \mathrm{Var}(\theta_{c+\ell}|c)$). This would be the case if the variance of initial skill levels were identical across cohorts ($\mathrm{Var}(\theta_{c+\ell}|c+\ell) = \mathrm{Var}(\theta_c|c)$) and if there were no early skill shocks over the first $\ell+1$ years of working careers. Alternatively, growth in the variance of initial skills across cohorts could offset growth in the variance of skills accumulated via labor market experience.

In the absence of initial cohort differences and early career skill shocks, changes in the variance of residuals should be the same over time whether we follow cohorts or experience groups: $\Delta_e(t,\ell) = \Delta_c(t,\ell)$. Put another way, we should observe similar growth over time in the variance when following cohorts or experience groups regardless of whether that growth is due to an increase in skill returns (i.e., the first term in equation (26)), the existence of skill growth shocks (i.e., the second term in equation (26)), or growth in transitory non-skill shocks (i.e., the third term in equation (26)). Thus, comparing growth in the variance of residuals for given cohorts or experience groups (as in Juhn, Murphy, and Pierce (1993)) is not directly informative about changes in the returns to skill unless there are no skill shocks and the variance of non-skill shocks is time-invariant. Stated differently, $\Delta_e(t,\ell) = \Delta_c(t,\ell) > 0$ is consistent with growth in skill returns, permanent skill shocks, or growth in the variance of non-skill shocks.

Finally, if cohorts are initially identical (i.e., $\mathrm{Var}(\theta_c|c)$ does not depend on c) and shocks depend only on time, then the variance of residuals will grow less (or decrease more) over time when following an experience group than when following a cohort if skill growth shocks are important early in the lifecycle (i.e., $\mathrm{Var}(\Delta\theta_\tau) > 0$ for $\tau = c+1, ..., c+\ell$).

A.2 Lemieux (2006)

We use data from the Health and Retirement Study (HRS), described in Appendix F, to test Lemieux (2006)'s assumption. We first residualize the word recall scores by regressing them on indicators of race, education, experience, and birth year. Next, we regress the squared residuals of test scores on the indicators of race, education, experience, and calendar year, and jointly test whether the estimated coefficients on year indicators are identical (or jointly equal to zero excluding the base year 1996).

Table A-1 reports the p-values of the Wald tests conducted on the full sample, and college- and non-college subsamples. Since all p-values are smaller than 0.05, we reject the hypothesis that the variance of unobserved skill stays constant over time at the 5% significance level.

Table A-1: Wald test p-values

Variables	All Men	Non-College	College
Year	0.0000	0.0003	0.0000

A.3 Castex and Dechter (2014)

We consider regression log wage residuals in period $t+\ell$ on lagged test score residuals, $\tilde{T}_{j,t}$, where we explicitly allow for different test measurements denoted by j. This yields

$$\hat{\beta}_{j,t,t+\ell} \xrightarrow{p} \frac{\mathrm{Cov}(w_{t+\ell}, \tilde{T}_{j,t})}{\mathrm{Var}(\tilde{T}_{j,t})} = \frac{\mu_{t+\ell}}{\tau_j} \underbrace{\left[1 + \frac{\mathrm{Cov}(\theta_{t+\ell} - \theta_t, \theta_t)}{\mathrm{Var}(\theta_t)}\right]}_{\text{Skill Dynamics } (SD_{t,t+\ell})} \underbrace{\left[\frac{\tau_j^2\,\mathrm{Var}(\theta_t)}{\tau_j^2\,\mathrm{Var}(\theta_t) + \mathrm{Var}(\eta_{j,t})}\right]}_{\text{Test Reliability Ratio } (R_{j,t})},$$

The ratio of these estimators (using the same test measurement j) for two different cohorts observed in years t and t', respectively, yields the following:

$$\frac{\hat{\beta}_{j,t,t+\ell}}{\hat{\beta}_{j,t',t'+\ell}} \xrightarrow{p} \frac{\mu_{t+\ell}}{\mu_{t'+\ell}} \left[\frac{SD_{t,t+\ell}}{SD_{t',t'+\ell}}\right] \left[\frac{R_{j,t}}{R_{j,t'}}\right]$$

Growth in skill returns is biased when skill dynamics or the reliability of measurements vary across cohorts. If test measurement error is time-invariant, i.e., $\mathrm{Var}(\eta_{j,t}) = \sigma_j^2$, then the reliability ratios will differ if and only if the variance of skills differs across the cohorts (i.e., $\mathrm{Var}(\theta_t) \neq \mathrm{Var}(\theta_{t'})$).

Using different test measurements across cohorts. Notice that the ratio of estimators for two different cohorts in t and t' using different measurements j and j' yields the following:

$$\frac{\hat{\beta}_{j,t,t+\ell}}{\hat{\beta}_{j',t',t'+\ell}} \xrightarrow{p} \frac{\mu_{t+\ell}}{\mu_{t'+\ell}} \left[\frac{SD_{t,t+\ell}}{SD_{t',t'+\ell}}\right] \left[\frac{R_{j,t}}{R_{j,t'}}\right] \left[\frac{\tau_{j'}}{\tau_j}\right] \left[\frac{R_{j,t'}}{R_{j',t'}}\right],$$

where additional bias arises due to differences in the test score measurement quality as determined by $\tau_j/\tau_{j'}$ and the reliability ratio, $R_{j,t'}/R_{j',t'}$.

Re-scaling different test measurements across cohorts. Deming (2017) scales measurements by their standard deviations, $\sigma_{\tilde{T}_j} = \sqrt{\tau_j^2 \operatorname{Var}(\theta_t) + \operatorname{Var}(\eta_{j,t})}$, before regressing log wage residuals on test score residuals. Denote these regression coefficients as

$$\begin{aligned}
\tilde{\beta}_{j,t,t+\ell} &\xrightarrow{p} \frac{\operatorname{Cov}(w_{t+\ell}, \tilde{T}_{j,t}/\sigma_{\tilde{T}_j})}{\operatorname{Var}(\tilde{T}_{j,t}/\sigma_{\tilde{T}_j})} \\
&= \frac{\mu_{t+\ell}}{\tau_j} SD_{t,t+\ell} R_{j,t} \sigma_{\tilde{T}_j} \\
&= \frac{\mu_{t+\ell}}{\tau_j} \left[\frac{\operatorname{Cov}(\theta_{t+\ell}, \theta_t)}{\operatorname{Var}(\theta_t)}\right] \left[\frac{\tau_j^2 \operatorname{Var}(\theta_t)}{\tau_j^2 \operatorname{Var}(\theta_t) + \operatorname{Var}(\eta_{j,t})}\right] \sqrt{\tau_j^2 \operatorname{Var}(\theta_t) + \operatorname{Var}(\eta_{j,t})} \\
&= \frac{\mu_{t+\ell}}{\tau_j} \left[\frac{\operatorname{Cov}(\theta_{t+\ell}, \theta_t)}{\operatorname{Var}(\theta_t)}\right] \left[\frac{\tau_j^2 \operatorname{Var}(\theta_t)}{\sqrt{\tau_j^2 \operatorname{Var}(\theta_t) + \operatorname{Var}(\eta_{j,t})}}\right] \\
&= \mu_{t+\ell}\, \tau_j \left[\frac{\operatorname{Cov}(\theta_{t+\ell}, \theta_t)}{\sqrt{\tau_j^2 \operatorname{Var}(\theta_t) + \operatorname{Var}(\eta_{j,t})}}\right].
\end{aligned}$$

Clearly, this re-scaling of measurements will not help eliminate any biases for $\mu_{t+\ell}/\mu_{t'+\ell}$ when taking the ratio $\tilde{\beta}_{j,t,t+\ell}/\tilde{\beta}_{j',t',t'+\ell}$.

No measurement error. In the absence of measurement error in test scores, we have

$$\hat{\beta}_{j,t,t+\ell} \xrightarrow{p} \frac{\mu_{t+\ell}}{\tau_j} SD_{t,t+\ell} \qquad \text{and} \qquad \tilde{\beta}_{j,t,t+\ell} \xrightarrow{p} \mu_{t+\ell}\, SD_{t,t+\ell} \sqrt{\operatorname{Var}(\theta_t)},$$

which implies

$$\frac{\hat{\beta}_{j,t,t+\ell}}{\hat{\beta}_{j',t',t'+\ell}} \xrightarrow{p} \frac{\mu_{t+\ell}}{\mu_{t'+\ell}} \left[\frac{\tau_{j'}}{\tau_j}\right] \left[\frac{SD_{t,t+\ell}}{SD_{t',t'+\ell}}\right] \qquad \text{and} \qquad \frac{\tilde{\beta}_{j,t,t+\ell}}{\tilde{\beta}_{j',t',t'+\ell}} \xrightarrow{p} \frac{\mu_{t+\ell}}{\mu_{t'+\ell}} \left[\frac{SD_{t,t+\ell}}{SD_{t',t'+\ell}}\right] \left[\frac{\sqrt{\mathrm{Var}(\theta_t)}}{\sqrt{\mathrm{Var}(\theta_{t'})}}\right].$$

This highlights that re-scaling does not help in addressing the challenge that without the same measure over time, it is impossible to sort out changes in skill variation across cohorts from the “strength” of skill measurements used for the different cohorts.

B Early Occupational Experiences in NLSY

We use the data provided by Castex and Dechter (2014) to calculate the fraction of years each individual in the NLSY79 and NLSY97 worked in different occupations over ages 17–26. We restrict the sample to male respondents who took the Armed Forces Vocational Aptitude Battery (ASVAB) tests between ages 16 and 17.5, so the skill measurements are comparable. The AFQT test is based on four ASVAB subtests: arithemtic reasoning, mathematics knowledge, word knowledge, and paragraph comprehension. Occupations are coded based on the current (or most recent) job at the time of each interview.

Table B-1 reports the average fraction of years individuals reported working in 6 different occupation categories over ages 17–26. Figure B-1 reports the fraction of years in different occupations separately by AFQT quintile.

Table B-1: Average fraction of years (over ages 17–26) spent working in occupations

NLSY Cohort	Clerical	Farm labor	Manager	Professional	Sales	Service
NLSY79	0.088	0.401	0.043	0.068	0.042	0.154
NLSY97	0.099	0.369	0.053	0.105	0.112	0.162

Notes: Sample sizes are 1,200 in NLSY79 and 1,007 in NLSY97.

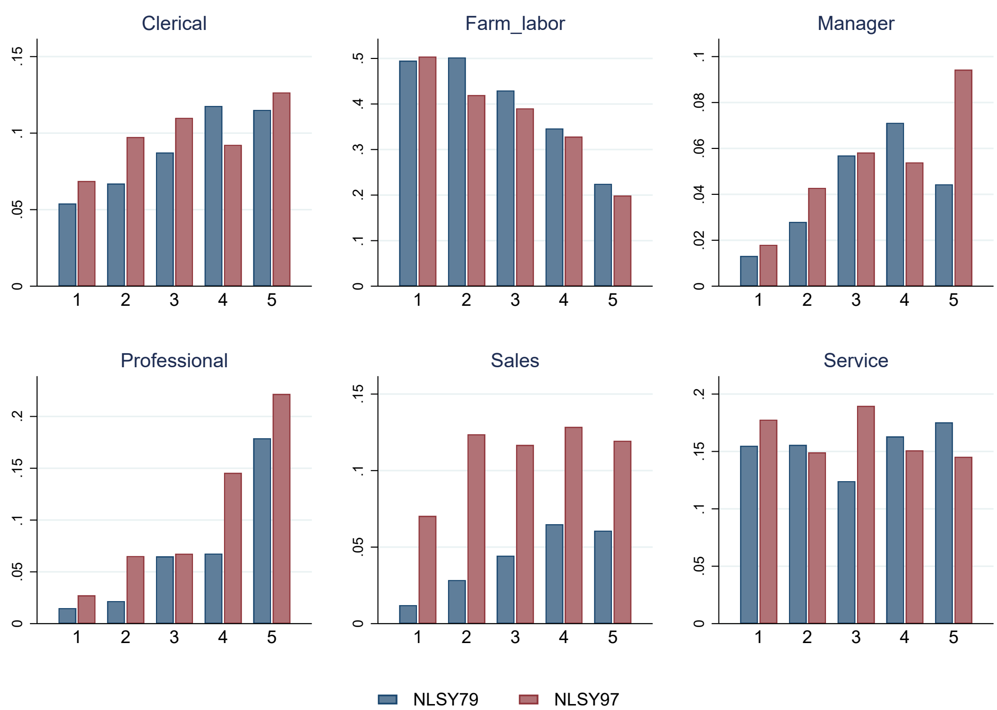


Figure B-1: Average fraction of years (over ages 17–26) spent working in occupations by AFQT quintile

C Identification Results

C.1 Identifying skill returns and distributions of skill and non-skill shocks by cohort

In addition to establishing identification for μ_t over time, the following proposition establishes identification of the variances of unobserved skills, skill growth innovations, and non-skill transitory shocks when these are all allowed to vary by cohort. To achieve the results, we refine Assumption 4 to ensure its applicability to each cohort, c. For the completeness of the arguments, we present the revised condition.

Assumption 4′. *Let c be a cohort denoting the year of labor market entry. For all cohorts, c: (i)* $\mathrm{Cov}(\Delta\theta_t, \theta_{t'}|c) = 0$ *for all $t - t' \geq 1$; (ii)* $\mathrm{Cov}(\theta_t, \varepsilon_{t'}|c) = 0$ *for all t, t'; and (iii) for known $k \geq 1$,* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = 0$ *for all $t - t' \geq k$.*

Proposition 3. *Suppose $\bar{t} - \underline{t} \geq 2k$ for some $k \geq 1$, and Assumption 4′ holds. Then, (i) μ_t is identified for all $t \geq \underline{t} + k$ up to a normalization $\mu_{t^*} = 1$ for some period $t^* \geq \underline{t} + k$, (ii)* $\mathrm{Var}(\theta_t|c)$ *and* $\mathrm{Var}(\varepsilon_t|c)$ *are identified for all (c, t) such that $\underline{t} + k \leq t \leq \bar{t} - k$ and cohort c is observed both in period t and some later period $t' \geq t + k$, and (iii)* $\mathrm{Var}(\Delta\theta_t|c)$ *is identified for all (c, t) such that* $\mathrm{Var}(\theta_t|c)$ *and* $\mathrm{Var}(\theta_{t-1}|c)$ *are identified.*

Proof. **(i) Identification of μ_t.** Without loss of generality, let $t^* = \underline{t} + k$ and $t' = \underline{t}$ so that $t^* - t' \geq k$ and $\mu_{t^*} = \mu_{k+\underline{t}} = 1$. We first proceed with the following derivation:

$$\begin{aligned}
\mathrm{Cov}(w_{t^*}, w_{t'}|c) &= \mu_{t^*}\mu_{t'}\,\mathrm{Cov}(\theta_{t^*}, \theta_{t'}|c) && \text{[Assum 4′ (ii)–(iii)]} \\
&= \mu_{t^*}\mu_{t'}\,\mathrm{Cov}(\theta_{t'} + \Delta\theta_{t'+1} + \cdots + \Delta\theta_{t^*}, \theta_{t'}|c) \\
&= \mu_{t^*}\mu_{t'}\,\mathrm{Var}(\theta_{t'}|c) && \text{[Assum 4′ (i)].}
\end{aligned} \tag{27}$$

Using the IV estimation formula and the normalization $\mu_{t^*} = 1$, we identify μ_{t^*+1}:

$$\begin{aligned}
\frac{\mathrm{Cov}(w_{t^*+1}, w_{t'}|c)}{\mathrm{Cov}(w_{t^*}, w_{t'}|c)} &= \frac{\mu_{t^*+1}\mu_{t'}\,\mathrm{Var}(\theta_{t'}|c)}{\mu_{t^*}\mu_{t'}\,\mathrm{Var}(\theta_{t'}|c)} \\
&= \frac{\mu_{t^*+1}}{\mu_{t^*}} \\
&= \mu_{t^*+1}.
\end{aligned}$$

We identify $\mu_{\underline{t}+k}, \ldots, \mu_{\bar{t}}$ by applying the above arguments recursively.

(ii) Identification of $\mathrm{Var}(\theta_t|c)$ **and** $\mathrm{Var}(\varepsilon_t|c)$**.** For any (t, t') such that $t' - t \geq k$, we can rearrange equation (27) and get the following expression:

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

Since μ_t is now known for $t \geq \underline{t} + k$, we can identify $\operatorname{Var}(\theta_t|c)$ for $t = \underline{t} + k, \ldots, \bar{t} - k$ by varying $(t, t') \in \{(\underline{t} + k, \underline{t} + 2k), \ldots, (\bar{t} - k, \bar{t})\}$. For the same time periods, $\operatorname{Var}(\varepsilon_t|c)$ is identified using:

$$\operatorname{Var}(\varepsilon_t|c) = \operatorname{Var}(w_t|c) - \mu_t^2 \operatorname{Var}(\theta_t|c).$$

(iii) Identification of $\operatorname{Var}(\Delta\theta_t|c)$**.** Assumption 4′(i) implies that

$$\operatorname{Var}(\Delta\theta_t|c) = \operatorname{Var}(\theta_t|c) - \operatorname{Var}(\theta_{t-1}|c).$$

Therefore, $\operatorname{Var}(\Delta\theta_t|c)$ is identified for $t = \underline{t} + k + 1, \ldots, \bar{t} - k$ since $\operatorname{Var}(\theta_t|c)$ is already determined in the previous step.

□

C.2 Identifying early skill returns, μ_t

The bias for $\Delta\mu_t/\mu_{t-1}$ when using future residuals as instruments (equation (16)) presents the key identification challenge for μ_t in early sample periods. Here, we show that the identification results can be extended to earlier years by utilizing additional cohort information.

Proposition 3′. *Suppose that Assumption 4′ holds for two cohorts c and $\tilde{c}$. Furthermore, the following two conditions hold: (a)* $\operatorname{Var}(\theta_{t-1}|c) \neq \operatorname{Var}(\theta_{t-1}|\tilde{c})$ *and (b)* $\operatorname{Var}(\Delta\theta_t|c) = \operatorname{Var}(\Delta\theta_t|\tilde{c})$*. Then, (i) μ_t is identified for all t up to a normalization for some period t^*, (ii)* $\operatorname{Var}(\theta_t|c)$ *and* $\operatorname{Var}(\varepsilon_t|c)$ *are identified for all (c, t) such that $t \leq \bar{t} - k$ and cohort c is observed both in period t and some later period $t' \geq t + k$, and (iii)* $\operatorname{Var}(\Delta\theta_t|c)$ *is identified for for all (c,t) such that* $\operatorname{Var}(\theta_t|c)$ *and* $\operatorname{Var}(\theta_{t-1}|c)$ *are identified.*

Proof. Assumption 4′ and conditions (a) and (b) imply that, for $t' \geq t + k$,

$$\frac{\operatorname{Cov}(w_t, w_{t'}|c) - \operatorname{Cov}(w_t, w_{t'}|\tilde{c})}{\operatorname{Cov}(w_{t-1}, w_{t'}|c) - \operatorname{Cov}(w_{t-1}, w_{t'}|\tilde{c})} = \frac{\mu_t\mu_{t'}\left[\operatorname{Var}(\theta_{t-1}|c) - \operatorname{Var}(\theta_{t-1}|\tilde{c})\right]}{\mu_{t-1}\mu_{t'}\left[\operatorname{Var}(\theta_{t-1}|c) - \operatorname{Var}(\theta_{t-1}|\tilde{c})\right]} = \frac{\mu_t}{\mu_{t-1}}, \tag{28}$$

which identifies μ_t/μ_{t-1} for all $t \leq \bar{t} - k$. We combine the identification result of μ_t in Proposition 3 with this, and the recursive arguments establish the desired results. The remaining identification results follow directly from the previous results in Proposition 3. □

Condition (a) on cohorts 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. Condition (b) 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.

C.3 Identification when skills have a permanent and AR(1) component

Here, we consider the case of Section 3.5.2 in which skills are characterized by a permanent component ψ_i and persistent component $\phi_{i,t}$ that follows an AR(1) process:

$$\begin{aligned} \theta_{i,t} &= \psi_i + \phi_{i,t}, \\ \phi_{i,t} &= \rho_t \phi_{i,t-1} + \nu_{i,t}, \end{aligned}$$

where we exclude the possibility of $\rho_t = 1$ because Assumption 1(i) holds in that case.

For $t' \leq t - k - 1$, Assumption 2 implies the following:

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\Delta \mu_t}{\mu_{t-1}} + \frac{\mu_t}{\mu_{t-1}} \left[\frac{(\rho_t - 1)\hat{\rho}_{t'+1,t-1} \mathrm{Var}(\phi_{t'})}{\mathrm{Var}(\psi) + \hat{\rho}_{t'+1,t-1} \mathrm{Var}(\phi_{t'})} \right],$$

where $\hat{\rho}_{t,t'} \equiv \prod_{j=t}^{t'} \rho_j$. Clearly, the IV estimator is not consistent when $\rho_t \neq 1$. For example, when $\mathrm{Var}(\psi) = 0$, the term in brackets simplifies to $\rho_t - 1$, which implies that the IV estimator converges to $\rho_t \mu_t / \mu_{t-1} - 1$. However, identification of skill returns over time for the case $\rho_t \neq 1$ is still feasible as long as $\mathrm{Var}(\psi) > 0$.

To show identification, it is convenient to re-write the log wage equation as follows:

$$\begin{aligned} w_{i,t} &= \mu_t \psi_i + \tilde{\phi}_{i,t} + \varepsilon_{i,t}, \\ \tilde{\phi}_{i,t} &= \tilde{\rho}_t \tilde{\phi}_{i,t-1} + \tilde{\nu}_{i,t}, \end{aligned}$$

where $\tilde{\phi}_{i,t} \equiv \mu_t \phi_{i,t}$, $\tilde{\rho}_t \equiv \rho_t \mu_t / \mu_{t-1}$, and $\tilde{\nu}_{i,t} \equiv \mu_t \nu_{i,t}$.

Notice that Assumption 2 and the AR(1) process modeled in Equation (14) imply the following orthogonality conditions in terms of the transformed variables:

Assumption 2′. *For all cohorts, c: (i)* $\mathrm{Cov}(\psi, \tilde{\phi}_t|c) = 0$ *for all t; (ii)* $\mathrm{Cov}(\psi, \varepsilon_{t'}|c) = \mathrm{Cov}(\tilde{\phi}_t, \varepsilon_{t'}|c) = 0$ *for all* t, t'*; (iii)* $\mathrm{Cov}(\tilde{\phi}_{t'}, \tilde{\nu}_t|c) = \mathrm{Cov}(\tilde{\nu}_{t'}, \tilde{\nu}_t|c) = 0$ *for all* $t - t' \geq 1$*; (iv) for known* $k \geq 1$, $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = 0$ *for all* $t - t' \geq k$.

Identification of $\tilde{\rho}_t$. Under Assumption 2′, we can construct the following moment condition: for $t' \leq t - k - 1$,

$$\mathrm{Cov}(w_{t'}, w_t|c) - \tilde{\rho}_t \,\mathrm{Cov}(w_{t'}, w_{t-1}|c) = \mu_{t'}(\mu_t - \tilde{\rho}_t\mu_{t-1}) \,\mathrm{Var}(\psi|c). \tag{29}$$

Suppose that there exist two cohorts c and $\tilde{c}$ such that $\mathrm{Var}(\psi|c) > 0$ and $\mathrm{Var}(\psi|\tilde{c}) > 0$. Taking the ratio of (29) for cohort c relative to $\tilde{c}$ yields

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

Similarly, for $t'' \leq t - k - 1$,

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

Combining these two equations yields

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

which becomes

$$A\tilde{\rho}_t^2 + B\tilde{\rho}_t + C = 0, \tag{30}$$

where

$$\begin{aligned}
A &= \mathrm{Cov}(w_{t'}, w_{t-1}|c) \,\mathrm{Cov}(w_{t''}, w_{t-1}|\tilde{c}) - \mathrm{Cov}(w_{t''}, w_{t-1}|c) \,\mathrm{Cov}(w_{t'}, w_{t-1}|\tilde{c}),\\
B &= \mathrm{Cov}(w_{t''}, w_t|c) \,\mathrm{Cov}(w_{t'}, w_{t-1}|\tilde{c}) + \mathrm{Cov}(w_{t''}, w_{t-1}|c) \,\mathrm{Cov}(w_{t'}, w_t|\tilde{c})\\
&\quad - \mathrm{Cov}(w_{t'}, w_t|c) \,\mathrm{Cov}(w_{t''}, w_{t-1}|\tilde{c}) - \mathrm{Cov}(w_{t'}, w_{t-1}|c) \,\mathrm{Cov}(w_{t''}, w_t|\tilde{c}),\\
C &= \mathrm{Cov}(w_{t'}, w_t|c) \,\mathrm{Cov}(w_{t''}, w_t|\tilde{c}) - \mathrm{Cov}(w_{t''}, w_t|c) \,\mathrm{Cov}(w_{t'}, w_t|\tilde{c}).
\end{aligned}$$

We can investigate some cases that equation (30) has a unique solution. First, if $A = 0$ and $B \neq 0$, then the unique solution is

$$\tilde{\rho}_t = -\frac{C}{B}.$$

Second, if $A \neq 0$ and $B^2 - 4AC = 0$, then the unique solutions becomes

$$\tilde{\rho}_t = -\frac{B}{2A}.$$

Notice that there exist other set of sufficient conditions, especially by constructing the additional moment conditions using different cohorts or applying instruments from different time periods.

Identification of μ_t. From Equation (29), we have, for $t \leq t' - k - 1$,

$$\frac{\mathrm{Cov}(w_t, w_{t'}|c) - \tilde{\rho}_{t'}\,\mathrm{Cov}(w_t, w_{t'-1}|c)}{\mathrm{Cov}(w_{t-1}, w_{t'}|c) - \tilde{\rho}_{t'}\,\mathrm{Cov}(w_{t-1}, w_{t'-1}|c)} = \frac{\mu_t}{\mu_{t-1}}. \tag{31}$$

Because $\tilde{\rho}_t$ is identified for all $t \geq \underline{t} + k + 1$, μ_t/μ_{t-1} is identified for all $t \leq \bar{t} - k - 1$.

Equation (29) also implies that, for $t' \leq t - k - 2$,

$$\frac{\mathrm{Cov}(w_{t'}, w_t|c) - \tilde{\rho}_t\,\mathrm{Cov}(w_{t'}, w_{t-1}|c)}{\mathrm{Cov}(w_{t'}, w_{t-1}|c) - \tilde{\rho}_{t-1}\,\mathrm{Cov}(w_{t'}, w_{t-2}|c)} = \frac{\mu_t - \tilde{\rho}_t\mu_{t-1}}{\mu_{t-1} - \tilde{\rho}_{t-1}\mu_{t-2}} = \frac{\mu_{t-1}}{\mu_{t-2}}\,\frac{\left(\frac{\mu_t}{\mu_{t-1}} - \tilde{\rho}_t\right)}{\left(\frac{\mu_{t-1}}{\mu_{t-2}} - \tilde{\rho}_{t-1}\right)}. \tag{32}$$

Because $\tilde{\rho}_t$ is identified for all $t \geq \underline{t} + k + 1$ and μ_t/μ_{t-1} is identified for all $t \leq \bar{t} - k - 1$ based on Equation (31), μ_t/μ_{t-1} for $t \geq \bar{t} - k$ is also identified from Equation (32) as long as $\bar{t} - k - 1 \geq \underline{t} + k + 1$. Therefore, μ_t is identified for all t (up to a normalization $\mu_{t^*} = 1$) if $\bar{t} - \underline{t} \geq 2(k + 1)$.

Identification of ρ_t. $\rho_t = \tilde{\rho}_t\mu_{t-1}/\mu_t$ is identified for $t \geq \underline{t} + k + 1$ because $\tilde{\rho}_t$ is identified for $t \geq \underline{t} + k + 1$ and μ_t is identified for all t.

Identification of $\mathrm{Var}(\psi|c)$. Equation (29) implies $\mathrm{Var}(\psi|c)$ is identified for all cohorts observed in t', $t - 1$, and t such that $t' \leq t - k - 1$.

Identification of $\mathrm{Var}(\phi_t|c)$**.** For $t' \geq t+k$,

$$\mathrm{Cov}(w_t, w_{t'}|c) = \mu_t \mu_{t'} \left[\mathrm{Var}(\psi|c) + \left(\sum_{j=t+1}^{t'} \rho_j \right) \mathrm{Var}(\phi_t|c) \right].$$

Since ρ_t is identified for $t \geq \underline{t}+k+1$, $\mathrm{Var}(\phi_t|c)$ is identified for $t \geq \underline{t}+k$.

Finally, we note that it is straightforward to identify the covariance structure for ε_t from "close" autocovariances given identification of everything else.

C.4 Identification with $\varepsilon_{i,t} \sim$ ARMA$(1,q)$

We demonstrate identification for the model in Sections 2.2 generalized so that the transitory component $\varepsilon_{i,t}$ include an ARMA(1,q) process $\varphi_{i,t}$:

$$\varepsilon_{i,t} = \varphi_{i,t} + \tilde{\varepsilon}_{i,t} \tag{33}$$

$$\varphi_{i,t} = \rho_t \varphi_{i,t-1} + \sum_{j=0}^{\min\{q,t-c_i-1\}} \beta_j \nu_{i,t-j}, \tag{34}$$

where $\beta_0 = 1$. Similar to Assumption 3, we assume that $\nu_{i,t}$ is an *i.i.d.* innovation term, not correlated with any other variables including $\nu_{i,t'}$ for $t' \neq t$ and $\varphi_{i,t'}$ for $t-t' \geq 1$. Without loss of generality, we assume that $k \geq q$. Otherwise, we can redefine $k' = \max\{k,q\}$ and use k' instead of k.

Identification of ρ_t**.** For $t' \geq t+k+1$, we have

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'} - \rho_{t'}\varepsilon_{t'-1}|c) = \mathrm{Cov}\left(\varphi_t + \tilde{\varepsilon}_t, \sum_{j=0}^{\min\{q,t-c-1\}} \beta_j \nu_{t'-j} + (\tilde{\varepsilon}_{t'} - \rho_{t'}\tilde{\varepsilon}_{t'-1}) \Big| c \right) = 0,$$

and therefore

$$\mathrm{Cov}(w_t, w_{t'}|c) - \rho_{t'} \mathrm{Cov}(w_t, w_{t'-1}|c) = \mu_t(\mu_{t'} - \rho_{t'}\mu_{t'-1}) \mathrm{Var}(\theta_t|c). \tag{35}$$

Taking the ratio of (35) for cohort c relative to $\tilde{c}$ yields

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

Similarly, for $t'' \geq t + k + 1$,

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

Combining these two equations yields

$$\frac{\mathrm{Cov}(w_t, w_{t'}|c) - \rho_{t'}\,\mathrm{Cov}(w_t, w_{t'-1}|c)}{\mathrm{Cov}(w_t, w_{t'}|\tilde{c}) - \rho_{t'}\,\mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})} = \frac{\mathrm{Cov}(w_t, w_{t''}|c) - \rho_{t''}\,\mathrm{Cov}(w_t, w_{t''-1}|c)}{\mathrm{Cov}(w_t, w_{t''}|\tilde{c}) - \rho_{t''}\,\mathrm{Cov}(w_t, w_{t''-1}|\tilde{c})}. \tag{36}$$

Equation (36) can be written as

$$A_1 \rho_{t'} \rho_{t''} + B_1 \rho_{t'} + C_1 \rho_{t''} + D_1 = 0, \tag{37}$$

where

$$\begin{aligned}
A_1 &= \mathrm{Cov}(w_t, w_{t'-1}|c)\,\mathrm{Cov}(w_t, w_{t''-1}|\tilde{c}) - \mathrm{Cov}(w_t, w_{t''-1}|c)\,\mathrm{Cov}(w_t, w_{t'-1}|\tilde{c}),\\
B_1 &= \mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})\,\mathrm{Cov}(w_t, w_{t''}|c) - \mathrm{Cov}(w_t, w_{t'-1}|c)\,\mathrm{Cov}(w_t, w_{t''}|\tilde{c}),\\
C_1 &= \mathrm{Cov}(w_t, w_{t'}|\tilde{c})\,\mathrm{Cov}(w_t, w_{t''-1}|c) - \mathrm{Cov}(w_t, w_{t'}|c)\,\mathrm{Cov}(w_t, w_{t''-1}|\tilde{c}),\\
D_1 &= \mathrm{Cov}(w_t, w_{t'}|c)\,\mathrm{Cov}(w_t, w_{t''}|\tilde{c}) - \mathrm{Cov}(w_t, w_{t'}|\tilde{c})\,\mathrm{Cov}(w_t, w_{t''}|c).
\end{aligned}$$

By changing t, c, or $\tilde{c}$, we can also construct an equation

$$A_2 \rho_{t'} \rho_{t''} + B_2 \rho_{t'} + C_2 \rho_{t''} + D_2 = 0, \tag{38}$$

where A_2, B_2, C_2, and D_2 are defined in a similar way.

We investigate some cases that this system of equations (37)–(38) has a unique solution. When $A_1 = A_2 = 0$, it becomes a system of linear equations. If $B_1 C_2 - B_2 C_1 \neq 0$, it has a unique solution

$$\begin{pmatrix} \rho_{t'} \\ \rho_{t''} \end{pmatrix} = -\begin{pmatrix} B_1 & C_1 \\ B_2 & C_2 \end{pmatrix}^{-1} \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}.$$

When $A_1 \neq 0$ and $A_2 \neq 0$, it becomes a set of rectangular hyperbolas. We first rearrange the equations:

$$\begin{aligned}
\left(\rho_{t'} + \frac{C_1}{A_1}\right)\left(\rho_{t''} + \frac{B_1}{A_1}\right) &= \frac{B_1 C_1 - A_1 D_1}{A_1^2}\\
\left(\rho_{t'} + \frac{C_2}{A_2}\right)\left(\rho_{t''} + \frac{B_2}{A_2}\right) &= \frac{B_2 C_2 - A_2 D_2}{A_2^2}.
\end{aligned}$$

If the constants on the right hand side have different signs and the graphs share only one asymptote, it always has a unique solution. Therefore, we conclude that $(\rho_{t'}, \rho_{t''})$ are jointly identified when (i) $(B_1C_1 - A_1D_1)(B_2C_2 - A_2D_2) < 0$; and (ii-1) $C_1/A_1 = C_2/A_2$ and $(B_1/A_1) \neq (B_2/A_2)$ or (ii-2) $C_1/A_1 \neq C_2/A_2$ and $(B_1/A_1) = (B_2/A_2)$. Once we identify $\rho_{t'}$ for some t', we can recover $\rho_{t^\dagger}$ recursively by constructing equation (37) with $t^\dagger$ instead of t'' or with different cohort if necessary. Then, we can identify ρ_t for all $t \geq \underline{t} + k + 1$. Notice that these conditions are sufficient for identification. Additional equations generated by varying t, c, or $\tilde{c}$ can provide a different set of sufficient conditions for identification.

Identification of μ_t. For $t' \geq t+k+1$, suppose that there exists $(c, \tilde{c})$ such that $\mathrm{Var}(\theta_{t-1}|c) \neq \mathrm{Var}(\theta_{t-1}|\tilde{c})$ and $\mathrm{Var}(\Delta\theta_t|c) = \mathrm{Var}(\Delta\theta_t|\tilde{c})$. Then, from Equation (35), we have

$$\begin{aligned}
&\frac{\left[\mathrm{Cov}(w_t, w_{t'}|c) - \rho_{t'}\mathrm{Cov}(w_t, w_{t'-1}|c)\right] - \left[\mathrm{Cov}(w_t, w_{t'}|\tilde{c}) - \rho_{t'}\mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})\right]}{\left[\mathrm{Cov}(w_{t-1}, w_{t'}|c) - \rho_{t'}\mathrm{Cov}(w_{t-1}, w_{t'-1}|c)\right] - \left[\mathrm{Cov}(w_{t-1}, w_{t'}|\tilde{c}) - \rho_{t'}\mathrm{Cov}(w_{t-1}, w_{t'-1}|\tilde{c})\right]} \\
=&\frac{\mu_t(\mu_{t'} - \rho_{t'}\mu_{t'-1})\left[\mathrm{Var}(\theta_{t-1}|c) - \mathrm{Var}(\theta_{t-1}|\tilde{c})\right]}{\mu_{t-1}(\mu_{t'} - \rho_{t'}\mu_{t'-1})\left[\mathrm{Var}(\theta_{t-1}|c) - \mathrm{Var}(\theta_{t-1}|\tilde{c})\right]} \\
=&\frac{\mu_t}{\mu_{t-1}}.
\end{aligned} \tag{39}$$

Since ρ_t is identified for all $t \geq \underline{t} + k + 1$, μ_t/μ_{t-1} is identified for all $t \leq \bar{t} - k - 1$.

Equation (35) also implies that, for $t' \leq t - k - 2$,

$$\frac{\mathrm{Cov}(w_{t'}, w_t|c) - \rho_t\mathrm{Cov}(w_{t'}, w_{t-1}|c)}{\mathrm{Cov}(w_{t'}, w_{t-1}|c) - \rho_{t-1}\mathrm{Cov}(w_{t'}, w_{t-2}|c)} = \frac{\mu_t - \rho_t\mu_{t-1}}{\mu_{t-1} - \rho_{t-1}\mu_{t-2}} = \frac{\mu_{t-1}}{\mu_{t-2}}\frac{\left(\frac{\mu_t}{\mu_{t-1}} - \rho_t\right)}{\left(\frac{\mu_{t-1}}{\mu_{t-2}} - \rho_{t-1}\right)} \tag{40}$$

Because ρ_t is identified for all $t \geq \underline{t} + k + 1$ and μ_t/μ_{t-1} is identified for all $t \leq \bar{t} - k - 1$ based on Equation (39), μ_t/μ_{t-1} for $t \geq \bar{t} - k$ is also identified from Equation (40) as long as $\bar{t} - k - 1 \geq \underline{t} + k + 1$. Therefore, μ_t is identified for all t (up to a normalization $\mu_{t^*} = 1$) if $\bar{t} - \underline{t} \geq 2(k+1)$.

Identification of $\mathrm{Var}(\theta_t|c)$. For $t' \geq t + k + 1$, Equation (35) implies

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

Because ρ_t is identified for all $t \geq \underline{t} + k + 1$ and μ_t is identified for all t, $\mathrm{Var}(\theta_t|c)$ is identified for all $t \leq \bar{t} - k - 1$.

Finally, we note that it is straightforward to identify $\{\beta_j\}$ and $\mathrm{Var}(\nu_t|c)$ from "close" autocovariances given identification of everything else.

C.5 Identification with Heterogeneous Skill Growth Rates

We now demonstrate identification for the model in Section 3.5.1 with systematic heterogeneity in lifecycle skill growth.

Letting ψ_i reflect the initial skill for an individual entering the labor market, the skill growth process (12) implies that the level of unobserved skill for individual i from cohort c_i in year t can be written as

$$\theta_{i,t} = \psi_i + \Lambda_t(c_i)\delta_i + \sum_{j=0}^{t-c_i-1} \nu_{i,t-j}, \tag{41}$$

where $\Lambda_t(c) \equiv \sum_{j=0}^{t-c-1} \lambda_{t-j}(c)$ reflects the accumulated influence of skill growth heterogeneity.

To facilitate an identification analysis, assume that idiosyncratic skill growth shocks $\nu_{i,t}$ are serially uncorrelated and uncorrelated with initial skills and systematic skill growth; however, we make no assumptions about the correlation between heterogeneous skill growth rates and initial skill levels. Consistent with the literature estimating HIP models, we strengthen conditions (ii) and (iii) of Assumption 1 to assume that non-skill shocks are uncorrelated with all skill-related components. Formally, we assume the following, explicitly conditioning on cohorts.

Assumption 7. *For all cohorts, c: (i)* $\mathrm{Cov}(\psi, \nu_t|c) = \mathrm{Cov}(\delta, \nu_t|c) = 0$ *for all t; (ii)* $\mathrm{Cov}(\psi, \varepsilon_{t'}|c) = \mathrm{Cov}(\delta, \varepsilon_{t'}|c) = \mathrm{Cov}(\nu_t, \varepsilon_{t'}|c) = 0$ *for all t, t'; (iii) for known $k \geq 1$,* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = 0$ *for all $t - t' \geq k$.*

Assumption 7 implies that the covariance between skills in periods t and $t' < t$ can be written as

$$\mathrm{Cov}(\theta_{t'}, \theta_t|c) = \mathrm{Var}(\psi|c) + \Lambda_{t'}(c)\Lambda_t(c)\,\mathrm{Var}(\delta|c) + [\Lambda_{t'}(c) + \Lambda_t(c)]\,\mathrm{Cov}(\psi, \delta|c) + \sum_{j=0}^{t'-c-1} \mathrm{Var}(\nu_{t-j}|c).$$

In addition to Assumption 7, we assume that there exists $\overline{e}$ such that $\lambda_t(c) = 0$ for $e = t - c \geq \overline{e}$.

Identification of μ_t. μ_t can be identified based on experienced workers with $\lambda_t(c) = 0$ for which Propositions 3 and 3′ can be applied. First, μ_t/μ_{t-1} for $t \geq \underline{t} + (k+1)$ is identified if there exists some cohort c such that (i) the cohort has experience $e = t - c \geq \overline{e}$ in year t and (ii) the cohort is observed in years $t' \leq t - k - 1$, $t - 1$, and t. Moreover, μ_t/μ_{t-1} for $t \leq k + 1$ is identified if there exist two cohorts c and $\tilde{c}$ such that (i) both cohorts have experience of at least $\overline{e}$ in year t, (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 $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|\tilde{c})$.

Identification of $\lambda_t(c)$**.** By dividing the residual by μ_t, we get

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

and its first difference is

$$\Delta\left(\frac{w_{i,t}}{\mu_t}\right) = \Delta\theta_{i,t} + \Delta\left(\frac{\varepsilon_{i,t}}{\mu_t}\right) = \lambda_t(c_i)\delta_i + \nu_{i,t} + \Delta\left(\frac{\varepsilon_{i,t}}{\mu_t}\right).$$

For $|t' - t| \geq k+1$, $\mathrm{Cov}(\Delta\varepsilon_t, \Delta\varepsilon_{t'}|c) = 0$ and

$$\mathrm{Cov}\left(\Delta\left(\frac{w_t}{\mu_t}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)\Big|c\right) = \mathrm{Cov}(\Delta\theta_t, \Delta\theta_{t'}|c) = \lambda_t(c)\lambda_{t'}(c)\,\mathrm{Var}(\delta|c).$$

Therefore, we can identify changes in $\lambda_t(c)$:

$$\frac{\lambda_t(c)}{\lambda_{t-1}(c)} = \frac{\mathrm{Cov}\left(\Delta\left(\frac{w_t}{\mu_t}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)|c\right)}{\mathrm{Cov}\left(\Delta\left(\frac{w_{t-1}}{\mu_{t-1}}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)|c\right)}, \qquad \forall (t,t') \text{ such that } t - t' \geq k+2 \text{ or } t' - t \geq k+1.$$

Normalizing $\lambda_{t^*(c)}(c) = 1$ for some $t^*(c)$, all $\lambda_t(c)$'s can be identified for which the covariances in (scaled) residual wage changes are observed.

Identification of $\mathrm{Var}(\delta|c)$**.** Once $\lambda_t(c)$'s have been identified, $\mathrm{Var}(\delta|c)$ is identified from

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

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

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

where

$$\mathrm{Cov}(\theta_t, \delta|c) = \mathrm{Cov}(\psi, \delta|c) + \mathrm{Var}(\delta|c) \sum_{j=0}^{t-c-1} \lambda_{t-j}(c). \tag{42}$$

Therefore,

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

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

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

Then,

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

Therefore,

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

Identification of $\mathrm{Var}(\nu_t|c)$**.** Note that

$$\mathrm{Var}(\theta_t|c) = \mathrm{Var}(\theta_{t-1}|c) + \mathrm{Var}(\delta|c)\lambda_t(c)^2 + 2\,\mathrm{Cov}(\theta_{t-1}, \delta|c)\lambda_t(c) + \mathrm{Var}(\nu_t|c),$$

from which $\mathrm{Var}(\nu_t|c)$ is identified once all the other components have been identified.

Finally, we note that it is straightforward to identify $\{\beta_j\}$ and $\mathrm{Var}(\xi_t|c)$ from "close" autocovariances given identification of everything else.

C.6 Identification with Occupations

Identification of μ_t^o**.** With Assumption 5(iii)–(iv), the long autocovariance for log wage residuals for $t - t' \geq k+1$ can be written as follows:

$$\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}, o_{t'}) = \mu_t^{o_t}\left[\mu_{t'}^{o_{t'}}\,\mathrm{Cov}(\theta_t, \theta_{t'}|o_t, o_{t-1}, o_{t'}) + \mathrm{Cov}(\theta_t, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'})\right], \tag{43}$$

$$\mathrm{Cov}(w_{t-1}, w_{t'}|o_t, o_{t-1}, o_{t'}) = \mu_{t-1}^{o_{t-1}}\left[\mu_{t'}^{o_{t'}}\,\mathrm{Cov}(\theta_{t-1}, \theta_{t'}|o_t, o_{t-1}, o_{t'}) + \mathrm{Cov}(\theta_{t-1}, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'})\right]. \tag{44}$$

Moreover, Assumption 5(i)–(ii) imply $\mathrm{Cov}(\theta_t, \theta_{t'}|o_t, o_{t-1}, o_{t'}) = \mathrm{Cov}(\theta_{t-1}, \theta_{t'}|o_t, o_{t-1}, o_{t'})$ and $\mathrm{Cov}(\theta_t, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'}$ $\mathrm{Cov}(\theta_{t-1}, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'})$, so equations (43) and (44) imply equation (18).

IV Estimator without Conditioning on $o_{t'}$**.** Next, we show that μ_t^o is identified based on covariances conditioned only on (o_t, o_{t-1}) when we assume $\mathrm{E}[\Delta\theta_t|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_{t-1}|o_t, o_{t-1}, o_{t'}] =$

0 in addition to Assumption 5. Consider the long autocovariance (43) that is not conditioned on $o_{t'}$:

$$\begin{aligned}\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}) =& \mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}, o_{t'})|o_t, o_{t-1}\right] \\ &+ \mathrm{Cov}\left(\mathrm{E}[w_t|o_t, o_{t-1}, o_{t'}], \mathrm{E}[w_{t'}|o_t, o_{t-1}, o_{t'}]|o_t, o_{t-1}\right).\end{aligned} \tag{45}$$

The second term in equation (45) is

$$\begin{aligned}&\mathrm{Cov}\left(\mathrm{E}[w_t|o_t, o_{t-1}, o_{t'}], \mathrm{E}[w_{t'}|o_t, o_{t-1}, o_{t'}]|o_t, o_{t-1}\right) \\ =& \mathrm{Cov}\left(\gamma_t^{o_t} + \mu_t^{o_t}\,\mathrm{E}[\theta_t|o_t, o_{t-1}, o_{t'}], \mathrm{E}[w_{t'}|o_t, o_{t-1}, o_{t'}]|o_t, o_{t-1}\right) \\ =& \mu_t^{o_t}\,\mathrm{Cov}\left(\mathrm{E}[\theta_t|o_t, o_{t-1}, o_{t'}], \mathrm{E}[w_{t'}|o_t, o_{t-1}, o_{t'}]|o_t, o_{t-1}\right).\end{aligned}$$

where we used the additional assumption $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = 0$.

Thus, the long autocovariances (43) and (44) that are not conditioned on $o_{t'}$ are given by

$$\begin{aligned}\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}) =& \mu_t^{o_t}\Xi_t^{o_t, o_{t-1}}, \\ \mathrm{Cov}(w_{t-1}, w_{t'}|o_t, o_{t-1}) =& \mu_{t-1}^{o_{t-1}}\Xi_{t-1}^{o_t, o_{t-1}},\end{aligned}$$

where

$$\begin{aligned}\Xi_t^{o_t, o_{t-1}} \equiv& \mathrm{E}\left[\mu_{t'}^{o_{t'}}\,\mathrm{Cov}(\theta_t, \theta_{t'}|o_t, o_{t-1}, o_{t'}) + \mathrm{Cov}(\theta_t, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'})|o_t, o_{t-1}\right] \\ &+ \mathrm{Cov}\left(\mathrm{E}[\theta_t|o_t, o_{t-1}, o_{t'}], \mathrm{E}[w_{t'}|o_t, o_{t-1}, o_{t'}]|o_t, o_{t-1}\right).\end{aligned}$$

Assumption 5(i)–(ii) and $\mathrm{E}[\Delta\theta_t|o_t, o_{t-1}, o_{t'}] = 0$ imply $\Xi_t^{o_t, o_{t-1}} = \Xi_{t-1}^{o_t, o_{t-1}}$, so

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'}|o_t, o_{t-1})}{\mathrm{Cov}(w_{t-1}, w_{t'}|o_t, o_{t-1})} = \frac{\mu_t^{o_t} - \mu_{t-1}^{o_{t-1}}}{\mu_{t-1}^{o_{t-1}}}.$$

The IV estimator for stayers in an occupation ($o_t = o_{t-1} = o$) identifies growth in returns to skill in that occupation. Moreover, the IV estimator for occupational switchers ($o_t \neq o_{t-1}$) identifies the differences in the level of returns to skill across occupations, given a normalization $\mu_{t^*}^{o^*} = 1$ for some (t^*, o^*).

Identification of γ_t^o. Given μ_t^o, we show that γ_t^o is identified under the assumptions $\mathrm{E}[\Delta\theta_t|o_t, o_{t-1}] = \mathrm{E}[\varepsilon_t|o_t, o_{t-1}] = \mathrm{E}[\varepsilon_{t-1}|o_t, o_{t-1}] = 0$.

Since μ_t^o is identified, we can use it to scale the average log wage residuals as follows:

$$\frac{\mathrm{E}\left[w_t | o_t, o_{t-1}\right]}{\mu_t^{o_t}} = \frac{\gamma_t^{o_t}}{\mu_t^{o_t}} + \mathrm{E}[\theta_t | o_t, o_{t-1}].$$

Using $\mathrm{E}[\Delta\theta_t | o_t, o_{t-1}] = 0$, the average growth of scaled log wage residual is

$$\frac{\mathrm{E}\left[w_t | o_t, o_{t-1}\right]}{\mu_t^{o_t}} - \frac{\mathrm{E}\left[w_{t-1} | o_t, o_{t-1}\right]}{\mu_{t-1}^{o_{t-1}}} = \frac{\gamma_t^{o_t}}{\mu_t^{o_t}} - \frac{\gamma_{t-1}^{o_{t-1}}}{\mu_{t-1}^{o_{t-1}}}.$$

Therefore, with a normalization $\gamma_{t^*}^{o^*} = 0$ for some (t^*, o^*), $\gamma_t^{o^*}$ for $t \neq t^*$ is identified based on stayers in occupation o^*. On the other hand, γ_t^o for $o \neq o^*$ is identified from occupation switchers.

C.7 Identification with Multiple Skills

Recall that we define $\overline{\theta}_{i,t} \equiv \sum_j \mu_{j,t}\theta_{i,j,t}$. Given Assumption 6, our IV estimator identifies the following:

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\sum_{j=1}^{J} \Delta\mu_{j,t}\,\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j'=1}^{J} \mu_{j',t-1}\,\mathrm{Cov}(\theta_{j',t'}, \overline{\theta}_{t'})}, \quad \text{for } t - t' \geq k + 1,$$

which implies Proposition 2.

Proof of Proposition 2. We consider skill return growth from period t_0 to t, where the text assumes $t_0 = t - 1$. More generally, we require $t' + k \leq t_0 \leq t - 1$. Empirically, we use $t_0 = t - 2$ given the biennial nature of the PSID in later years.

Assumption 6 implies the following:

$$\begin{aligned}
\frac{\operatorname{Cov}(w_t - w_{t_0}, w_{t'})}{\operatorname{Cov}(w_{t_0}, w_{t'})} &= \frac{\sum_{j=1}^{J} (\mu_{j,t} - \mu_{j,t_0}) \operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j=1}^{J} \mu_{j,t_0} \operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})} \\
&= \frac{\sum_{j=1}^{J} \left(\frac{\mu_{j,t} - \mu_{j,t_0}}{\mu_{j,t_0}}\right) \mu_{j,t_0} \operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j=1}^{J} \mu_{j,t_0} \operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})} \\
&= \sum_{j=1}^{J} \omega_{j,t',t_0} \left(\frac{\mu_{j,t} - \mu_{j,t_0}}{\tilde{\mu}_{j,t_0}}\right)
\end{aligned} \tag{46}$$

where the weights,

$$\omega_{j,t',t_0} \equiv \frac{\operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}) \mu_{j,t_0}}{\sum_{j'=1}^{J} \operatorname{Cov}(\theta_{j',t'}, \overline{\theta}_{t'}) \mu_{j',t_0}}.$$

Notice that $\operatorname{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$, implies that $\operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}) \geq 0, \forall j$. Since $\mu_{j,t} \geq 0, \forall j, t$, the weights $\omega_{j,t',t_0} \geq 0$ whenever $\operatorname{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$. Since the weights sum to one, non-negativity of the weights further implies that none exceeds one.

□

We can re-write the weights in terms of linear projections:

$$\omega_{j,t',t_0} = \frac{L(\theta_{j,t'} | \overline{\theta}_{t'}) \mu_{j,t_0}}{\sum_{j'=1}^{J} L(\theta_{j',t'} | \overline{\theta}_{t'}) \tilde{\mu}_{j',t_0}} = \frac{L(\theta_{j,t_0} | \overline{\theta}_{t'}) \mu_{j,t_0}}{\sum_{j'=1}^{J} L(\theta_{j',t_0} | \overline{\theta}_{t'}) \tilde{\mu}_{j',t_0}} = \frac{L(\mu_{j,t_0} \theta_{j,t_0} | \overline{\theta}_{t'})}{\sum_{j'=1}^{J} L(\mu_{j',t_0} \theta_{j',t_0} | \overline{\theta}_{t'})}$$

where $L(a|b) = \operatorname{Cov}(a,b)b/\operatorname{Var}(b)$ is the linear projection of a onto b. The second equality follows from condition (i) of Assumption 6, which implies that $L(\theta_{j,t} | \overline{\theta}_{t'}) = L(\theta_{j,t'} | \overline{\theta}_{t'})$ for all $t \geq t'$. Thus, the weight on growth in returns to skill j depends on the (linearly) predicted rewards from skill j in period t_0, $\mu_{j,t_0} \theta_{j,t_0}$, given total worker productivity in period t', $\overline{\theta}_{t'}$.

Proposition 4. *If Assumption 6 holds, then for all $t - t' \geq k + 1$, the IV estimator identifies growth in the* *weighted-average return to skills,* $m_{t,t'} = \sum_{j=1}^{J} \varphi_{j,t'}\mu_{j,t}$*:*

$$\frac{\mathrm{Cov}(w_t - w_{t_0}, w_{t'})}{\mathrm{Cov}(w_{t_0}, w_{t'})} = \frac{m_{t,t'} - m_{t_0,t'}}{m_{t_0,t'}}$$

with weights

$$\varphi_{j,t'} \equiv \frac{\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j'=1}^{J} \mathrm{Cov}(\theta_{j',t'}, \overline{\theta}_{t'})}, \quad \textit{for } j = 1, ..., J. \tag{47}$$

If $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$*, then the weights* $\varphi_{j,t'} \in [0, 1], \forall j$.

Proof of Proposition 4. Using the definitions of $m_{t,t'}$ and $\varphi_{j,t'}$, growth in the weighted-average return to skills can be written as

$$\frac{m_{t,t'} - m_{t_0,t'}}{m_{t_0,t'}} = \frac{\sum_{j=1}^{J} \varphi_{j,t'}(\mu_{j,t} - \mu_{j,t_0})}{\sum_{j=1}^{J} \varphi_{j,t'}\mu_{j,t_0}} = \frac{\sum_{j=1}^{J} \mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})(\mu_{j,t} - \mu_{j,t_0})}{\sum_{j'=1}^{J} \mathrm{Cov}(\theta_{j',t'}, \overline{\theta}_{t'})\mu_{j',t_0}} = \frac{\mathrm{Cov}(w_t - w_{t_0}, w_{t'})}{\mathrm{Cov}(w_{t_0}, w_{t'})},$$

where the last equality reflects equation (46).

Notice that $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$, implies that $\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}) \geq 0, \forall j$. So the weights $\varphi_{j,t'} \geq 0$. Since the weights sum to one, non-negativity of the weights further implies that none exceeds one.

□

Following the argument above, we can also write these weights in terms of linear projections:

$$\varphi_{j,t'} \equiv \frac{L(\theta_{j,t'}|\overline{\theta}_{t'})}{\sum_{j'=1}^{J} L(\theta_{j',t'}|\overline{\theta}_{t'})}.$$

Condition (i) of Assumption 6 implies that $L(\theta_{j,t}|\overline{\theta}_{t'}) = L(\theta_{j,t'}|\overline{\theta}_{t'}), \forall t \geq t'$, so the weights are proportional to the predicted level of skill j in periods $t \geq t'$ conditional on total worker productivity in period t', $\overline{\theta}_{t'}$.

C.7.1 Occupations as Bundles of Skills

We now consider the case in which log wage residuals are given by equation (23), where we define $\tilde{\mu}^{o}_{j,t} \equiv \mu_{j,t}\alpha^{o}_{j,t}$ and $\overline{\theta}^{o_{i,t}}_{i,t} \equiv \sum_{j=1}^{J} \tilde{\mu}^{o_{i,t}}_{j,t}\theta_{i,j,t}$.

Focusing on occupation stayers, we make the following assumption to accommodate multiple skills and occupations.

Assumption 8. *(i)* $\mathrm{Cov}(\Delta\theta_{j,t}, \theta_{j',t'}|o_t = o_{t-1}, o_{t'}) = 0$ *for all* j, j'*, and* $t - t' \geq 1$*; for known* $k \geq 1$ *and for all* $t - t' \geq k + 1$*: (ii)* $\mathrm{Cov}(\theta_{j,t}, \varepsilon_{t'}|o_t = o_{t-1}, o_{t'}) = \mathrm{Cov}(\theta_{j,t-1}, \varepsilon_{t'}|o_t = o_{t-1}, o_{t'}) = 0$ *for all* j*; (iii)* $\mathrm{Cov}(\varepsilon_t, \theta_{j,t'}|o_t = o_{t-1}, o_{t'}) = \mathrm{Cov}(\varepsilon_{t-1}, \theta_{j,t'}|o_t = o_{t-1}, o_{t'}) = 0$ *for all* j*; and (iv)* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}|o_t = o_{t-1}, o_{t'}) = \mathrm{Cov}(\varepsilon_{t-1}, \varepsilon_{t'}|o_t = o_{t-1}, o_{t'}) = 0$.

IV Estimator Conditional on $o_t = o_{t-1} = o$ and $o_{t'} = o'$. Given Assumption 8, our IV estimator conditional on $o_t = o_{t-1} = o$ and $o_{t'} = o'$ identifies the following for $t - t' \geq k + 1$:[63]

$$\begin{aligned}
\frac{\mathrm{Cov}(\Delta w_t, w_{t'}|o, o')}{\mathrm{Cov}(w_{t-1}, w_{t'}|o, o')} &= \frac{\sum_{j=1}^{J} \Delta\tilde{\mu}^o_{j,t}\, \mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o')}{\sum_{j'=1}^{J} \tilde{\mu}^o_{j',t-1}\, \mathrm{Cov}(\theta_{j',t'}, \overline{\theta}^{o'}_{t'}|o, o')} \\
&= \frac{\sum_{j=1}^{J} \left(\frac{\Delta\tilde{\mu}^o_{j,t}}{\tilde{\mu}^o_{j,t-1}}\right) \tilde{\mu}^o_{j,t-1}\, \mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o')}{\sum_{j'=1}^{J} \tilde{\mu}^o_{j',t-1}\, \mathrm{Cov}(\theta_{j',t'}, \overline{\theta}^{o'}_{t'}|o, o')} \\
&= \sum_{j=1}^{J} \upsilon^{o,o'}_{j,t',t-1} \left(\frac{\Delta\tilde{\mu}^o_{j,t}}{\tilde{\mu}^o_{j,t-1}}\right) \\
&= \sum_{j=1}^{J} \upsilon^{o,o'}_{j,t',t-1} \left(\frac{\Delta\mu_{j,t}}{\mu_{j,t-1}} + \frac{\mu_{j,t}}{\mu_{j,t-1}} \frac{\Delta\alpha^o_{j,t}}{\alpha^o_{j,t-1}}\right),
\end{aligned}$$

where

$$\upsilon^{o,o'}_{j,t',t-1} \equiv \frac{\tilde{\mu}^o_{j,t-1}\, \mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o')}{\sum_{j'=1}^{J} \tilde{\mu}^o_{j',t-1}\, \mathrm{Cov}(\theta_{j',t'}, \overline{\theta}^{o'}_{t'}|o, o')}.$$

Therefore, if $\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o') \geq 0$ for all j and (o, o'), the IV estimator for all occupational stayers reflects weighted average of the growth rate of skill-specific returns in occupation o. Notice that $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}|o, o') \geq 0, \forall j, j'$, implies that $\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o') \geq 0, \forall j$.

[63]To simplify notation, we let $\mathrm{Cov}(x, y|o, o')$ represent $\mathrm{Cov}(x, y|o_t = o_{t-1} = o, o_{t'} = o')$, $\mathrm{Cov}(x, y|o)$ represent $\mathrm{Cov}(x, y|o_t = o_{t-1} = o)$, and $\mathrm{E}[x|o]$ represent $\mathrm{E}[x|o_t = o_{t-1} = o]$.

IV Estimator for Stayers in Occupation o. Next, we show the IV estimator formula based on covariances conditioned only on $o_t = o_{t-1} = o$. Consider the long residual autocovariance that is not conditioned on $o_{t'}$:

$$\mathrm{Cov}(w_t, w_{t'}|o) = \mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o, o_{t'})|o\right] + \mathrm{Cov}\left(\mathrm{E}[w_t|o, o_{t'}], \mathrm{E}[w_{t'}|o, o_{t'}]|o\right). \tag{48}$$

With Assumption 8, the first term in equation (48) is

$$\mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o, o_{t'})|o\right] = \sum_{j=1}^{J} \tilde{\mu}^o_{j,t}\, \mathrm{E}\left[\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o_{t'}}_{t'}|o, o_{t'})\middle|o\right].$$

With additional assumptions $\mathrm{E}[\theta_{j,t}-\theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = 0$ for all j and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_{t'}|o_t, o_{t-1}, o_{t'}] = 0$, the second term in equation (48) is

$$\mathrm{Cov}\left(\mathrm{E}[w_t|o, o_{t'}], \mathrm{E}[w_{t'}|o, o_{t'}]|o\right) = \sum_{j=1}^{J} \tilde{\mu}^o_{j,t}\, \mathrm{Cov}\left(\mathrm{E}[\theta_{j,t'}|o, o_{t'}], \mathrm{E}[\overline{\theta}^{o_{t'}}_{t'}|o, o_{t'}]|o\right).$$

Therefore,

$$\begin{aligned}\mathrm{Cov}(w_t, w_{t'}|o) &= \sum_{j=1}^{J} \tilde{\mu}^o_{j,t}\left\{\mathrm{E}\left[\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o_{t'}}_{t'}|o, o_{t'})\middle|o\right] + \mathrm{Cov}\left(\mathrm{E}[\theta_{j,t'}|o, o_{t'}], \mathrm{E}[\overline{\theta}^{o_{t'}}_{t'}|o, o_{t'}]|o\right)\right\} \\ &= \sum_{j=1}^{J} \tilde{\mu}^o_{j,t}\, \mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o_{t'}}_{t'}|o).\end{aligned}$$

Altogether, Assumption 8, $\mathrm{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\theta_{j,t-1} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = 0$ for all j, and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_{t-1}|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_{t'}|o_t, o_{t-1}, o_{t'}] = 0$ imply that the IV estimator conditional on $o_t = o_{t-1} = o$ is

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'}|o)}{\mathrm{Cov}(w_{t-1}, w_{t'}|o)} = \frac{\sum_{j=1}^{J} \Delta\tilde{\mu}^o_{j,t}\, \mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o_{t'}}_{t'}|o)}{\sum_{j'=1}^{J} \tilde{\mu}^o_{j',t-1}\, \mathrm{Cov}(\theta_{j',t'}, \overline{\theta}^{o_{t'}}_{t'}|o)} = \sum_{j=1}^{J} \tilde{\upsilon}^o_{j,t',t-1}\left(\frac{\Delta\tilde{\mu}^o_{j,t}}{\tilde{\mu}^o_{j,t-1}}\right) = \sum_{j=1}^{J} \tilde{\upsilon}^o_{j,t',t-1}\left(\frac{\Delta\mu_{j,t}}{\mu_{j,t-1}} + \frac{\mu_{j,t}}{\mu_{j,t-1}}\frac{\Delta\alpha^o_{j,t}}{\alpha^o_{j,t-1}}\right),$$

where

$$\tilde{\upsilon}^{o}_{j,t',t-1} \equiv \frac{\tilde{\mu}^{o}_{j,t-1}\operatorname{Cov}(\theta_{j,t'}, \overline{\theta}^{o_{t'}}_{t'}|o)}{\sum_{j'=1}^{J}\tilde{\mu}^{o}_{j',t-1}\operatorname{Cov}(\theta_{j',t'}, \overline{\theta}^{o_{t'}}_{t'}|o)}.$$

If $\operatorname{Cov}(\theta_{j,t'}, \overline{\theta}^{o_{t'}}_{t'}|o) \geq 0$ for all j and o, the IV estimator for stayers in occupation o reflects a weighted average of the growth rate of skill-specific returns.

IV Estimator for All Occupation Stayers. Finally, consider the IV estimator for all occupation stayers (i.e., $o_t = o_{t-1}$), regardless of occupation. The log wage residual autocovariance for occupational stayers is

$$\begin{aligned}\operatorname{Cov}(w_t, w_{t'}|o_t = o_{t-1}) =& \operatorname{E}\left[\operatorname{Cov}(w_t, w_{t'}|o_t, o_{t-1})\middle|o_t = o_{t-1}\right] \\ &+ \operatorname{Cov}\left(\operatorname{E}[w_t|o_t, o_{t-1}], \operatorname{E}[w_{t'}|o_t, o_{t-1}]\middle|o_t = o_{t-1}\right).\end{aligned}$$

The first term is

$$\operatorname{E}\left[\operatorname{Cov}(w_t, w_{t'}|o_t, o_{t-1})\middle|o_t = o_{t-1}\right] = \sum_{j=1}^{J}\mu_{j,t}\operatorname{E}\left[\alpha^{o_t}_{j,t}\operatorname{Cov}(\theta_{j,t'}, \overline{\theta}^{o_{t'}}_{t'}|o_t, o_{t-1})|o_t = o_{t-1}\right].$$

Assuming $\operatorname{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}] = 0$ for all j and $\operatorname{E}[\varepsilon_t|o_t, o_{t-1}] = \operatorname{E}[\varepsilon_{t'}|o_t, o_{t-1}] = 0$, the second term is

$$\begin{aligned}&\operatorname{Cov}\left(\operatorname{E}[w_t|o_t, o_{t-1}], \operatorname{E}[w_{t'}|o_t, o_{t-1}]\middle|o_t = o_{t-1}\right) \\ =& \operatorname{E}\left[\operatorname{E}[w_t|o_t, o_{t-1}]\left(\operatorname{E}[w_{t'}|o_t, o_{t-1}] - \operatorname{E}[w_{t'}|o_t = o_{t-1}]\right)\middle|o_t = o_{t-1}\right] \\ =& \sum_{j=1}^{J}\mu_{j,t}\operatorname{E}\left[\alpha^{o_t}_{j,t}\operatorname{E}\left[\theta_{j,t'}|o_t, o_{t-1}\right]\left(\operatorname{E}[\overline{\theta}^{o_{t'}}_{t'}|o_t, o_{t-1}] - \operatorname{E}[\overline{\theta}^{o_{t'}}_{t'}|o_t = o_{t-1}]\right)\middle|o_t = o_{t-1}\right] \\ =& \sum_{j=1}^{J}\mu_{j,t}\operatorname{Cov}\left(\alpha^{o_t}_{j,t}\operatorname{E}\left[\theta_{j,t'}|o_t, o_{t-1}\right], \operatorname{E}[\overline{\theta}^{o_{t'}}_{t'}|o_t, o_{t-1}]\middle|o_t = o_{t-1}\right).\end{aligned}$$

Altogether,

$$
\begin{aligned}
\operatorname{Cov}(w_t, w_{t'}|o_t = o_{t-1}) =& \sum_{j=1}^{J} \mu_{j,t} \Big\{ \mathrm{E}[\alpha_{j,t}^{o_t} \operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t, o_{t-1}) | o_t = o_{t-1}] \\
& + \operatorname{Cov}\left(\alpha_{j,t}^{o_t} \mathrm{E}[\theta_{j,t'}|o_t, o_{t-1}], \mathrm{E}[\overline{\theta}_{t'}^{o_{t'}} | o_t, o_{t-1}] \big| o_t = o_{t-1}\right) \Big\} \\
=& \sum_{j=1}^{J} \mu_{j,t} \operatorname{Cov}(\alpha_{j,t}^{o_t} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1}).
\end{aligned}
$$

Therefore, Assumption 8, $\mathrm{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}] = \mathrm{E}[\theta_{j,t-1} - \theta_{j,t'}|o_t, o_{t-1}] = 0$ for all j, and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}] = \mathrm{E}[\varepsilon_{t-1}|o_t, o_{t-1}] = \mathrm{E}[\varepsilon_{t'}|o_t, o_{t-1}] = 0$ imply that the IV estimator for all stayers is

$$
\begin{aligned}
\frac{\operatorname{Cov}(\Delta w_t, w_{t'}|o_t = o_{t-1})}{\operatorname{Cov}(w_{t-1}, w_{t'}|o_t = o_{t-1})} =& \frac{\sum_{j=1}^{J} \mu_{j,t} \operatorname{Cov}(\alpha_{j,t}^{o_t} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})}{\sum_{j'=1}^{J} \mu_{j',t-1} \operatorname{Cov}(\alpha_{j',t-1}^{o_{t-1}} \theta_{j',t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})} - 1 \\
=& \sum_{j=1}^{J} \hat{\upsilon}_{j,t',t-1} \left(\frac{\mu_{j,t}}{\mu_{j,t-1}} \frac{\operatorname{Cov}(\alpha_{j,t}^{o_t} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})}{\operatorname{Cov}(\alpha_{j,t-1}^{o_{t-1}} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})} - 1 \right) \\
=& \sum_{j=1}^{J} \hat{\upsilon}_{j,t',t-1} \left(\frac{\Delta \mu_{j,t}}{\mu_{j,t-1}} + \frac{\mu_{j,t}}{\mu_{j,t-1}} \frac{\operatorname{Cov}(\Delta \alpha_{j,t}^{o_t} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})}{\operatorname{Cov}(\alpha_{j,t-1}^{o_{t-1}} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})} \right),
\end{aligned}
$$

where

$$
\hat{\upsilon}_{j,t',t-1} \equiv \frac{\mu_{j,t-1} \operatorname{Cov}(\alpha_{j,t-1}^{o_{t-1}} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})}{\sum_{j'=1}^{J} \mu_{j',t-1} \operatorname{Cov}(\alpha_{j',t-1}^{o_{t-1}} \theta_{j',t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1})}.
$$

The weights are positive if and only if

$$
\begin{aligned}
\operatorname{Cov}(\alpha_{j,t-1}^{o_{t-1}} \theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1}) =& \mathrm{E}\left[\alpha_{j,t-1}^{o_{t-1}} \operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t, o_{t-1}) | o_t = o_{t-1}\right] \\
& + \operatorname{Cov}\left(\alpha_{j,t-1}^{o_{t-1}} \mathrm{E}[\theta_{j,t'}|o_t, o_{t-1}], \mathrm{E}[\overline{\theta}_{t'}^{o_{t'}} | o_t, o_{t-1}] | o_t = o_{t-1}\right) \geq 0.
\end{aligned}
$$

The first term is positive if $\operatorname{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1} = o) \geq 0$ for all (j, o), which is the condition for positive weights among stayers in occupation o as shown above. The second term is zero when $\mathrm{E}[\overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1} = o]$ does not vary with o.

C.8 Identification with Occupation-Specific Wage Functions and Multiple Skills

We now consider the case in which log wage residuals are given by equation (24), where we now define $\tilde{\mu}^o_{j,t} \equiv \mu^o_t \alpha^o_{j,t}$ and $\overline{\theta}^{o_{i,t}}_{i,t} \equiv \sum_{j=1}^J \tilde{\mu}^{o_{i,t}}_{j,t} \theta_{i,j,t}$.

Identification of μ^o_t. With Assumption 8, the long autocovariance of log wage residuals for stayers in occupation $o_t = o_{t-1} = o$ and $o_{t'} = o'$ can be written as follows for $t - t' \geq k + 1$:

$$\text{Cov}(w_t, w_{t'}|o, o') = \mu^o_t \sum_{j=1}^J \alpha^o_{j,t} \text{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o'), \tag{49}$$

$$\text{Cov}(w_{t-1}, w_{t'}|o, o') = \mu^o_{t-1} \sum_{j=1}^J \alpha^o_{j,t-1} \text{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o'). \tag{50}$$

Together, equations (49) and (50) imply

$$\frac{\text{Cov}(\Delta w_t, w_{t'}|o, o')}{\text{Cov}(w_{t-1}, w_{t'}|o, o')} = \frac{\mu^o_t}{\mu^o_{t-1}} \frac{\sum_{j=1}^J \alpha^o_{j,t} \text{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o')}{\sum_{j'=1}^J \alpha^o_{j',t-1} \text{Cov}(\theta_{j',t'}, \overline{\theta}^{o'}_{t'}|o, o')} - 1 = \frac{\Delta \mu^o_t}{\mu^o_{t-1}} + \frac{\mu^o_t}{\mu^o_{t-1}} \left(\sum_{j=1}^J \tilde{\omega}^{o,o'}_{j,t',t-1} \frac{\Delta \alpha^o_{j,t}}{\alpha^o_{j,t-1}} \right),$$

where

$$\tilde{\omega}^{o,o'}_{j,t',t-1} \equiv \frac{\alpha^o_{j,t-1} \text{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o')}{\sum_{j'=1}^J \alpha^o_{j',t-1} \text{Cov}(\theta_{j',t'}, \overline{\theta}^{o'}_{t'}|o, o')}.$$

Occupation-specific returns, μ^o_t, are identified based on stayers in occupation o if $\alpha^o_{j,t}$ does not change over time. Furthermore, $\text{Cov}(\theta_{j,t'}, \theta_{j',t'}|o, o') \geq 0, \forall j, j'$, implies that $\text{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o') \geq 0, \forall j$, in which case the weights $\tilde{\omega}^{o,o'}_{j,t',t-1}$ are non-negative.

IV Estimator for Stayers in Occupation o. Next, we show that μ^o_t is identified based on covariances conditioned only on $o_t = o_{t-1} = o$ when we assume $\text{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = \text{E}[\theta_{j,t-1} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = 0$ for all j and $\text{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = \text{E}[\varepsilon_{t-1}|o_t, o_{t-1}, o_{t'}] = \text{E}[\varepsilon_{t'}|o_t, o_{t-1}, o_{t'}] = 0$ in addition to Assumption 8.

Consider the long autocovariance that is not conditioned on $o_{t'}$, i.e., equation (48). Given Assumption 8 and equation (49), the first term in equation (48) is

$$\mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o, o_{t'})|o\right] = \mu_t^o \,\mathrm{E}\left[\sum_{j=1}^{J} \alpha_{j,t}^o \,\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o, o_{t'})\Big|o\right].$$

The second term in equation (48) is

$$\begin{aligned}\mathrm{Cov}\left(\mathrm{E}[w_t|o, o_{t'}], \mathrm{E}[w_{t'}|o, o_{t'}]|o\right) =& \mathrm{Cov}\left(\gamma_t^o + \mu_t^o \sum_{j=1}^{J} \alpha_{j,t}^o \,\mathrm{E}[\theta_{j,t}|o, o_{t'}], \mathrm{E}[\overline{\theta}_{t'}^{o_{t'}}|o, o_{t'}]\Big|o\right)\\ =& \mu_t^o \sum_{j=1}^{J} \alpha_{j,t}^o \,\mathrm{Cov}\left(\mathrm{E}[\theta_{j,t'}|o, o_{t'}], \mathrm{E}[\overline{\theta}_{t'}^{o_{t'}}|o, o_{t'}]|o\right),\end{aligned}$$

using additional assumptions $\mathrm{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = 0$ and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_{t'}|o_t, o_{t-1}, o_{t'}] = 0$.

Therefore, the long autocovariances in equations (49) and (50) that are not conditioned on $o_{t'}$ are given by

$$\begin{aligned}\mathrm{Cov}(w_t, w_{t'}|o) =& \mu_t^o \sum_{j=1}^{J} \alpha_{j,t}^o \,\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o),\\ \mathrm{Cov}(w_{t-1}, w_{t'}|o) =& \mu_{t-1}^o \sum_{j=1}^{J} \alpha_{j,t-1}^o \,\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o),\end{aligned}$$

where we use the law of total covariance:

$$\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o) = \mathrm{E}\left[\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o, o_{t'})\Big|o\right] + \mathrm{Cov}\left(\mathrm{E}[\theta_{j,t'}|o, o_{t'}], \mathrm{E}[\overline{\theta}_{t'}^{o_{t'}}|o, o_{t'}]|o\right).$$

Combining the two equations gives

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'}|o)}{\mathrm{Cov}(w_{t-1}, w_{t'}|o)} = \frac{\Delta\mu_t^o}{\mu_{t-1}^o} + \frac{\mu_t^o}{\mu_{t-1}^o}\left(\sum_{j=1}^{J} \iota_{j,t',t-1}^o \frac{\Delta\alpha_{j,t}^o}{\alpha_{j,t-1}^o}\right),$$

where

$$\iota_{j,t',t-1}^o \equiv \frac{\alpha_{j,t-1}^o \,\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o)}{\sum_{j'=1}^{J} \alpha_{j',t-1}^o \,\mathrm{Cov}(\theta_{j',t'}, \overline{\theta}_{t'}^{o_{t'}}|o)}.$$

If $\text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o) \geq 0$ for all j, then these weights are non-negative.

IV Estimator for All Occupation Stayers. Finally, consider the IV estimator for all occupational stayers (i.e., $o_t = o_{t-1}$), regardless of occupation. We show that if $\gamma_t^o = 0$, then this estimator identifies a weighted average of occupation-specific skill returns.

The log wage residual autocovariance for occupational stayers is

$$\begin{aligned}\text{Cov}(w_t, w_{t'}|o_t = o_{t-1}) =& \,\text{E}\left[\text{Cov}(w_t, w_{t'}|o_t, o_{t-1})\middle|o_t = o_{t-1}\right]\\ &+ \text{Cov}\left(\text{E}[w_t|o_t, o_{t-1}], \text{E}[w_{t'}|o_t, o_{t-1}]\middle|o_t = o_{t-1}\right).\end{aligned}$$

Under Assumption 8, the first term is

$$\text{E}\left[\text{Cov}(w_t, w_{t'}|o_t, o_{t-1})\middle|o_t = o_{t-1}\right] = \text{E}\left[\mu_t^{o_t}\sum_{j=1}^{J}\alpha_{j,t}^{o_t}\text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}\middle|o_t, o_{t-1})|o_t = o_{t-1}\right].$$

If $\gamma_t^o = 0$ and $\text{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = 0$ for all j, the second term is

$$\begin{aligned}&\text{Cov}\left(\text{E}[w_t|o_t, o_{t-1}], \text{E}[w_{t'}|o_t, o_{t-1}]\middle|o_t = o_{t-1}\right)\\ =&\,\text{E}\left[\text{E}[w_t|o_t, o_{t-1}]\left(\text{E}[w_{t'}|o_t, o_{t-1}] - \text{E}[w_{t'}|o_t = o_{t-1}]\right)\middle|o_t = o_{t-1}\right]\\ =&\,\text{E}\left[\mu_t^{o_t}\sum_{j=1}^{J}\alpha_{j,t}^{o_t}\text{E}\left[\theta_{j,t'}|o_t, o_{t-1}\right]\left(\text{E}[\overline{\theta}_{t'}^{o_{t'}}|o_t, o_{t-1}] - \text{E}[\overline{\theta}_{t'}^{o_{t'}}|o_t = o_{t-1}]\right)\middle|o_t = o_{t-1}\right]\end{aligned}$$

Therefore, the long autocovariances for occupational stayers are

$$\begin{aligned}\text{Cov}(w_t, w_{t'}|o_t = o_{t-1}) &= \sum_o \Pr(o_t = o_{t-1} = o|o_t = o_{t-1})\mu_t^o\Psi_{t',t}^o,\\ \text{Cov}(w_{t-1}, w_{t'}|o_t = o_{t-1}) &= \sum_o \Pr(o_t = o_{t-1} = o|o_t = o_{t-1})\mu_{t-1}^o\Psi_{t',t-1}^o,\end{aligned}$$

where $\Pr(o_t = o_{t-1} = o|o_t = o_{t-1})$ denotes the share of stayers in occupation o among all occupation stayers and

$$\Psi_{t',t}^o \equiv \sum_{j=1}^{J}\alpha_{j,t}^o\left\{\text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}}|o_t = o_{t-1} = o) + \text{E}[\theta_{j,t'}|o_t = o_{t-1} = o]\left(\text{E}[\overline{\theta}_{t'}^{o_{t'}}|o_t = o_{t-1} = o] - \text{E}[\overline{\theta}_{t'}^{o_{t'}}|o_t = o_{t-1}]\right)\right\}.$$

The term $\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1} = o)$ is positive if $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'} | o_t = o_{t-1} = o) \geq 0$ for all j', while the second term is zero when $\mathrm{E}[\overline{\theta}_{t'}^{o_{t'}} | o_t = o_{t-1} = o]$ does not vary with o.

The two long residual autocovariances imply

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'} | o_t = o_{t-1})}{\mathrm{Cov}(w_{t-1}, w_{t'} | o_t = o_{t-1})} = \sum_o \zeta^o_{t',t-1} \left(\frac{\Delta \mu_t^o}{\mu_{t-1}^o} + \frac{\mu_t^o}{\mu_{t-1}^o} \frac{\Psi^o_{t',t} - \Psi^o_{t',t-1}}{\Psi^o_{t',t-1}} \right),$$

where

$$\zeta^o_{t',t-1} \equiv \frac{\Pr(o_t = o_{t-1} = o | o_t = o_{t-1}) \Psi^o_{t',t-1} \mu^o_{t-1}}{\sum_{o'} \Pr(o_t = o_{t-1} = o' | o_t = o_{t-1}) \Psi^{o'}_{t',t-1} \mu^{o'}_{t-1}}.$$

The weights are non-negative if and only if $\Psi^o_{t',t-1} \geq 0$ for all o.

If $\alpha^o_{j,t} = \alpha^o_{j,t-1}$, then $\Psi^o_{t',t} = \Psi^o_{t',t-1}$ and the IV estimator identifies a weighted average of $\Delta \mu_t^o / \mu_{t-1}^o$ across occupations.

D MD Estimation and Standard Errors

D.1 MD Estimation

For a given parameter vector $\mathbf{\Lambda}$, we can compute theoretical counterparts for $\mathrm{Cov}(w_t, w_{t'} | s, c)$, where $s \in \{\text{Non-college, College}\}$ indicates non-college and college status, implied by any specific model 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 generally partition the cohort set into ten-year cohort groups (e.g., C_1, C_2, C_3, and C_4 corresponding to cohorts born 1942–1951, 1952–1961, 1962–1971, and 1972–1981, respectively) or the experience set into 10-year experience groups E_1, E_2, E_3, and E_4, corresponding to 1–10, 11–20, 21–30, and 31–40 years, respectively, aggregating within these cohort or experience groups.

In the case of cohort grouping in Section 3.4, the minimum distance estimator $\hat{\mathbf{\Lambda}}$ solves

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

where Γ is described in Table 4; $\widehat{\mathrm{Cov}}(w_t, w_{t'} | s, C_j)$ is the sample covariance for residuals in years t and t' conditional on education group s and cohort group C_j; and $\mathrm{Cov}(w_t, w_{t'} | s, C_j, \mathbf{\Lambda})$ is the corresponding theoretical covariance given parameter vector $\mathbf{\Lambda}$.

In the case of experience grouping in Section 3.5.2, the minimum distance estimator $\hat{\mathbf{\Lambda}}$ solves

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

where $\Gamma = \{s, j, t, t'|1970 \le t' \le t \le 2012, t - t' \ge 6, s \in \{\text{Non-college, College}\}, j \in \{3, 4\}\}$; $\widehat{\mathrm{Cov}}(w_t, w_{t'}|s, E_j)$ is the sample covariance for residuals in years t and t' conditional on education group s and experience group E_j; and $\mathrm{Cov}(w_t, w_{t'}|s, E_j, \mathbf{\Lambda})$ is the corresponding theoretical covariance given parameter vector $\mathbf{\Lambda}$. In Sections 3.6 and 4, we also include covariance moments for less-experienced workers, E_1 and E_2, covering the same time periods.

D.2 Standard Errors

Consider the case of experience-based moments, and let $m = 1, 2, \ldots, M$ be the index of all moments. Let $d_{i,m}$ be the indicator of whether individual i contributes to the m^{th} moment $\mathrm{Cov}(w_t, w_{t'}|s, E_j)$. 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 $p_m(\mathbf{\Lambda}) = \mathrm{Cov}(w_t, w_{t'}|s, E_j, \mathbf{\Lambda})$. Then, we can write

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

where z_i includes $w_{i,t}$ $d_{i,m}$ for all t and m for individual i. Let $\boldsymbol{h}(z, \mathbf{\Lambda}) = [h_1(z, \mathbf{\Lambda})\ h_2(z, \mathbf{\Lambda})\ \ldots\ h_M(z, \mathbf{\Lambda})]^\top$. Then the following moment condition holds for the true parameter $\mathbf{\Lambda}_0$:

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

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

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

where $\boldsymbol{W} = \mathrm{diag}(\frac{N^2}{N_1^2}, \frac{N^2}{N_2^2}, \ldots, \frac{N^2}{N_M^2})$ and $N_m = \sum_{i=1}^{N} d_{i,m}$.

The GMM estimator $\hat{\mathbf{\Lambda}}$ is asymptotically normal with a variance matrix

$$\boldsymbol{V} = (\boldsymbol{H}^\top \boldsymbol{W}\boldsymbol{H})^{-1}(\boldsymbol{H}^\top \boldsymbol{W}\Omega\boldsymbol{W}\boldsymbol{H})(\boldsymbol{H}^\top \boldsymbol{W}\boldsymbol{H})^{-1},$$

where $\boldsymbol{H}$ is the Jacobian of the vector of moments, $\mathrm{E}[\partial \boldsymbol{h}(z, \mathbf{\Lambda}_0)/\partial \mathbf{\Lambda}^\top]$, and $\mathbf{\Omega} = \mathrm{E}[\boldsymbol{h}(z, \mathbf{\Lambda}_0)\boldsymbol{h}(z, \mathbf{\Lambda}_0)^\top]$.

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

$$\begin{aligned}\hat{\boldsymbol{H}} =& \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \boldsymbol{h}(z_i,\hat{\boldsymbol{\Lambda}})}{\partial \boldsymbol{\Lambda}^\top} = \boldsymbol{W}^{-\frac{1}{2}}\frac{\partial \boldsymbol{p}(\hat{\boldsymbol{\Lambda}})}{\partial \boldsymbol{\Lambda}^\top},\\ \hat{\boldsymbol{\Omega}} =& \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}(z_i,\hat{\boldsymbol{\Lambda}})\boldsymbol{h}(z_i,\hat{\boldsymbol{\Lambda}})^\top,\end{aligned}$$

where $\boldsymbol{W}^{-\frac{1}{2}} = \mathrm{diag}(\frac{N_1}{N}, \frac{N_2}{N}, \ldots, \frac{N_M}{N})$.

E PSID Data Details and Additional Results

E.1 Data Description

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.

Our sample is restricted to male heads of households from the core (SRC) sample and excludes those from any PSID oversamples (SEO, Latino) as well as those with zero individual weights.[64] 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 hours worked, and were not enrolled as a student.

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.

The wage measure we use divides annual earnings by annual hours worked, trimming the top and bottom 1% of all wages within year and college/non-college status by ten-year experience cells. The resulting sample contains 3,766 men and 44,547 person-year observations.

Figure E-1 shows the widening of the residual distribution over time, reporting average log wage residuals within each quartile. Consistent with Figure 1, the distribution widened most during the early 1980s and then again after 2000.

To examine whether attrition affects the residual autocovariances reported in Figure 3, Figure E-2 shows the autocovariances, $\mathrm{Cov}(w_b, w_t)$ for $6 \le t - b \le 16$, 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 E-2, 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 E-3 shows the residual autocovariances for individuals with 1–14 years of experience in the

[64] 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).

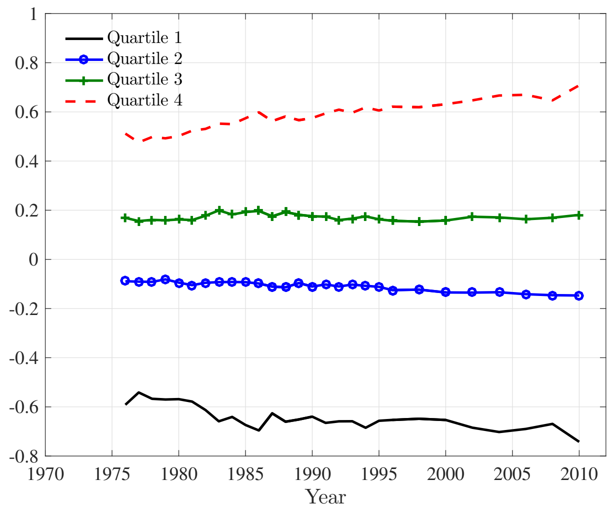


Figure E-1: Average Log Wage Residuals by Quartile

base years. Regardless of the base year, the autocovariances are typically declining from late 1980s through the 1990s as in Figures 3(a) (full sample) and 3(b) (men with 15–30 years experience) in the text. The lines also shift upwards over time, consistent with rising skill variances.

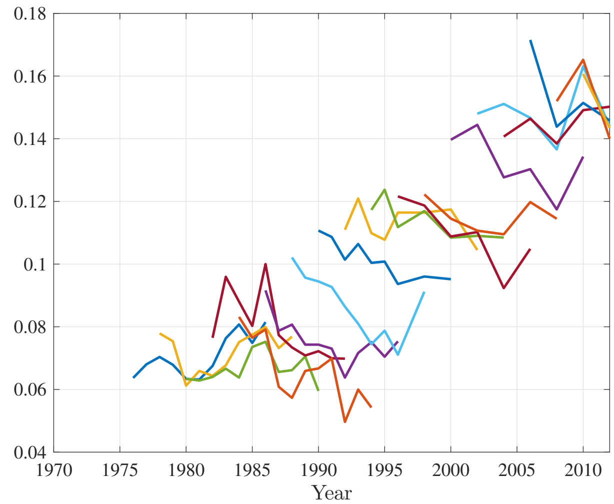


Figure E-2: Log Wage Residual Autocovariances (‘Balanced’ Sample)

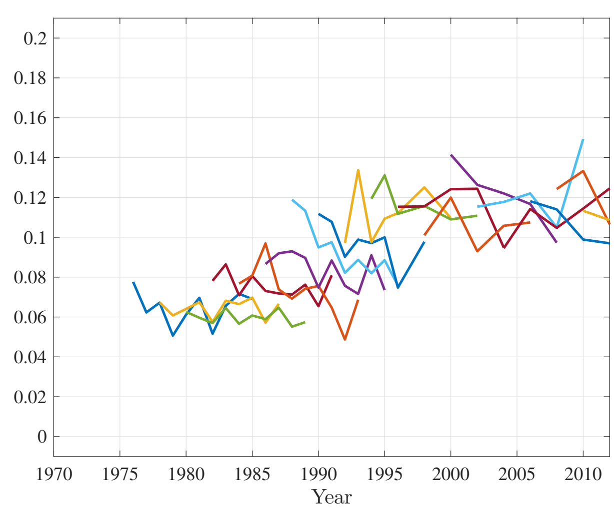


Figure E-3: Autocovariances for Log Wage Residuals (1–14 Years of Experience)

E.2 2SLS Estimates of Skill Returns by Education (PSID)

Table E-1: 2SLS estimates of $\Delta_2\mu_t/\mu_{t-2}$ for men with 21–40 years experience, 1979–1995

	1979–1980	1981–1983	1984–1986	1987–1989	1990–1992	1993–1995
			A. All men			
$\Delta_2\mu_t/\mu_{t-2}$	-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
1st stage F-Statistic	117.23	132.19	66.26	130.53	132.83	201.62
			B. Non-college men			
$\Delta_2\mu_t/\mu_{t-2}$	-0.108	0.009	-0.019	-0.101	-0.051	-0.105
	(0.061)	(0.062)	(0.072)	(0.070)	(0.066)	(0.065)
Observations	552	777	678	609	555	542
1st stage F-Statistic	66.06	59.12	24.04	55.22	65.32	72.32
			C. College men			
$\Delta_2\mu_t/\mu_{t-2}$	-0.031	-0.166**	-0.003	-0.088	-0.024	-0.104
	(0.068)	(0.053)	(0.074)	(0.060)	(0.059)	(0.060)
Observations	314	491	509	524	594	758
1st stage F-Statistic	73.87	90.56	99.30	71.46	66.14	142.24

Notes: Estimates from 2SLS regression of $w_{i,t} - w_{i,t-2}$ on $w_{i,t-2}$ using instruments $(w_{i,t-8}, w_{i,t-9})$. Experience restrictions based on year t. * denotes significance at 0.05 level.

Table E-2: 2SLS estimates of $\Delta_2\mu_t/\mu_{t-2}$ for men with 21–40 years experience, 1996–2012

	1996–2000	2002–2006	2008–2012
	A. All men		
$\Delta_2\mu_t/\mu_{t-2}$	-0.084*	-0.040	-0.058
	(0.030)	(0.032)	(0.031)
Observations	1,427	1,591	1,493
1st stage F-Statistic	295.75	281.91	267.83
	B. Non-college men		
$\Delta_2\mu_t/\mu_{t-2}$	-0.073	-0.064	0.011
	(0.053)	(0.046)	(0.082)
Observations	589	624	481
1st stage F-Statistic	96.00	126.69	114.93
	C. College men		
$\Delta_2\mu_t/\mu_{t-2}$	-0.094**	-0.040	-0.074*
	(0.036)	(0.042)	(0.032)
Observations	834	960	866
1st stage F-Statistic	212.60	169.90	163.07

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. Experience restrictions based on year t. * denotes significance at 0.05 level.

E.3 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 3.3 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 (10) as follows:

$$\Delta_2 w_{i,t} = \left(\frac{\Delta_2 \mu_t}{\mu_{t-2}}\right) w_{i,t-2} + u_{i,t}, \tag{51}$$

where $u_{i,t} \equiv \varepsilon_{i,t} - \frac{\mu_t}{\mu_{t-2}} \varepsilon_{i,t-2} + \mu_t \Delta_2 \theta_{i,t}$.

Serially uncorrelated skill shocks implies the following moment condition:

$$\mathrm{E}[w_{t'} u_t] = 0, \qquad \text{for } t' \leq t-2-k. \tag{52}$$

Under the stronger assumption that $\mathrm{Var}(\Delta\theta_t) = 0, \forall t$, the following additional moment condition holds:

$$\mathrm{E}[w_{t''} u_t] = 0, \qquad \text{for } t'' \geq t+k. \tag{53}$$

Equation (53) will not hold when $\mathrm{Var}(\Delta_2\theta_t) > 0$, and the IV estimate using future residuals as instruments is asymptotically biased with probability limit

$$\frac{\mathrm{Cov}(\Delta_2 w_t, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})} = \frac{\Delta_2 \mu_t}{\mu_{t-2}} + \frac{\mu_t}{\mu_{t-2}} \frac{\mathrm{Var}(\Delta_2\theta_t)}{\mathrm{Var}(\theta_{t-2})} > \frac{\Delta_2 \mu_t}{\mu_{t-2}}, \qquad \text{for } t' \geq t+k.$$

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 $t'' \geq t+k$,

$$\frac{\mathrm{Var}(\Delta_2\theta_t)}{\mathrm{Var}(\theta_{t-2})} = \left[\frac{\mathrm{Cov}(\Delta_2 w_t, w_{t''})}{\mathrm{Cov}(w_{t-2}, w_{t''})} - \frac{\mathrm{Cov}(\Delta_2 w_t, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right] \left[1 + \frac{\mathrm{Cov}(\Delta_2 w_t, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right]^{-1}. \tag{54}$$

E.3.1 Overidentification Tests

We begin by testing the moments in equation (53) 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 (52) and the first two available out of $w_{i,t+6}, w_{i,t+7}, w_{i,t+8}, w_{i,t+9}$ for (53).

Table E-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 (51). Panel A reports estimates when moments from both equations (52) and (53) are used (i.e., lags and leads), while Panel B reports estimates when only the moment condition from equation (52) is used (i.e., lags only). The sample is restricted to be the same in both panels.[65]

Comparing the J-statistics in Panels A and B in Table E-3, we can test the validity of using leads as instruments (i.e. moments in equation (53)). Since the differences are greater than 5.991 (critical value for χ^2_2 at significance level 0.05) except for 1979–1980 and 2002–2004, we reject the 'leads' moments in equation (53) at 5% significance level for 1981–2000. (See Panel C for p-values of these tests.) Moreover, all J-statistics in Panel B are smaller than 3.841 (critical value for χ^2_1 at significance level 0.05), implying that we cannot reject the lags as instruments (i.e. moments in equation (52)) at the 5% level. Altogether, these results suggest that the lagged residuals are valid instruments, while the leads are not (in most years).

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

[65]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 2 and 3). Panel A of Table E-4 below reports GMM estimates when this sample selection is not imposed. Those results are directly comparable and quite similar to those in Tables 2 and 3.

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

	1979–80	1981–83	1984–86	1987–89	1990–92	1993–95	1996–2000	2002–04
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
C. p-Values for J-Tests of the Validity of Leads as Instruments								
Leads	0.111	0.006	0.004	0.013	0.009	0.034	0.012	0.443
Lags	0.924	0.665	0.427	0.351	0.800	0.626	0.744	0.899

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. Panel C reports p-values based on a comparison of J-statistics from Panels A and B. * denotes significance at 0.05 level.

E.3.2 Inferring Relative Magnitude of Skill Shocks

Table E-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 (52) are used (i.e., 2 nearest valid lags). These results are analogous to the 2SLS estimates in Tables 2 and 3, 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 (53) 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 (54) 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 E-4: GMM Estimates of Skill Return Growth using Leads vs. Lags as Instruments and Relative Skill Shock Variance (Unbalanced Samples)

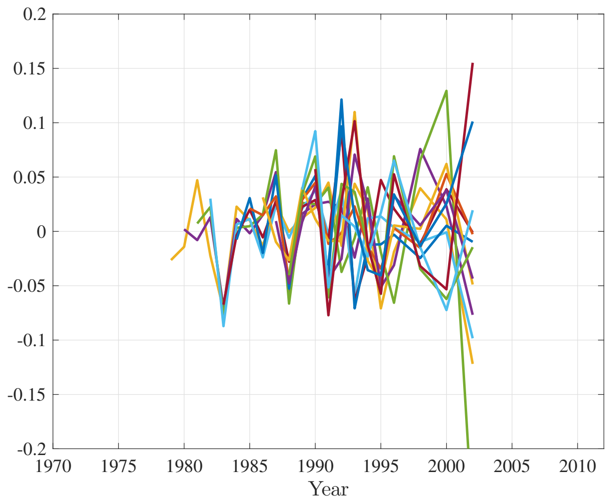
	1979–80	1981–83	1984–86	1987–89	1990–92	1993–95	1996–2000	2002–04
A. 2 Nearest Valid Lags as Instruments								
Coeff. on $w_{i,t-2}$	-0.033	-0.045	-0.044	-0.084*	-0.083*	-0.067	-0.076*	-0.090*
	(0.045)	(0.038)	(0.038)	(0.033)	(0.035)	(0.035)	(0.025)	(0.035)
Observations	1,349	2,077	2,188	2,245	2,189	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								
$\mathrm{Var}(\Delta_2\theta_t)/\mathrm{Var}(\theta_{t-2})$	.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 (54).

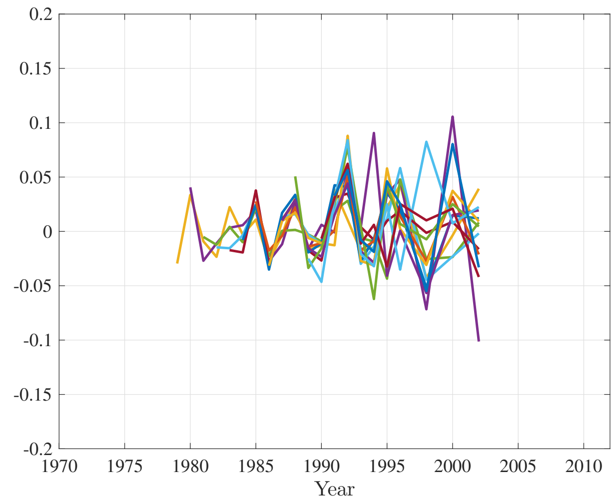
* denotes significance at 0.05 level.

E.4 Testing HIP based on growth in log wage residuals

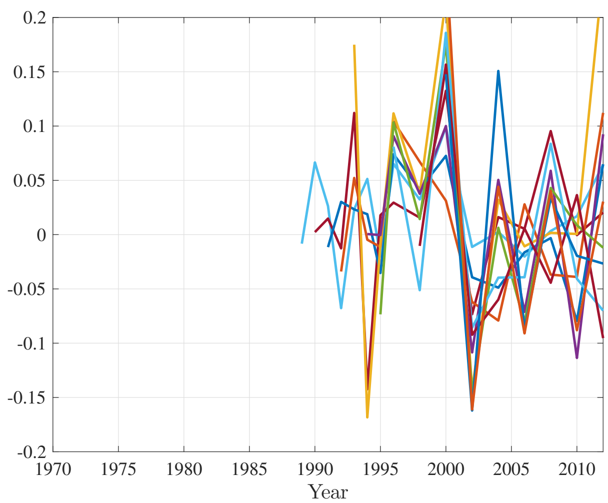
This appendix shows results for $\mathrm{Cov}(\Delta(w_t/\mu_t), w_{t'})$ in PSID.



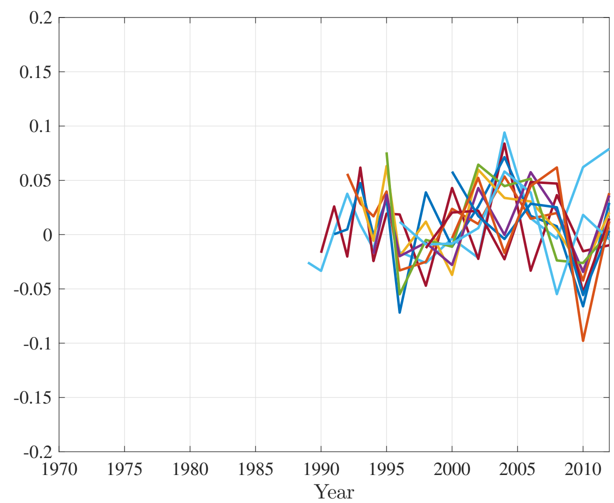
(a) Cohort Group 3, Non-College



(b) Cohort Group 3, College



(c) Cohort Group 4, Non-College



(d) Cohort Group 4, College

Figure E-4: $\mathrm{Cov}(\Delta(w_t/\mu_t), w_{t'})$ for each t, t' by Cohort Group

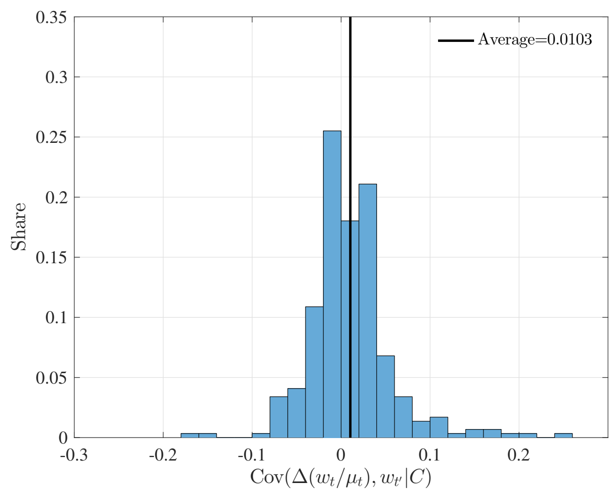
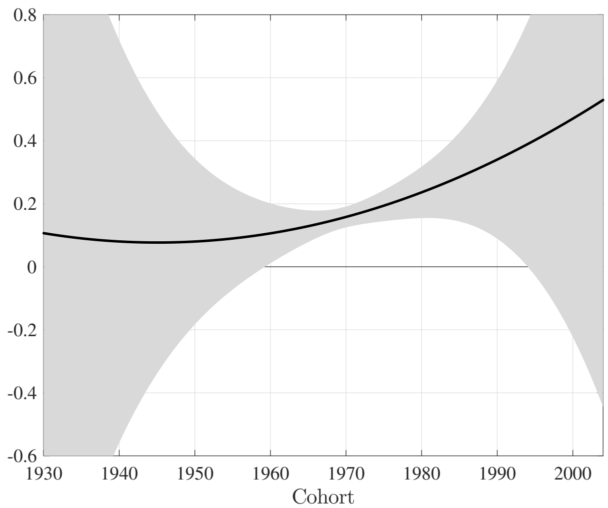


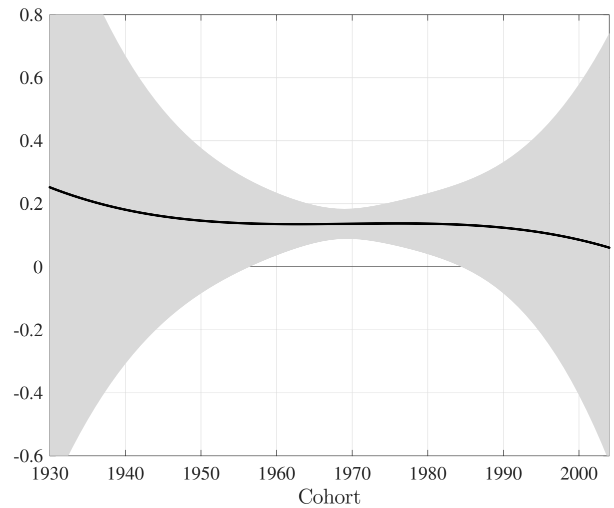
Figure E-5: Distribution of $\mathrm{Cov}(\Delta(w_t/\mu_t), w_{t'})$ for all (t, t', C) for for Low-Experience Men

E.5 Additional estimates for model with AR(1) skill dynamics

Figure E-6 reports estimated $\mathrm{Var}(\psi|c)$ when allowing for time-varying AR(1) skill shocks as discussed in Section 3.5.2. See the text for additional details on the specification.



(a) Non-College Men



(b) College Men

Figure E-6: $\mathrm{Var}(\psi|c)$ implied by MD estimates allowing for time-varying AR(1) skill shocks, 21–40 years of experience

E.6 Comparison with specifications used in literature on earnings dynamics

As noted in the text, the literature on earnings dynamics estimates similar models; although, it rarely considers the returns to unobserved skills. Haider (2001) and Moffitt and Gottschalk (2012) are notable exceptions. Since their estimates suggest qualitatively different patterns for the returns to skill over time, this appendix explores whether those can be explained by seemingly modest differences in specifications.

As with most of the earnings dynamics literature, Haider (2001) and Moffitt and Gottschalk (2012) assume $\varepsilon_t \sim \mathrm{ARMA}(1,1)$, which is very similar, though not identical, to our specification in equation (15) with $k = 2$. We prefer the latter given our desire to maintain a completely flexible (time-varying) autocorrelation structure for transitory shocks, $\tilde{\varepsilon}_t$. Regardless, the two specifications yield very similar estimated μ_t series when using the same long autocovariances (i.e., $\mathrm{Cov}(w_t, w_{t'})$ for $|t - t'| \geq 6$) in MD estimation, as can be seen from comparing the estimated returns given by the blue lines with circles in Figure E-7 and the returns reported in Figure 13. The former assumes $\varepsilon_t \sim \mathrm{ARMA}(1,1)$, while the latter assumes ε_t contains an AR(1) component, φ_t, and transitory component (with $k = 6$), $\tilde{\varepsilon}_t$, as described by equation (15).[66] Figure E-7 also reproduces estimates from our baseline specification, which assumes that ε_t only contains a transitory component (with $k = 6$). Altogether, these estimates indicate that accounting for persistent non-skill shocks has little effect on estimated returns to skills over time.

The red dashed lines in Figure E-7 show that estimating the $\varepsilon_t \sim \text{ARMA}(1,1)$ specification using all autocovariances, as in Haider (2001) and Moffitt and Gottschalk (2012) and the earnings dynamics literature more generally, yields quite different μ_t patterns for college men. This suggests that accounting for a flexible short-term autocorrelation structure – accommodated by using only long autocovariances – has important implications for skill returns among college men. (Failing to allow for time variation in the short-term autocorrelation structure for ε_t is primarily responsible for the different patterns.) The remaining two estimated μ_t series in Figure E-7 impose additional assumptions about skills made by Haider (2001) and Moffitt and Gottschalk (2012): time-invariance of non-skill growth innovations and cohort-invariance of initial skill variation. Neither of these assumptions has major implications for our estimated μ_t series for non-college or college men.

Since Haider (2001) and Moffitt and Gottschalk (2012) estimate their models for all men (rather than by education), we reproduce our analysis pooling all non-college and college men. The estimated μ_t series is reported in Figure E-8. In this case, we find that their assumptions imposing stability of initial skill distributions and of the distribution of skill growth innovations produce upward biased estimates in skill return growth beginning in the mid-1990s. Restrictions on the short-term autocorrelation structure for non-skill shocks have relatively modest effects on their estimated μ_t series, similar to patterns observed for the non-college sample.

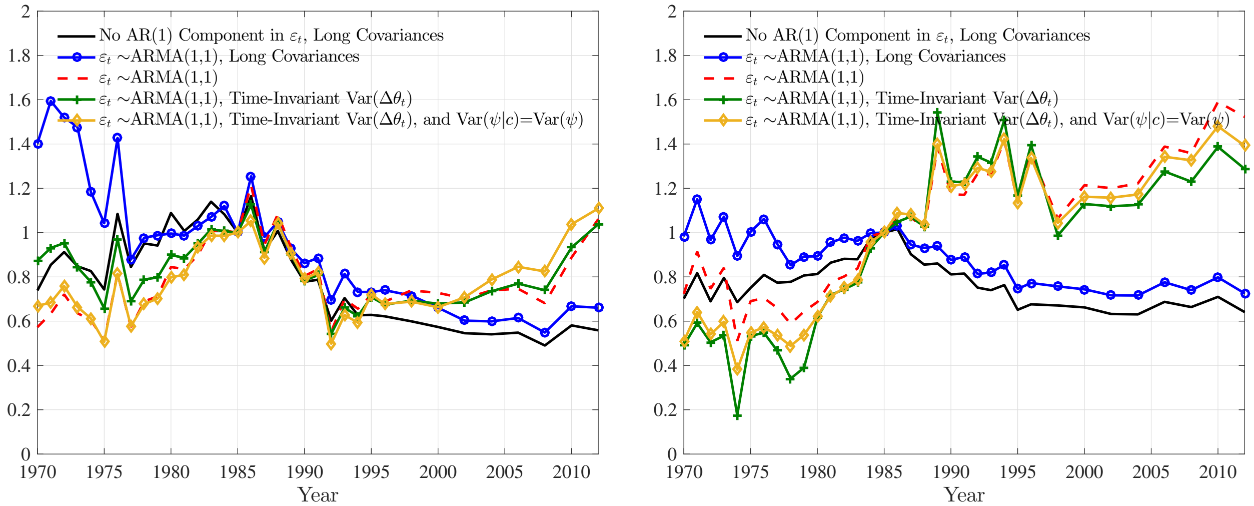
E.7 Estimation with Multiple Occupations in PSID

In creating occupation codes for our sample period, we combine retrospective (1968–1980) and original (1981–2013) measures, which creates a break in occupational mobility trends (in 1981) due to lower measurement error in the retrospective measures (Kambourov and Manovskii, 2008). The 3-digit codes are based on the 1970 Census classification prior to 2002, while they are based on the 2000 Census classification afterwards. Therefore, we do not measure occupation changes between years 2000 and 2002.[67] We create 1- and 2-digit codes from the first and first two digits of the 3-digit codes, respectively.

We use the 3-digit codes to create 3 broad and exclusive occupation categories (cognitive, routine, and manual) considered by Cortes (2016). Given small sample sizes for manual occupations in the PSID, our analysis focuses on cognitive and routine occupations.

[66] Formally, we specify ARMA(1,1) non-skill shocks as $\varepsilon_{i,t} = \nu_{i,t}$ for $t = c_i + 1$ and $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + \nu_{i,t} + \beta_1\nu_{i,t-1}$ for $t > c_i + 1$, with $\text{Cov}(\nu_t, \nu_{t'}) = 0$ for all $t \neq t'$.

[67] Since we pool observations across several years (assuming constant growth of skill returns within each pooled sample) for 2SLS estimation, the change in skill return between 2000 and 2002 reflects an extrapolation from adjacent years.



(a) Non-College (b) College

Figure E-7: Estimated μ_t under Different Restrictions: Non-College and College

We also estimate skill returns for those who stay in occupations with high social skill requirements, as measured by data from the Occupational Information Network (O*NET). The O*NET is a survey of U.S. workers that asks about the abilities, skills, knowledge, and work activities required in an occupation. Following Deming (2017), we measure an occupation's social skill intensity as the average of the four items in the 1998 O*NET module on "social skills" (coordination, negotiation, persuasion, and social perceptiveness). The social skill intensity measures are then assigned to individuals in the PSID sample based on their current 3-digit occupation in each year. We define social occupations as occupations that fall in the top third of the social skill intensity distribution in the pooled sample of worker-year observations. As noted by Deming (2017), cognitive occupations are also very likely to be social occupations. Among worker-year observations in cognitive occupations, around 59% are in social occupations. Conversely, around 76% of observations in social occupations are also in cognitive occupations.

E.7.1 GMM Estimation using Occupation Stayers and Switchers

We estimate occupation-specific γ_t^o and μ_t^o for routine and cognitive occupations (normalizing $\gamma_{1985}^o = 0$ and $\mu_{1985}^o = 1$ for routine occupations) using optimal two-step GMM. Because we use the PSID data, we use the following moments based on equation (19):

$$\mathrm{E}\left[z_t \left\{ \Delta_2 w_t - \left(\gamma_t^{o_t} - \frac{\mu_t^{o_t}}{\mu_{t-2}^{o_{t-2}}} \gamma_{t-2}^{o_{t-2}} \right) - \left(\frac{\mu_t^{o_t}}{\mu_{t-2}^{o_{t-2}}} - 1 \right) w_{t-2} \right\} \middle| o_t, o_{t-2} \right] = \mathbf{0}, \quad \forall (t, o_t, o_{t-2}),$$

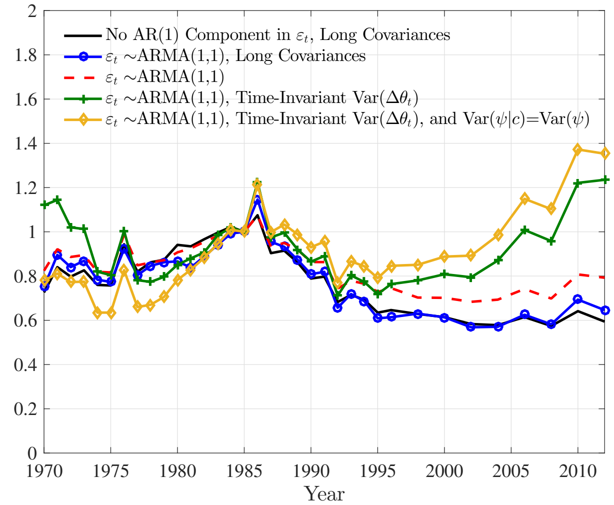


Figure E-8: Estimated μ_t under Different Restrictions: All Men

where $z_{i,t} = (1, w_{i,t-8}, w_{i,t-9})^\top$ (or $(1, w_{i,t-8}, w_{i,t-10})^\top$ in later sample years). We use linear splines for γ_t^o and μ_t^o, each with 14 knots in t, restricting to moment conditions with at least 20 observations (9 switcher moments are dropped). Altogether, there are 54 parameters with 303 moment conditions.

We also estimate the model imposing equal skill returns, $\mu_t^{\text{routine}} = \mu_t^{\text{cognitive}} = \mu_t$ for all t. The estimated μ_t and γ_t^o series are nearly identical to their counterparts reported in Figure 18, while the J-statistic comparing the unrestricted and restricted criterion functions equals 20.08 and is distributed χ^2_{14} under the null hypothesis of equal skill returns. Thus, we cannot reject the null of identical returns at conventional levels (p-value = 0.128).

For comparison with Cortes (2016), Figure E-9 reports estimates of γ_t^o and $\mathrm{E}[\theta_t|o_t]$ when imposing time-invariant $\mu_t^o = \mu^o$. Estimated time profiles for γ_t^o and $\mathrm{E}[\theta_t|o_t]$ are notably flatter than their counterparts allowing for variation in skill returns (see Figures 18(b) and 19(b)).

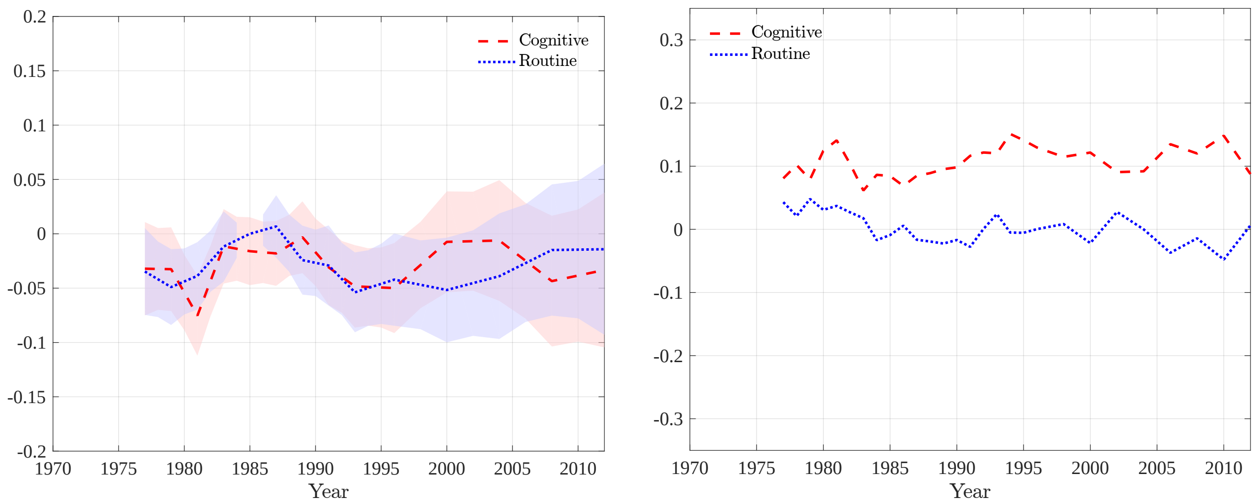
Finally, we use the same estimation strategy, restricting the sample to men who had 21–40 years of experience in year t. Due to fewer observations, this reduces the number of moments used in estimation to 261. These results are reported in Figures E-10 and E-11.

E.7.2 2SLS Estimated Returns for Occupational Stayers

Tables E-5 and E-6 report 2SLS estimates (and standard errors) for skill return growth underlying Figures 17 and 20 in the main text. First-stage F-statistics for the instruments are also reported.

Figure E-12 reports 2SLS estimates based on occupation-stayers in years $t-2$ to t based on different occupation classifications.[68] In all cases, $w_{i,t-8}$ and $w_{i,t-9}$ (or $w_{i,t-10}$) residuals are used as instruments.

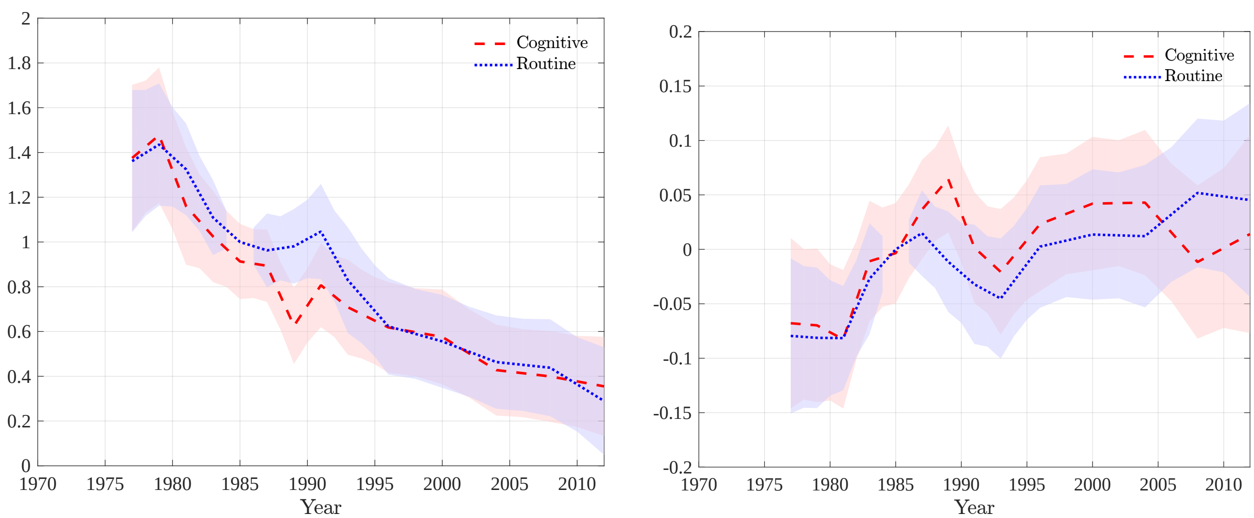
[68]These estimates are based on the same subperiods as reported in Tables E-5 and E-6. The 3-category occupation estimates



(a) γ_t^o (b) $\mathrm{E}\left[\theta_t|o_t\right]$

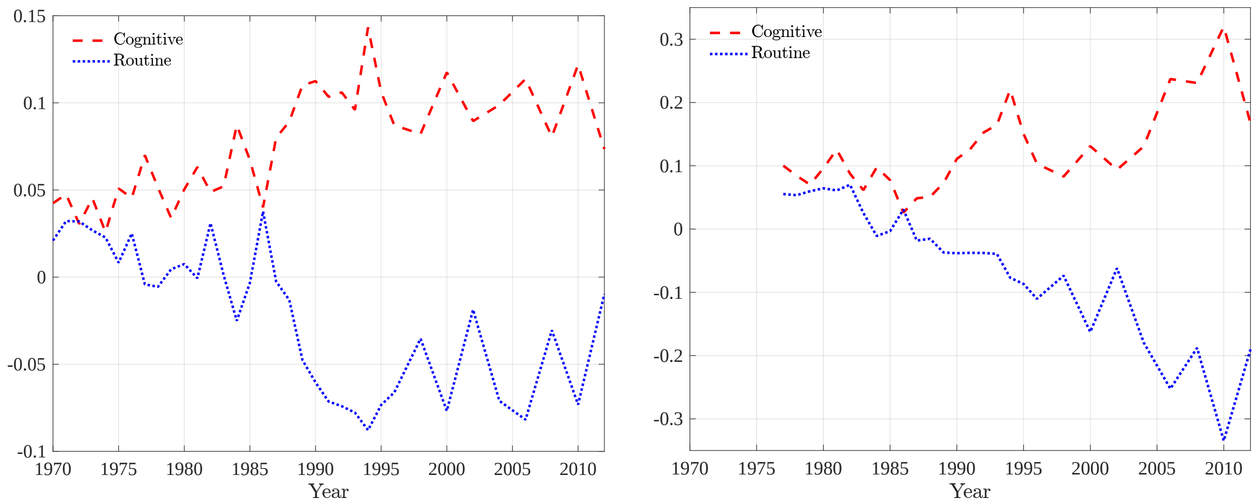
Figure E-9: GMM estimates imposing time-invariant μ_t^o

Notes: Imposing $\mu^{routine} = 1$, $\mu^{cognitive}$ is estimated to be 0.946 (SE=0.026).



(a) μ_t^o (b) γ_t^o

Figure E-10: GMM estimates of μ_t^o and γ_t^o, 21–40 years of experience in t



(a) $\mathrm{E}[w_t|o_t]$ (b) $\mathrm{E}[\theta_t|o_t]$

Figure E-11: Average log wage residual and skill by occupation, 21–40 years of experience in t

Estimated skill return patterns are very similar regardless of how narrowly we define occupations.

As noted in the text, estimated return growth for stayers in occupation $o_t = o_{t-1} = o$ should not depend on earlier occupation ($o_{t'}$) under Assumption 5. Estimates reported in Figures E-13 and E-14 indicate very similar estimated skill return profiles for occupation stayers with $o_{t'} = o$ vs. $o_{t'} \neq o$.

reflect those remaining within occupations classified as either cognitive, routine, or manual based on Cortes (2016).

Table E-5: 2SLS Estimates for Occupational Stayers, 1979–1995

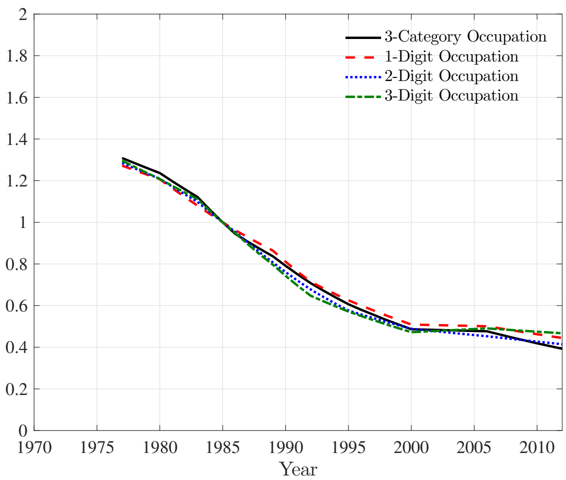
	1979–1980	1981–1983	1984–1986	1987–1989	1990–1992	1993–1995
			A. Non-College Men			
$\Delta_2\mu_t/\mu_{t-2}$	-0.045	0.024	-0.090	-0.171*	-0.081	-0.115
	(0.060)	(0.064)	(0.060)	(0.053)	(0.059)	(0.075)
Observations	509	598	548	519	511	423
1st stage F-Statistic	62.246	66.535	95.190	54.104	65.388	65.562
			B. College Men			
$\Delta_2\mu_t/\mu_{t-2}$	-0.104	-0.116	-0.083	-0.046	-0.157*	-0.063
	(0.058)	(0.065)	(0.058)	(0.055)	(0.046)	(0.064)
Observations	369	511	591	694	665	731
1st stage F-Statistic	88.830	76.805	86.170	55.780	95.572	132.245
			C. Cognitive Occupations			
$\Delta_2\mu_t/\mu_{t-2}$	-0.084	-0.075	-0.107*	-0.097	-0.136*	-0.102
	(0.071)	(0.062)	(0.053)	(0.050)	(0.043)	(0.056)
Observations	506	776	869	905	895	914
1st stage F-Statistic	92.258	75.839	37.259	74.934	123.756	144.648
			D. Routine Occupations			
$\Delta_2\mu_t/\mu_{t-2}$	0.020	-0.059	-0.129*	-0.073	-0.084	-0.107
	(0.070)	(0.049)	(0.047)	(0.048)	(0.065)	(0.056)
Observations	648	915	908	929	899	801
1st stage F-Statistic	81.071	85.482	85.385	108.442	71.017	101.606
			E. Social Occupations			
$\Delta_2\mu_t/\mu_{t-2}$	0.062	-0.067	-0.108	-0.083	-0.202*	-0.048
	(0.097)	(0.069)	(0.068)	(0.068)	(0.054)	(0.085)
Observations	374	525	573	594	598	589
1st stage F-Statistic	55.417	81.778	85.375	48.250	78.289	72.354

Notes: Estimates from 2SLS regression of $\Delta_2 w_{i,t}$ on $w_{i,t-2}$ using instruments $(w_{i,t-8}, w_{i,t-9})$.
* denotes significance at 0.05 level.

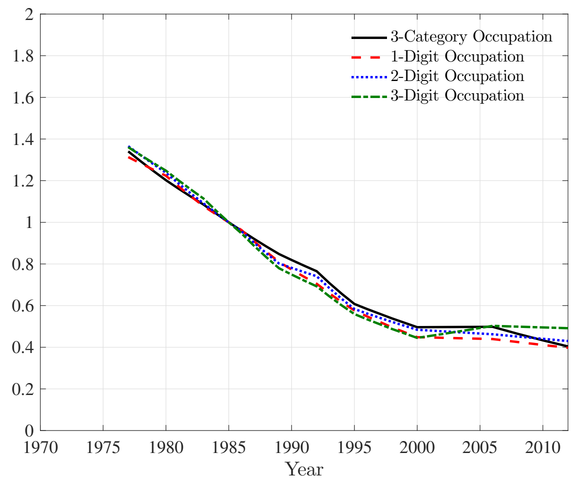
Table E-6: 2SLS Estimates for Occupational Stayers, 1996–2012

	1996–2000	2002–2006	2008–2012
	A. Non-College Men		
$\Delta_2\mu_t/\mu_{t-2}$	-0.088	0.017	0.032
	(0.051)	(0.079)	(0.082)
Observations	407	252	325
1st stage F-Statistic	68.156	30.257	105.090
	B. College Men		
$\Delta_2\mu_t/\mu_{t-2}$	-0.070	0.006	-0.035
	(0.045)	(0.049)	(0.042)
Observations	706	443	662
1st stage F-Statistic	155.501	99.896	133.535
	C. Cognitive Occupations		
$\Delta_2\mu_t/\mu_{t-2}$	-0.110*	-0.002	-0.097*
	(0.038)	(0.052)	(0.041)
Observations	882	836	794
1st stage F-Statistic	153.995	127.048	127.603
	D. Routine Occuaptions		
$\Delta_2\mu_t/\mu_{t-2}$	-0.068	-0.018	-0.078
	(0.049)	(0.039)	(0.054)
Observations	810	801	707
1st stage F-Statistic	112.638	145.430	124.283
	E. Social Occupations		
$\Delta_2\mu_t/\mu_{t-2}$	-0.122*	-0.010	-0.038
	(0.049)	(0.043)	(0.042)
Observations	580	606	587
1st stage F-Statistic	94.614	123.053	93.845

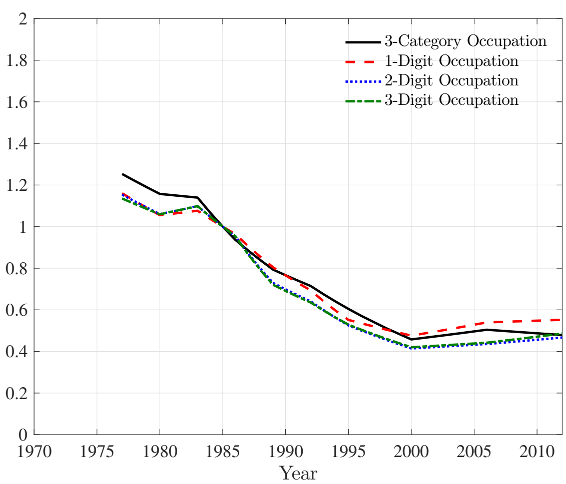
Notes: Estimates from 2SLS regression of $\Delta_2 w_{i,t}$ on $w_{i,t-2}$ using instruments (w_{t-8}, w_{t-9}) for 1996–2000 and (w_{t-8}, w_{t-10}) for 2002–2006 and 2008–2012. * denotes significance at 0.05 level.



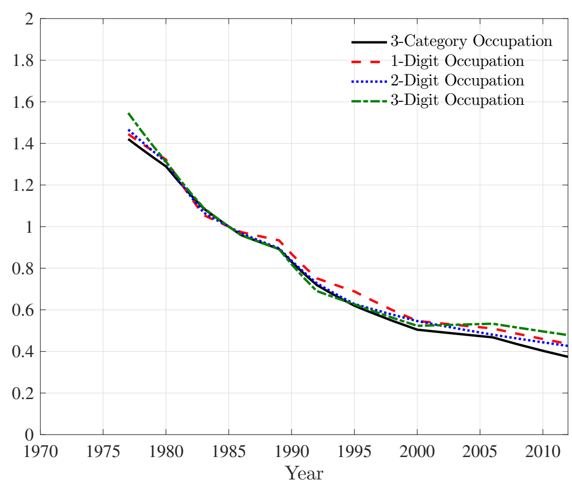
(a) Constant Occupation between $t-2$ and t



(b) Constant Occupation between $t-2$ and t: 21–40 Years of Experience

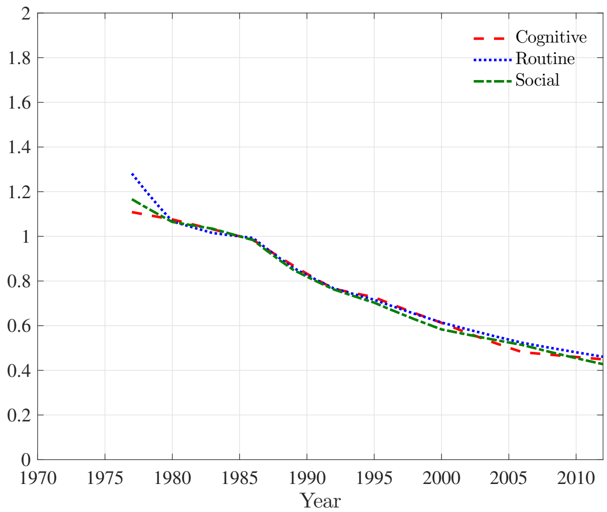


(c) Constant Occupation between $t-2$ and t: Non-College

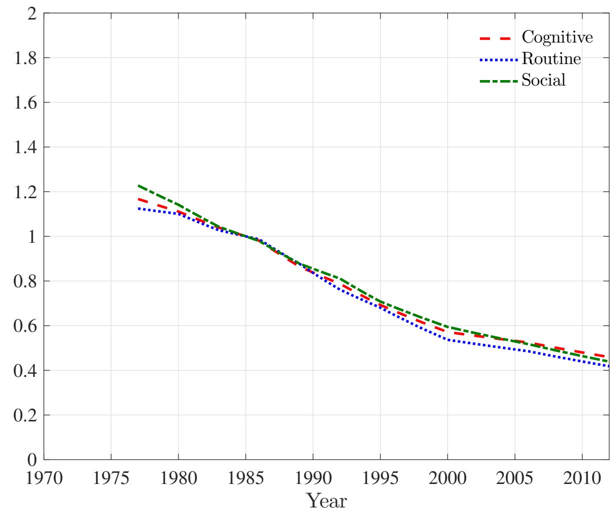


(d) Constant Occupation between $t-2$ and t: College

Figure E-12: μ_t Implied by 2SLS Estimates for Occupational Stayers from $t-2$ to t

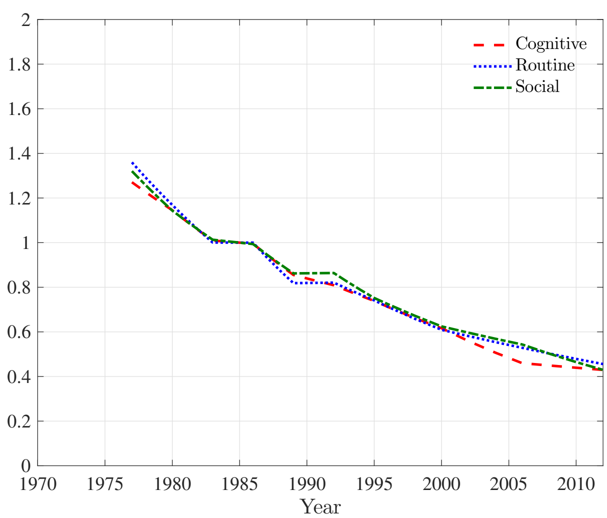


(a) Same Occupation in t' ($o_t = o_{t-2} = o_{t'}$)

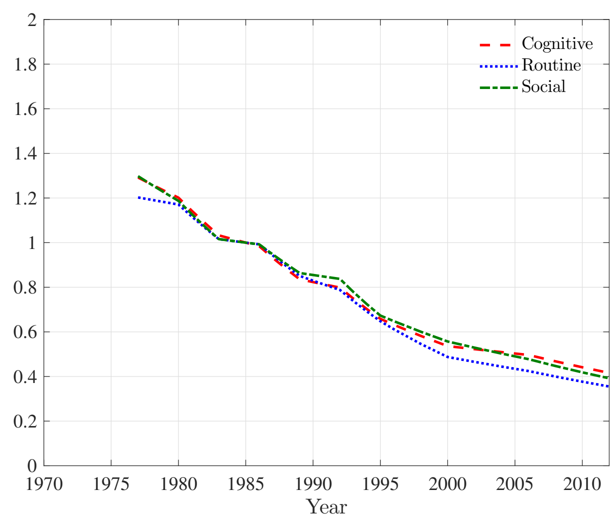


(b) Different Occupation in t' ($o_t = o_{t-2} \neq o_{t'}$)

Figure E-13: μ_t^o/μ_{1985}^o implied by 2SLS estimates for occupation-stayers: All experience levels



(a) Same Occupation in t' ($o_t = o_{t-2} = o_{t'}$)



(b) Different Occupation in t' ($o_t = o_{t-2} \neq o_{t'}$)

Figure E-14: μ_t^o/μ_{1985}^o implied by 2SLS estimates for occupation-stayers: 21–40 years of experience

F HRS Data and Results

F.1 Data Description

We use data from the Health and Retirement Study (HRS), a national U.S. panel survey of individuals over age 50 and their spouses.[69] We use data from six cohorts incorporated over time, beginning with the first cohort surveyed in 1992. (New cohorts of individuals were added in 1998, 2004, 2010, and 2016.)[70] The survey has been fielded every two years since 1992, and it provides information about demographics, income, and cognition, making it ideal data for the purpose of our study. Because one of the cognitive tests (word recall) in 1992 and 1994 differs from that of later years, we use data collected from 1996 to 2018.[71]

The HRS records the respondent's and spouse's wage rates if they are working at the time of the interview. We use the hourly wage rate, deflating nominal values to 1996 dollars using the Consumer Price Index.[72] The HRS also provides various cognitive functioning measures. We use word recall in our analysis, but report below on its correlation with two other measures available in several years: serial 7's and quantitative reasoning. Table F-1 provides a brief summary of these measures. The word recall test evaluates the memory of the respondents by reading a list of 10 words and asking them to recall immediately (immediate recall) and after a delay of about 5 minutes (delayed recall). We sum the number of words the respondent recalled in the two tasks and obtain a score of 21 different values. The serial 7's test asks the respondent to subtract 7 from the previous number, starting with 100 for five trials. This test score is the number of trials that the respondent answered correctly, and it has 6 different values. Quantitative reasoning consists of three simple arithmetic questions assessing the numeracy of the respondent. We construct a test score based on the answers and the resulting score ranges from 0 to 4. Additional details about these measures are provided below.

Our sample is restricted to age-eligible (i.e. born in eligible years when first interviewed) men. We use observations when men are ages 50–70 if their potential labor market experience is between 30 and

[69] More precisely, the sample does include some individuals age 50. For example, someone from the original cohort (born in 1931-1941) who was born late in 1941 may have been age 50 at the date of their first interview in 1992 if they were interviewed earlier in the calendar year.

[70] The HRS sample was built up over time. The initial cohort consisted of persons born between 1931 and 1941 (aged 51 to 61 at first interview in 1992). The Asset and Health Dynamics Among the Oldest Old (AHEAD) cohort, born before 1924, was added in 1993. Given the ages of respondents from this cohort (over 70 by 1994), it is excluded from our analysis. In 1998, two new cohorts were enrolled: the Children of the Depression (CODA) cohort, born 1924 to 1930, and the War Baby (WB) cohort, born 1942 to 1947. Early Baby Boomer (EBB, born 1948 to 1953) cohort was added in 2004, Mid Baby Boomer (MBB, born 1954 to 1959) cohort was added in 2010, and Late Baby Boomer (LBB, born 1960 to 1965) cohort was added in 2016. In addition to respondents from eligible birth years, the survey interviewed the spouses of married respondents or the partner of a respondent, regardless of age.

[71] The word recall test contained a list of 20 words in 1992 and 1994, while it was reduced to 10 words in later years.

[72] https://www.bls.gov/cpi/research-series/home.htm#CPI-U-RS20Data

Table F-1: Summary of Cognitive Measures

	Meant to measure	Number of values	Available years
Word recall	Memory	21 (0–20)	1996–2018
Serial 7's	Numeracy	6 (0–5)	1996–2018
Quantitative reasoning	Numeracy	5 (0–4)	2002–2018

50 years.[73] In estimation, we use non-imputed wages and cognitive measures only. The sample contains 10,151 individuals and 43,096 person-year observations.

Our sample consists of 70% white, 16% black, 11% Hispanic, and 4% other races with an average age of 60 years. We create five education categories based on years of completed schooling: 0-11 years (less than high school graduate), 12 years (high school graduate), 13-15 years (some college), 16 years (college graduate), and 17 or more years (above college). In our sample, 15% had less than 12 years of schooling, 30% had 12 years of schooling, 25% had some college, 15% completed college, and 15% had more than 16 years of schooling. Table F-2 shows the mean and the standard deviation of cognitive scores, the log hourly wage, and the log wage residual, along with correlations between these variables. The correlations between test scores range from nearly 0.29 to 0.48. All three test scores exhibit similar positive correlations with log wages and log wage residuals. Most relevant to our analysis, word recall has a correlation of 0.21 with log wages and 0.07 with the log wage residual.

Table F-2: Mean, standard deviation (S.D.), and correlations between cognitive scores and log wage residuals

	Num. of obs.	Mean	S.D.	Correlations			
				WR	S7	QR	w
Word recall (WR)	39,222	10.33	3.15	1.00			
Serial 7's (S7)	39,865	3.91	1.46	0.29	1.00		
Quantitative reasoning (QR)	17,828	2.03	1.26	0.33	0.48	1.00	
Log wage residual (w)	23,027	0.00	0.75	0.07	0.05	0.08	1.00
Log wage ($\ln W$)	23,042	2.45	0.84	0.21	0.20	0.27	0.89

F.2 Detailed Description of Cognitive Measures

Word recall. The HRS contains two separate tasks to assess respondent's memory: immediate word recall and delayed word recall. During the interview, the interviewer read a list of 10 nouns to the

[73] We use age recorded at the end of the interview (sometimes interviews occur over multiple dates). Potential experience is defined as age minus 6 minus years of schooling.

respondent and asked the respondent to recall as many words as possible from the list in any order. After approximately 5 minutes of answering other survey questions, the respondent was asked to recall the nouns previously presented. We construct a single measure which is the sum of the number of nouns that the respondent recalled in the two tasks. This measure ranges from 0 to 20.

Serial 7's. This test asks the respondent to subtract 7 from the prior number, beginning with 100 for five trials. Correct subtractions are based on the prior number given, so that even if one subtraction is incorrect subsequent trials are evaluated on the given (perhaps wrong) answer. This test score ranges from 0 to 5.

Quantitative reasoning. In the 2002 wave of HRS, three questions were added to the core survey to assess respondents' numerical ability:

1. "Next I would like to ask you some questions which assess how people use numbers in everyday life. If the chance of getting a disease is 10 percent, how many people out of 1,000 would be expected to get the disease?"

2. "If 5 people all have the winning numbers in the lottery and the prize is two million dollars, how much will each of them get?"

3. "Let's say you have $200 in a savings account. The account earns ten percent interest per year. How much would you have in the account at the end of two years?"

We construct a single measure called quantitative reasoning using the answers from these three questions. For each of the first two questions, the respondent earns 1 point if the answer is correct and 0 otherwise. For the last question, the respondent earns 2 points if the answer is correct; 1 point if the respondent used 10% as a simple interest rate rather than a compound interest rate (i.e., answered 240 instead of 242); otherwise, he earns 0 points. The quantitative reasoning measure is the sum of points earned on all three questions, ranging from 0 to 4.

G Survey of Income and Program Participation (SIPP) linked with W-2 Forms

This appendix describes data from Internal Revenue Service (IRS)/Social Security Administration (SSA) W-2 Forms linked with the Survey of Income and Program Participation (SIPP), referred to as the Gold Standard File (GSF) by the Census Bureau (U.S. Census Bureau, 2018). These data include the full SSA history of annual earnings (i.e., wage and salary) for all linked respondents from 1951 to 2011.[74] Because we use annual earnings from administrative records, annual hours of work and hourly wages are not available.

Our analysis is based on 16–69 year-old, US-born white men who could be linked to any of nine SIPP panels (1984, 1990, 1991, 1992, 1993, 1996, 2001, 2004, and 2008). The highest level of education achieved at the time of survey (asked only once in each panel) is available in 5 categories: no high school degree, high school degree, some college, college degree, and graduate degree. We map these categories to 10, 12, 14, 16, and 18 years of completed schooling in order to calculate potential experience (age - years of education - 6). Since some individuals were still young and unlikely to have completed their schooling at the time of survey, we exclude those who were under 30 years old or were enrolled in school when their education level was measured.

We focus mainly on results using Detailed Earnings Records (DER), which are uncapped and available from 1978 onward; however, we also take advantage of Summary Earnings Records (SER) available since 1951, which report earnings capped at the FICA taxable maximum. We work with log earnings residuals constructed as with the PSID and restrict observations to years when individuals were no longer enrolled in school. We trim the top and bottom 1% of DER-based earnings within year and college/non-college status by five-year experience cells, and residualize log DER-based earnings by regressing on experience indicators and interactions between education indicators and a third order polynomial in experience, separately by year and college/non-college status. Log SER-based earnings – used only as instruments in our analysis – are residualized by subtracting median values conditional on year, education, and five-year experience cells.

Based on a worker's primary job (i.e., the job with the highest earnings), the Census Bureau classified workers into 24 occupation categories each survey wave. Table G-1 reports these occupation codes, along with our 3-category grouping of occupations (cognitive, manual, and routine). Since respondents can report different occupations in each of 3 survey waves each year, we define occupation stayers between two years as those who reported any occupation in both years.

[74] This analysis was first performed using the SIPP Synthetic Beta (SSB), while final results were obtained by Census Bureau staff using the SIPP Completed Gold Standard Files. See Reeder, Stanley, and Vilhuber (2018) and Benedetto, Stanley, and Totty (2018) for additional details on the data.

Table G-1: SIPP/W-2 Occupation Codes and 3-Category Grouping

Code	Occupation	3-Category Grouping
1	Management	Cognitive
2	Business and financial operations	
3	Computer and mathematical	
4	Architecture and engineering	
5	Life, physical, and social science	
6	Community and social service	
7	Legal	
8	Education, training, and libraries	
9	Arts, design, entertainment, sports, and media	
10	Healthcare practitioner and technical	
11	Healthcare support	Manual
12	Protective service	
13	Food prep and service	
14	Building and grounds cleaning and maintenance	
15	Personal care and service	
16	Sales	Routine
17	Office and administrative support	
18	Farming, fishing, and forestry	Not classified
19	Construction and extraction	Routine
20	Installation, maintenance, and repairs	
21	Production	
22	Transportation	
23	Material moving	
24	Military	Not classified

Figure G-1 reports log earnings inequality, along with between-group and within-group (residual) inequality, based on DER wage measures in the SIPP/W-2. The general trends are qualitatively similar to those for the PSID reported in Figure 1; although, the variance of total log earnings inequality and residual inequality is notably higher than their counterparts for log wages in the PSID.

Figure G-2 shows $\mathrm{E}\left[w_t | w_b \in Q_b^j\right]$ for different t years where Q_b^j reflects quartile j in 'base' year b, while Figure G-3 shows residual autocovariances $\mathrm{Cov}(w_t, w_b)$ over years $t \geq b + 6$ for fixed base year b. Both figures are based on samples of non-college and college men with 21–25 years of experience in each base year, b. Together, these indicate declines in the return to skills over the late-1980s and 1990s, consistent with our PSID-based results.

Tables G-2 and G-3 report 2SLS estimates of skill return growth rates using SER- and DER-based lagged log earnings residuals (w_{t-7}), respectively, as instruments. (See Figure 21 in the paper.) Corresponding standard errors and sample sizes are also reported.

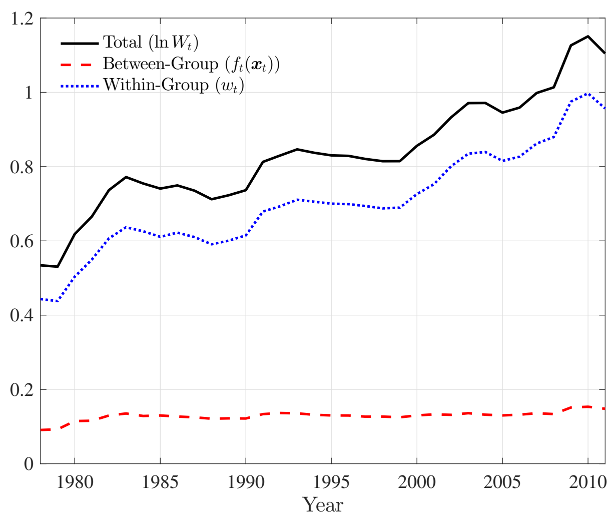
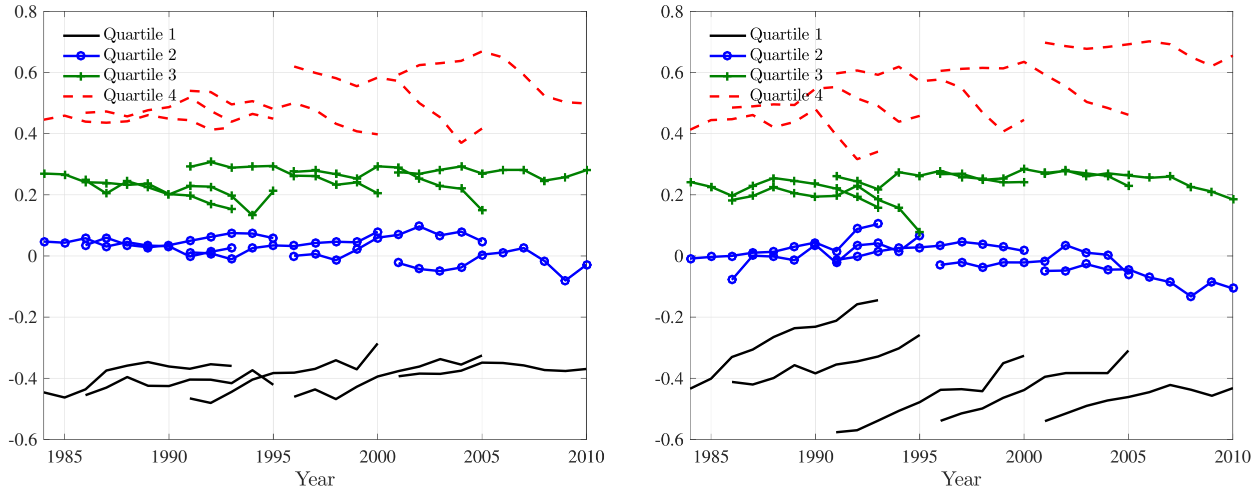
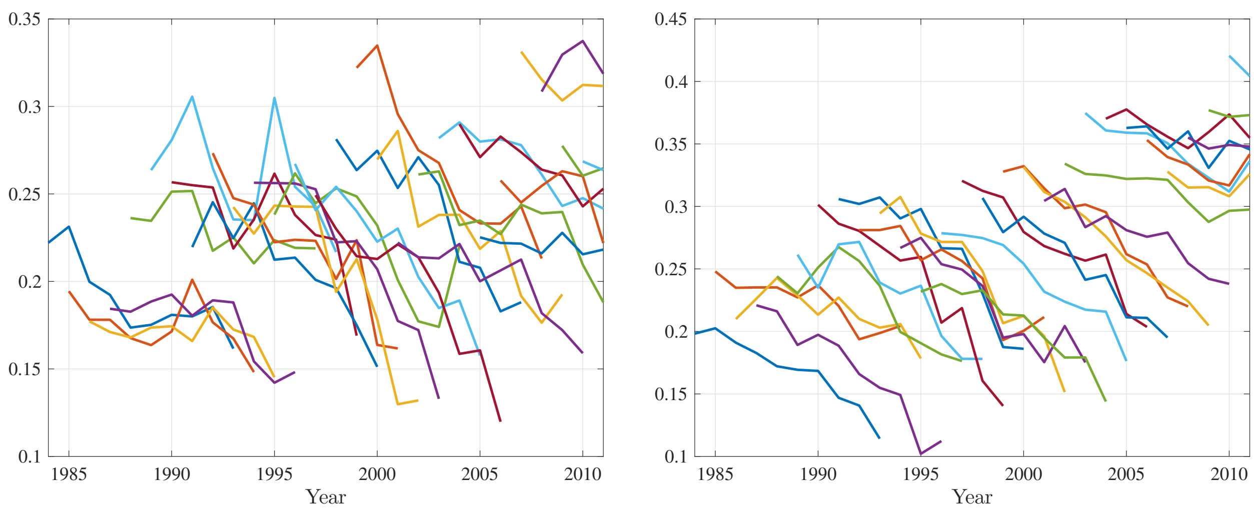


Figure G-1: Between- and within-group variances of log earnings, ages 16–64 with 5–40 years of experience (SIPP/W-2)



(a) Non-College (b) College

Figure G-2: Average predicted log earnings residuals by baseline residual quartile, 21–25 years of experience in base year (SIPP/W-2)



(a) Non-College (b) College

Figure G-3: Autocovariances for log earnings residuals, 21–25 years of experience in base year (SIPP/W-2)

Table G-2: 2SLS estimates of $\Delta\mu_t/\mu_{t-1}$ (instrument: w_{t-7} SER earnings), 32–40 years of experience in year t (SIPP/W-2)

Year	Non-College			College		
	Estimate	Standard Error	Observations	Estimate	Standard Error	Observations
1979	0.024	0.059	3,600	0.122	0.075	1,900
1980	-0.037	0.043	3,700	0.047	0.063	2,100
1981	0.053	0.051	3,800	-0.135	0.077	2,200
1982	0.077	0.045	3,800	0.015	0.063	2,400
1983	-0.044	0.043	3,700	-0.002	0.059	2,500
1984	-0.024	0.041	3,700	0.010	0.069	2,600
1985	-0.068	0.039	3,600	-0.024	0.058	2,700
1986	-0.039	0.040	3,600	0.010	0.050	2,900
1987	-0.139	0.037	3,700	-0.041	0.032	3,000
1988	-0.043	0.038	3,800	-0.101	0.030	3,100
1989	0.020	0.033	3,800	-0.043	0.043	3,100
1990	-0.078	0.033	3,800	-0.009	0.044	3,200

Notes: Reports coefficient estimates from 2SLS regression of Δw_t on w_{t-1} using w_{t-7} as an instrument. The number of observations is rounded to the nearest 100 due to confidentiality.

Table G-3: 2SLS estimates of $\Delta\mu_t/\mu_{t-1}$ (instrument: w_{t-7} DER earnings), 32–40 years of experience in year t (SIPP/W-2)

Year	Non-College			College		
	Estimate	Standard Error	Observations	Estimate	Standard Error	Observations
1985	-0.060	0.029	3,800	-0.026	0.032	2,900
1986	-0.072	0.030	3,800	-0.101	0.026	3,100
1987	-0.074	0.030	3,900	-0.049	0.027	3,200
1988	-0.055	0.037	4,000	-0.084	0.028	3,400
1989	0.016	0.030	3,900	-0.025	0.028	3,400
1990	-0.065	0.029	4,000	0.012	0.030	3,500
1991	0.010	0.032	4,000	-0.009	0.028	3,600
1992	-0.061	0.026	4,100	-0.096	0.025	3,800
1993	-0.073	0.023	4,100	-0.040	0.024	3,900
1994	-0.025	0.025	4,200	-0.055	0.027	4,200
1995	0.012	0.036	4,300	-0.045	0.023	4,400
1996	-0.018	0.028	4,300	-0.094	0.022	4,700
1997	-0.058	0.024	4,400	-0.020	0.021	5,000
1998	-0.088	0.019	4,400	-0.073	0.020	5,500
1999	-0.057	0.023	4,400	-0.084	0.020	6,000
2000	-0.107	0.028	4,400	-0.048	0.022	6,600
2001	-0.102	0.029	4,400	-0.039	0.018	7,100
2002	-0.045	0.029	4,300	-0.034	0.021	7,700
2003	-0.085	0.028	4,400	-0.045	0.018	8,300
2004	-0.010	0.029	4,800	-0.019	0.016	8,900
2005	-0.052	0.024	5,000	-0.064	0.015	9,400
2006	-0.002	0.024	5,100	-0.011	0.014	9,800
2007	-0.031	0.023	5,400	-0.023	0.015	10,000
2008	-0.011	0.022	5,600	-0.014	0.015	10,500
2009	0.006	0.024	5,700	-0.042	0.015	10,500
2010	-0.001	0.024	5,700	0.004	0.014	10,000
2011	-0.048	0.023	5,700	0.003	0.013	9,900

Notes: Reports coefficient estimates from 2SLS regression of Δw_t on w_{t-1} using w_{t-7} as an instrument. The number of observations is rounded to the nearest 100 due to confidentiality.