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

Lance Lochner
University of Western Ontario

Youngmin Park
Bank of Canada

Youngki Shin
McMaster University

February 16, 2025

Abstract

Economists disagree about the factors driving the substantial increase in residual wage inequality in the US over the past few decades. To identify changes in the returns to unobserved skills, we make a novel assumption about the lifecycle dynamics of skills, which we validate using data on test score dynamics for older workers in the HRS. Using survey data from the PSID and administrative data from the IRS and SSA, we estimate that the returns to unobserved skills *declined* substantially in the late-1980s and 1990s despite an increase in residual inequality. Accounting for firm-specific pay differences yields similar results. Extending our framework to consider occupational differences in returns to skill and multiple unobserved skills, we further show that skill returns display similar patterns for workers employed in each of cognitive, routine, and social occupations. Finally, our results suggest that increasing skill dispersion, driven by rising skill volatility, explains most of the growth in residual wage inequality since the 1980s.

*For valuable comments on this line of research, we thank Pat Bayer, Magne Mogstad, Terry Moon, Fabrizio Perri, Gianluca Violante and Thomas Lemieux as well as numerous seminar and conference participants. Lochner and Shin gratefully acknowledge generous support from SSHRC. 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.

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]

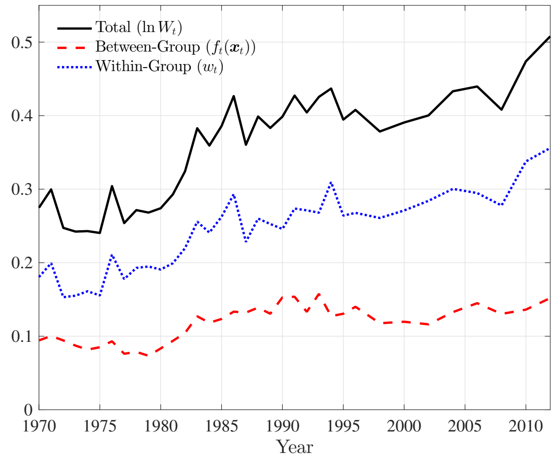


Figure 1: Between- and Within-Group Variances of Log Wages

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]

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 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. The extent to which the rise in residual inequality

[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). More recent task-based models of the labor market explore the influence of automation and globalization on wage and employment inequality between and within groups by altering the demand for skills (e.g., Autor, Levy, and Murnane, 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2022; Acemoglu and Loebbing, 2022).

reflects an increase in returns to unobserved skills, growing unobserved skill inequality, or increased wage volatility unrelated to skills is critical to understanding both the economic causes and welfare consequences of rising inequality. This paper develops 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.[3] 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.[4] 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, challenging measurement issues 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 the wages of workers in their late-20s, thereby ignoring early-career lifecycle skill accumulation that may vary across workers and over time. The vast majority of studies taking a task-based approach do not use individual-level data on skills or job tasks, implicitly assuming that worker skills and tasks within each occupation are time-invariant.[5] As a result, these studies attribute all time variation in wages across occupations to changes in the returns to skills/tasks.

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 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 year should continue, on average, to earn a high wage many years later (even after

[3] A broader literature (see, e.g., Cawley, Heckman, and Vytlacil, 2001; Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2005; Heckman, Stixrud, and Urzua, 2006) takes advantage of direct skill measurements to estimate log wages (or earnings) as a function of latent factors or skills; however, most studies follow a single cohort over time and do not distinguish between lifecycle skill growth associated with age and evolving skill returns over time.

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

[5] Recent studies call these assumptions into question. 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., while Spitz-Oener (2006) shows that most task changes in Germany over the 1980s and 1990s occurred within occupations.

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, with those predicted differences growing (shrinking) as the returns to skill rise (fall).

Categorizing workers based on their log wage residual quartile in 3 different base years, Figure 2 reports their average residuals 6–20 years later. 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.[6] The strong convergence in average residuals (conditional on base-year quartiles) over the late-1980s and 1990s indicates that either the persistence of skill differences across workers fell (e.g., greater skill depreciation at the top of the skill distribution relative to the bottom) or that the returns to skill declined over this period. This paper shows that the latter explanation is most consistent with a broad array of evidence: the returns to unobserved skill *fell* sharply over the late-1980s and 1990s. By contrast, returns were more stable in earlier and later years.

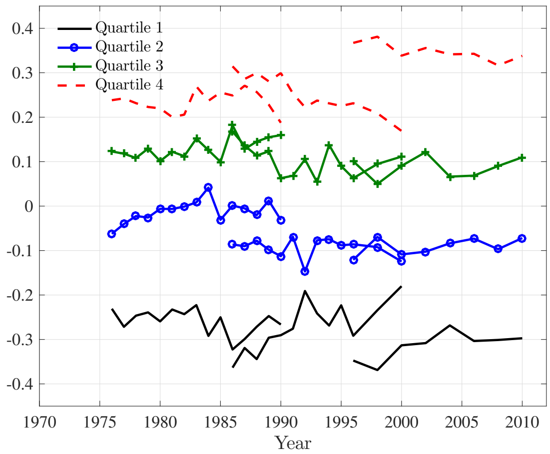


Figure 2: Average Predicted Log Wage Residuals by Baseline Residual Quartile

Notes: Each line reflects the average log wage residual in year $t = b + 6, \ldots, b + 20$ conditional on log wage residual quartile in base years $b = 1970$, 1980, or 1990.

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 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 this assumption, as do several other specification tests based on the autocovariance structure for residuals. We also show that this assumption, along with the assumption of limited persistence in non-skill wage shocks, can be relaxed; although, estimation then requires a more general moment-based approach that

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

exploits more of the autocovariance structure for log wage residuals. Once the returns to skill have been estimated (from, e.g., older workers), it is straightforward to estimate the distributions 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, so our approach is applicable 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 slightly stronger for non-college workers, consistent with the recent literature on polarization (Autor, Levy, and Murnane, 2003; 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 most of 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 firms and/or occupations, as well as the potential for different trends in the returns to heterogeneous skills (see, e.g., Acemoglu and Autor (2011), Sanders and Taber (2012), Kline (2024), and Woessmann (2024) for recent reviews). We first show that heterogeneity in firm-specific pay cannot account for our estimated declines in skill returns. We then 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 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

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

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 previous studies. 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. Section 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 firms, occupations, and multiple skills. 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 $i = 1, \ldots, N$ in period $t = \underline{t}, \ldots, \bar{t}$, $f_t(\boldsymbol{x}_{i,t})$ 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 $\mathrm{E}[\theta_{i,t}|\boldsymbol{x}_{i,t}] = \mathrm{E}[\varepsilon_{i,t}|\boldsymbol{x}_{i,t}] = 0$. 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 include measurement error.[8] Note that average unobserved skill growth, which may vary by observable characteristics, is subsumed in changes in $f_t(\boldsymbol{x}_{i,t})$.[9] 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,

[8] 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.

[9] The assumption of separability between $\boldsymbol{x}_{i,t}$ and $\theta_{i,t}$ is both common and convenient, though not necessary. One can condition our analysis on $\boldsymbol{x}_{i,t}$. Empirically, we separately study non-college and college educated workers.

reflected in $\mu_t > 0$, may evolve quite differently over time from the returns to observed measures of skills, reflected in $f_t(\cdot)$.[10] 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, $\eta_{i,t}$, are serially uncorrelated and are uncorrelated with other observed variables $\boldsymbol{x}_{i,t}$, unobserved skills $\theta_{i,t}$, and non-skill wage shocks $\varepsilon_{i,t}$. It is useful to define test score residuals, $\tilde{T}_{i,t} \equiv T_{i,t} - g_t(\boldsymbol{x}_{i,t}) = \tau\theta_{i,t} + \eta_{i,t}$. For a neater representation, we drop subscript i hereafter when it is clear from context.

2.1 Prior Assumptions in the Literature

By equating the increase in residual inequality with an increase in skill returns, Juhn, Murphy, and Pierce (1993) assume that the cross-sectional distributions of unobserved skills and non-skill shocks are time-invariant.[13] Lemieux (2006) relaxes this assumption, instead assuming only that the variance of skills conditional on observed characteristics is time-invariant, i.e.,

[10]There are two ways to think about $\theta_{i,t}$: it may reflect variation in a homogeneous skill (produced through schooling and other investments) conditional on educational attainment (and other factors in $\boldsymbol{x}_{i,t}$), or it may reflect an unobserved skill completely distinct from any skills produced through schooling. Thus, μ_t could reflect returns to a single skill, with estimated 'returns' to schooling over time, $f_t(\boldsymbol{x}_t)$, capturing returns to that skill as well as differences in the average amount of skill produced through schooling. Or, μ_t could reflect returns to a distinct skill not produced through education. The latter is a more natural interpretation when returns to education follow a different time path from μ_t; however, such differences could also be explained by changes in how education produces skills. We generally take the interpretation of Juhn, Murphy, and Pierce (1993) and most subsequent studies, treating $\theta_{i,t}$ as its own unique unobserved skill.

[11]Equations (1) and (2) imply wage *levels* that are non-linear in unobserved skill, 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 $g_t(\cdot)$ 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 $\tilde{g}_t(\boldsymbol{x}_{i,t}) + \theta_{i,t}$, in which case $g_t(\boldsymbol{x}_{i,t}) = \tau\tilde{g}_t(\boldsymbol{x}_{i,t})$).

[13]To support this assumption, they show that growth in the residual variance from 1963 to 1989 is similar when following cohorts or experience groups. However, Lochner, Park, and Shin (2025) show that this may indicate stability of skill variation across cohorts (within periods), but it does not say anything about the evolution of skill variation or returns over time. Equal within-cohort and within-experience growth in the residual variance is consistent with growth in skill returns, in the variance of skill growth, or in the variance of non-skill wage shocks.

$\mathrm{Var}(\theta_t|\boldsymbol{x}_t = \boldsymbol{x}) = \sigma^2(\boldsymbol{x})$, $\forall(t, \boldsymbol{x})$. Re-weighting $\mathrm{Var}(w_t|\boldsymbol{x}_t)$ each year to account for composition shifts in the labor market (i.e., an aging and increasingly more educated workforce), he estimates that the variance of skills increased over time, while changes in the returns to skill were modest.[14]

Lochner, Park, and Shin (2025) show that Lemieux's assumption can be tested using repeated cross-section or panel data with the same skill measure over time. 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. Applying this test to measures of cognitive memory for men with 30–50 of years of experience in the 1996–2018 HRS, Lochner, Park, and Shin (2025) strongly reject time-invariance of $\mathrm{Var}(T_t|\boldsymbol{x}_t)$, indicating significant changes in within-group skill inequality over time.

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 measurement error, these estimates do not necessarily identify the evolution of returns to contemporaneous skills, μ_t, because they are confounded by cross-cohort changes in the relationship between adolescent skills and adult skills. To see this, notice that OLS regression of log wage residuals in year $t + \ell$ on test residuals in year t identifies

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

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 period $t + \ell$ up to the test score scale: $\mu_{t+\ell}/\tau$. Unfortunately, identification is complicated by the terms in brackets related to the dynamics of unobserved skill (between the date skills vs. earnings are measured) and test measurement error.

Following a single cohort over time (i.e., varying ℓ for fixed t) confounds systematic heterogeneity in skill growth with changes in the returns to skill, as reflected in the "Skill Dynamics" term.[15] 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

[14]Lemieux's re-weighting approach ignores time variation in $\mathrm{Var}(\varepsilon_t)$; although, he provides a separate analysis that shows increasing measurement error in wages over time (in the March Current Population Survey).

[15]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.

$t' \approx 1997$, respectively. They use the same cognitive test measure, AFQT, in both periods with wages reported roughly 10–15 years after the tests were administered.[16] 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 the 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. Lochner, Park, and Shin (2025) document 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.

Even ignoring these concerns, Castex and Dechter (2014) only 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.

Deming (2017) applies this same approach to estimate the returns to social skills over time; however, he faces an additional challenge due to the lack of a consistent measure of social skills across NLSY cohorts. In using different measurements for each cohort, he is unable to 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 sales, professional, and management occupations.[17]

2.2 Identification using Panel Data on Wages

Previous efforts to estimate returns to unobserved skills rely on assumptions about the stability of skill distributions or early skill dynamics across cohorts. Using panel data, we introduce a 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 unrelated to skills. Although

[16] 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.

[17] Deming (2017) normalizes his available measures of social skills to have a standard deviation of one in both cohorts; however, this does not eliminate bias coming from differences in τ across measurements. See Lochner, Park, and Shin (2025) for details. Edin et al. (2022) benefit from more consistent measures of cognitive and social/leadership skills across cohorts in Sweden; however, they still rely on assumptions that early-career skill dynamics are identical across cohorts and that measurement reliability ratios are stable over time. Their estimates suggest modest reductions in returns to cognitive skills and increases in returns to social/leadership skills.

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.[18] We make the following assumption (letting Δ reflect the first-difference operator), which imposes restrictions on the lifecycle dynamics of skill, the interaction between skills and non-skill shocks, and the persistence of non-skill shocks.

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 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 persistence. 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 $\bar{t} - \underline{t} \geq k + 1$.

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 $t - t' \geq k$,

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

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

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

□

[18] 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.

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).

In general, 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.[19]

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.

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 log wage residuals conditional on prior residual quartiles over the late-1980s and 1990s shown in Figure 2 indicates a sharp decline in the

[19] Lochner, Park, and Shin (2025) discuss conditions under which different cohorts may be used to eliminate this bias, enabling estimation of skill returns over the full sample period. Given the lengthy period covered by many panel data sets, we focus on identification and estimation for periods $t \geq \underline{t} + k$.

returns to skill over those years.

In Section 3.5, we relax condition (i) of Assumption 1 related to the dynamics of skills. Here, we briefly make a few 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 $w_{i,t} = \mu_t(\theta_{i,t} + \varepsilon_{i,t})$. Whether transitory shocks are assumed to be related or unrelated to skills 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 analysis relies 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. Lochner, Park, and Shin (2025) show 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. They further show that μ_t profiles estimated when assuming ε_t contains an AR(1) component are quite similar to those presented in Section 3 obtained under Assumption 1.

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

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

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 $\mu_{\underline{t}}, \ldots, \mu_{\bar{t}}$ would be identified along with $\mathrm{Var}(\theta)$. 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).[20] Our empirical analysis below pays particular attention to experienced workers, estimating returns to skill based on this subgroup.

[20] 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., skill depreciation may systematically differ across workers with different skill levels). With panel data on skill measurements, it is possible to test whether $\text{Cov}(\Delta_2\theta_{t+2}, \theta_{t-\ell}) = 0$ using the following moments:

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

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.) Similar moments using log wage residuals $w_{t-\ell}$ as additional instruments allow us to further test whether $\text{Cov}(\Delta_2\theta_{t+2}, \varepsilon_{t-\ell}) = 0$.

In Appendix D.2, we implement these tests (estimating ϱ using GMM) for various k values using measures of cognitive memory for a sample of men with 30–50 years of experience in the 1996–2018 HRS. Using $\tilde{T}_{t-\ell}$ and $w_{t-\ell}$ lags with $\ell = 6, 8$ as instruments, estimated ϱ is only 0.018, insignificantly different from zero. Estimates 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.[21] 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.

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.

[21]While 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 $\text{Cov}(\theta_t, \nu_{t'}) = 0$ for all $t' \geq t + 1$), then these estimates imply ρ values of 0.97–1.02, very close to a random walk. For all lags, estimates are nearly identical whether or not we include $w_{t-\ell}$ as instruments in addition to $\tilde{T}_{t-\ell}$.

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 C 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

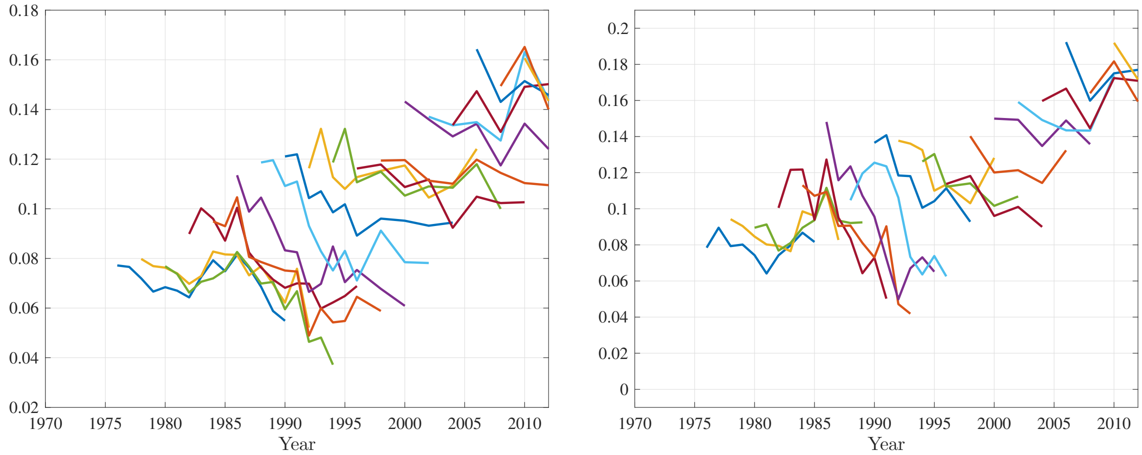
Figure 1 documents that log wage inequality has increased substantially since 1970, with particularly strong growth in the early-1980s and after 2000. The evolution of residual inequality closely mirrors this pattern, explaining a larger share of the total variance than the between-group variance.

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. 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 $\text{Cov}(w_b, w_t)$ for $t = b+6, \ldots, b+20$ with each line reporting autocovariances for a different 'base' year b and 15 subsequent years.[22] 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 $\text{Cov}(w_b, w_t) = \mu_t \Omega_b$ (see equation (5)) and following each line over t is directly informative about the evolution of μ_t. The sharply declining autocovariances over the late-1980s and 1990s (regardless of base year) suggest that the returns to unobserved skill fell substantially over this period, while the more stable autocovariances during earlier and later years are consistent with

[22] Lochner, Park, and Shin (2025) show that attrition due to non-response or aging/retirement does not affect these autocovariance patterns.

more modest changes in returns during those years. As discussed in Section 4, the upward-shifting lines beginning in the 1980s signal a period of rising skill dispersion. Figure 3(b) reveals very similar autocovariance patterns when restricting the sample to men with 21+ years of experience in years $t \geq b + 6$. (Similar patterns are also evident for less-experienced men.)



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

Figure 3: Autocovariances for Log Wage Residuals

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. Do these trends reflect differences in the evolution of returns to skill by educational attainment, as is often implicitly assumed? Figure 5 shows that long-autocovariances declined sharply during the late-1980s and 1990s (given any base year) for both non-college- and college-educated men, indicating qualitatively similar declines in the returns to skill for both education groups over that period. These autocovariance patterns are central to our empirical approach and underpin our main findings.

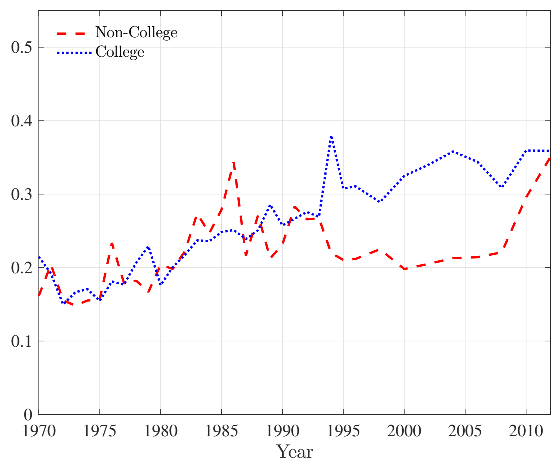


Figure 4: Variance of Log Wage Residuals by Education

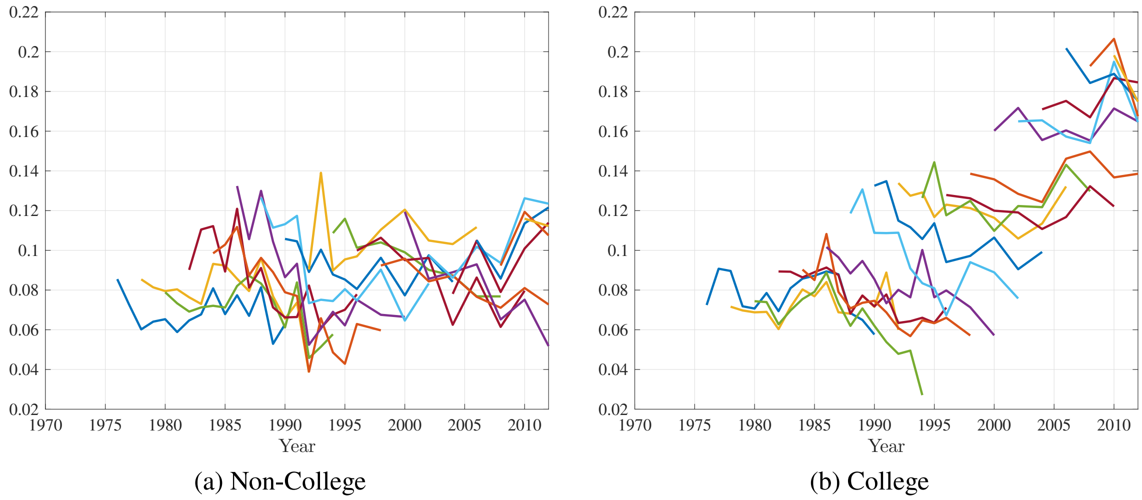


Figure 5: Autocovariances for Log Wage Residuals by Education, All Experience levels

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 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:

$$\Delta_2 w_{i,t} = \left(\frac{\Delta_2\mu_t}{\mu_{t-2}}\right) w_{i,t-2} + \left[\mu_t(\Delta\theta_{i,t-1} + \Delta\theta_{i,t}) + \varepsilon_{i,t} - \frac{\mu_t}{\mu_{t-2}}\varepsilon_{i,t-2}\right]. \tag{9}$$

Under Assumption 1, we can obtain consistent estimates of $\Delta_2\mu_t/\mu_{t-2}$ by estimating equation (9) via 2SLS using lags $w_{i,t'}$ for $t' \leq t-k-2$ as instrumental variables.

Table 1 reports 2SLS estimates of skill return growth rates using equation (9) for years t covering 1979–1995, assuming that skill return growth rates are constant within two- or three-year periods. Assuming $k = 6$, we use $(w_{i,t-8}, w_{i,t-9})$ as instruments. Table 2 reports 2SLS estimates for the later years of the PSID (t covering 1996–2012) when observations become biennial.[23] In all specifications, the instruments are 'strong' with very large first-stage F-statistics.

Panel A of Tables 1 and 2 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 C-1 and C-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

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

Table 1: 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.

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 estimated by Castex and Dechter (2014).

Since our model is overidentified with multiple instruments, Appendix C.3 examines the validity of Assumption 1 by testing the exogeneity of sufficiently lagged residuals as instruments in equation (9). We cannot reject exogeneity at conventional levels in any year, consistent with Assumption 1 (for $k = 6$). By contrast, the appendix shows 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 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.

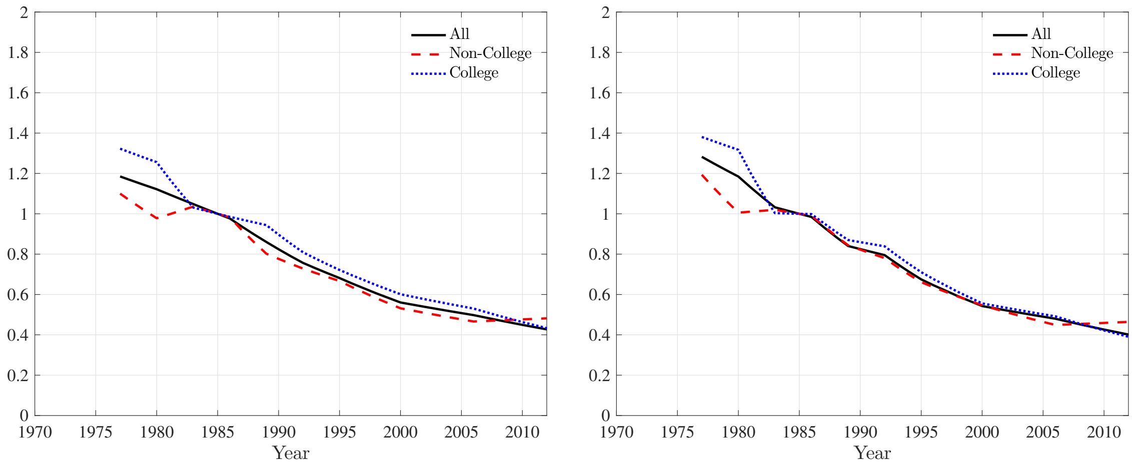
3.4 MD Estimation of Skill Returns using Long Autocovariances

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')

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

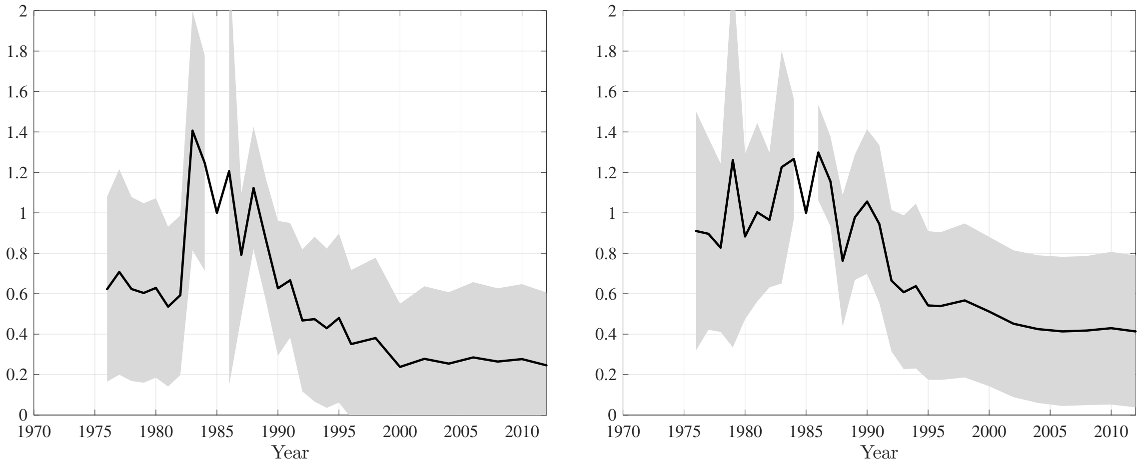
	1996–2000	2002–2006	2008–2012
	A. All men		
$\Delta_2\mu_t/\mu_{t-2}$	-0.075*	-0.039	-0.050
	(0.025)	(0.028)	(0.027)
Observations	2,122	2,129	1,968
1st stage F-Statistic	369.09	344.25	341.36
	B. Non-college men		
$\Delta_2\mu_t/\mu_{t-2}$	-0.087*	-0.043	0.011
	(0.043)	(0.047)	(0.075)
Observations	862	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 ($\mu_{1985} = 1$)



(a) Non-College

(b) College

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

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 (1942–1951,..., 1972–1981).[24] Table B-1 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 B for details.

Figure 7 reports MD estimates of μ_t separately for non-college and college men. (Shaded areas in this and subsequent figures reflect 95% confidence intervals.) Like their 2SLS counterparts, MD estimates of μ_t indicate substantial declines in the returns to skill over the late-1980s and 1990s, contrasting 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 relatively stable prior to 1985. They also suggest that returns fell by at least 40% for non-college men and 20% for college men. See Figure C-1 for estimated $\Omega_{C,t'}$.

3.5 Relaxing our Assumption on Skill Growth

Our analysis, thus far, has relied on the assumption 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) and (iii) of Assumption 1 to $\text{Cov}(\theta_t, \varepsilon_{t'}) = 0$ for all t, t', while maintaining condition (iv) (i.e., limited persistence of non-skill shocks). In this case,

[24] We estimate $\Omega_{C,t'} \equiv \mu_{t'} \text{Cov}(\theta_{t'+k-1}, \theta_{t'}|c \in C) + \text{Cov}(\theta_{t'+k-1}, \varepsilon_{t'}|c \in C)$ 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.

our IV estimator, using *past* or *future* residuals as instruments, converges to

$$\gamma_{t,t'} \equiv \frac{\mathrm{Cov}(\Delta w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\Delta\mu_t}{\mu_{t-1}} + \frac{\mu_t}{\mu_{t-1}} \frac{\mathrm{Cov}(\Delta\theta_t, \theta_{t'})}{\mathrm{Cov}(\theta_{t-1}, \theta_{t'})}, \quad \text{for } t'-t \geq k \text{ or } t-t' \geq k+1, \quad (10)$$

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

3.5.1 Heterogeneity in Lifecycle Skill Growth

We first explore the possibility that unobserved skill growth innovations are correlated over time as in the heterogeneous income profile (HIP) models estimated in, e.g., Lillard and Weiss (1979), Haider (2001), Baker and Solon (2003), Guvenen (2009), and Moffitt and Gottschalk (2012). Consider a more flexible process governing this skill growth heterogeneity, assuming

$$\Delta\theta_{i,t} = \lambda_t(c_i)\delta_i + \nu_{i,t}, \quad (11)$$

where δ_i is a mean zero individual-specific lifecycle growth rate factor with $\mathrm{Var}(\delta|c) > 0$, and the $\lambda_t(c) \geq 0$ terms allow for variation in systematic skill growth across time and cohorts/experience. This skill process generally violates condition (i) of Assumption 1 when $\lambda_t(c) > 0$, where the bias for IV estimator $\gamma_{t,t'}$ depends on

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

where 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 very 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 3 and 4 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

Table 3: 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 Test p-value	0.945		0.756		0.044	

Notes: * denotes significance at 0.05 level. Wald Test p-value reported for test of equivalence of coefficients on w_{t-2} across instrument years (i.e., across columns).

Table 4: 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 Test p-value	0.080		0.354		0.007	

Notes: * denotes significance at 0.05 level. Wald Test p-value reported for test of equivalence of coefficients on w_{t-2} across instrument years (i.e., across columns).

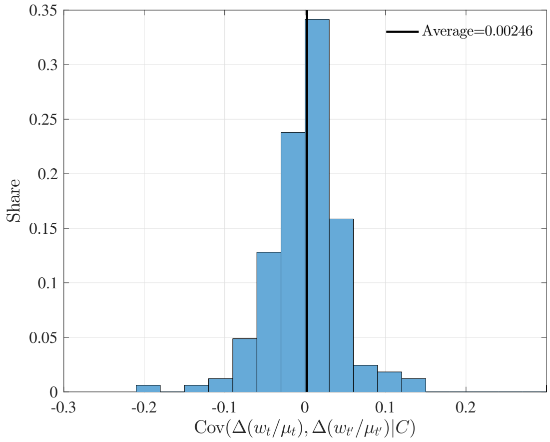


Figure 8: Distribution of $\mathrm{Cov}(\Delta(w_t/\mu_t), \Delta(w_{t'}/\mu_{t'}))$ for all (t, t', C) for Low-Experience Men

Notes: Figure reports distribution of covariances based on (t, t') satisfying $t \geq t' + 7$ for cohort groups $C \in \{1962 - 1971, 1972 - 1981\}$ when some individuals in these cohort groups had less than 21 years of experience in year t (i.e., $t \leq 1991$ for 1962–1971 cohorts or $t \leq 2001$ for 1972–1981 cohorts).

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., $\mathrm{Var}(\nu_t) > 0$).[25] Finally, we note that GMM estimates using only sufficiently lagged residuals as instruments (specification 1 of Tables 3 and 4) 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 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), baseline μ_t estimates (using only experienced workers) reported in Figure 7 can be used to scale log wage residuals to estimate

$$\mathrm{Cov}\left(\Delta\left(\frac{w_t}{\mu_t}\right), \Delta\left(\frac{w_{t'}}{\mu_{t'}}\right)\middle|c\right) = \lambda_t(c)\lambda_{t'}(c)\,\mathrm{Var}(\delta|c), \quad \text{for } t - t' \geq k + 1, \tag{12}$$

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 C-2 shows that for recent cohort groups, the covariances in equation (12) fluctuate around zero for all ages. Figure 8 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

[25] Exogeneity tests reported in Section 3.3 and Appendix C.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.

for $\text{Cov}(\Delta(w_t/\mu_t), w_{t'}|c) = \lambda_t(c)\mu_{t'}\text{Cov}(\delta, \theta_{t'}|c)$ reported in Appendix C.5 further support this conclusion.

Altogether, these results support condition (i) of Assumption 1: skill growth is not systematically related to past 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, $\phi_{i,t}$:

$$\theta_{i,t} = \psi_i + \phi_{i,t}, \qquad \text{where} \quad \phi_{i,t} = \rho_t\phi_{i,t-1} + \nu_{i,t}. \tag{13}$$

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)* $\text{Cov}(\psi, \phi_t|c) = 0$ *for all t; (ii)* $\text{Cov}(\psi, \varepsilon_{t'}|c) = \text{Cov}(\phi_t, \varepsilon_{t'}|c) = 0$ *for all* t, t'*; (iii)* $\text{Cov}(\phi_{t'}, \nu_t, |c) = \text{Cov}(\nu_{t'}, \nu_t, |c) = 0$ *for all* $t - t' \geq 1$*; (iv) for known* $k \geq 1$, $\text{Cov}(\varepsilon_t, \varepsilon_{t'}|c) = 0$ *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 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 $\tilde{\nu}_{i,t} \equiv \mu_t\nu_{i,t}$. The distinction between skill vs. non-skill persistent shocks is not important from a statistical point of view nor for identification of μ_t.

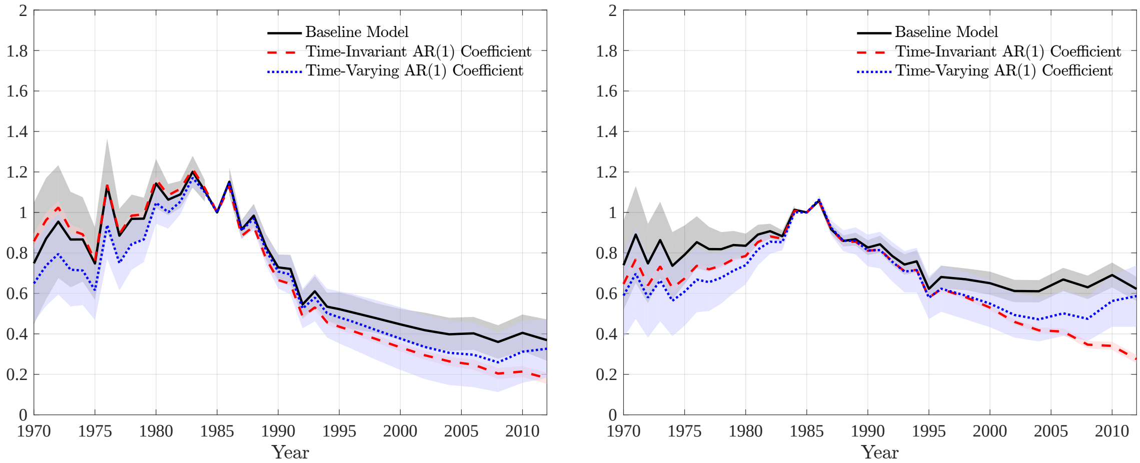
Due to the correlation between skill growth and past skill levels, our IV estimator will generally produce biased estimates for growth rates in skill returns when $\rho_t \neq 1$.[26] However, we show in Appendix A.1 that if $\text{Var}(\psi) > 0$, the evolution of both ρ_t and μ_t over time can still be identified under Assumption 2 and other mild conditions.

Identification breaks down when $\text{Var}(\psi) = 0$. In this knife-edge 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

[26]The IV estimator converges to $\gamma_{t,t'}$ in (10), 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 3 and 4 do not support this dependence of $\gamma_{t,t'}$ on t'.

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 estimate the model with AR(1) skills as defined in equation (13) 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 $\mu_{1985} = 1$), $\mathrm{Var}(\nu_t|c)$, and $\mathrm{Var}(\psi|c)$ separately for non-college and college men. To improve precision and facilitate estimation, we assume that $\mathrm{Var}(\psi|c)$ is a cubic polynomial in entry cohort c and that $\mathrm{Var}(\nu_t|c)$ is a cubic time trend multiplied by a quadratic experience trend. Long autocovariances for workers with at least 21 years of experience are targeted.[27]



(a) Non-College (b) College

Figure 9: μ_t implied by MD estimates allowing for time-varying vs. time-invariant AR(1) skill shocks, 21–40 years of experience

Importantly, estimates for $\mathrm{Var}(\psi|c)$ are always positive, with estimates significantly greater than zero for cohorts at the heart of our sample (see Appendix Figure C-5). This means that it is possible to separately identify the process for skill dynamics from the returns to skill over time. We find 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 1.07 for non-college men and 1.06 for college men. Based on the more general time-varying ρ_t case, Figure C-6 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 9 shows that the estimated μ_t series (for both fixed and time-varying ρ_t) are very similar to our baseline case

[27]Specifically, we target $\widehat{\mathrm{Cov}}(w_t, w_{t'}|E_j)$ 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 B for additional details.

assuming $\rho = 1$ and estimates reported in Figure 7.

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 strengthen conditions (i)–(iii) of Assumption 1 but maintain the limited persistence of non-skill shocks (Assumption 1(iv)).

Assumption 3. *(i)* $\mathrm{Cov}(\Delta\theta_t, \theta_{t'}) = 0$ *for all* $t - t' \geq 1$*; (ii)* $\mathrm{Cov}(\theta_t, \varepsilon_{t'}) = 0$ *for all* t, t'*; and (iii) for known* $k \geq 1$*,* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$ *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 $\mathrm{Cov}(w_t, w_{t'}) = \mu_t \mu_{t'} \mathrm{Var}(\theta_{t'})$ for $t - t' \geq k$. As in Proposition 1, the IV estimator of equation (4) can be used to identify $\mu_{\underline{t}+k}, \ldots, \mu_{\bar{t}}$ (with one normalization). Given skill returns and sufficient panel length, the variance of unobserved skills, $\mathrm{Var}(\theta_{t'}) = \mathrm{Cov}(w_t, w_{t'})/\mu_t \mu_{t'}$, can be identified for all but the first and last k periods. This variance is generally not identified for earlier periods, because it cannot be separated from (unidentified) early skill returns – only $\mu_t \mathrm{Var}(\theta_t)$ 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.

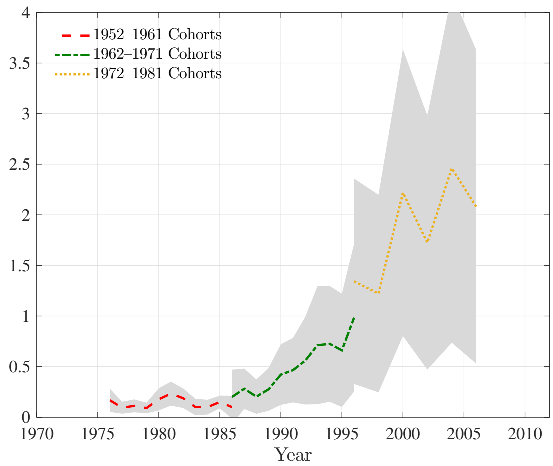
Having identified the variance of unobserved skills over time, it is straightforward to then identify variation in skill growth, $\mathrm{Var}(\Delta\theta_t) = \mathrm{Var}(\theta_t) - \mathrm{Var}(\theta_{t-1})$, and the variance of non-skill shocks, $\mathrm{Var}(\varepsilon_t) = \mathrm{Var}(w_t) - \mu_t^2 \mathrm{Var}(\theta_t)$, for all but the first and last k periods.[28]

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 3, it is straightforward to show that IV regression using future wage residuals as instruments identifies:

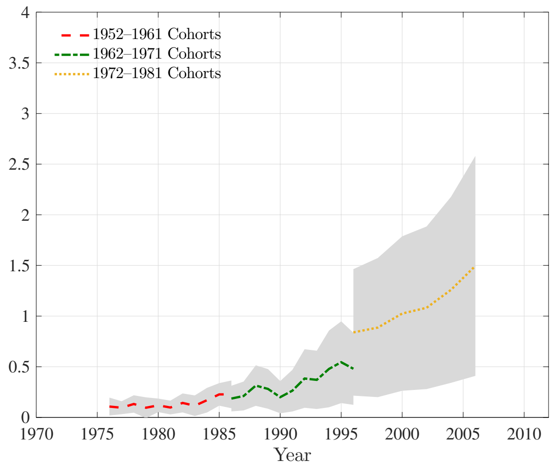
$$\frac{\mathrm{Cov}(\Delta w_t, w_{t''})}{\mathrm{Cov}(w_{t-1}, w_{t''})} = \frac{\Delta\mu_t}{\mu_{t-1}} + \frac{\mu_t}{\mu_{t-1}} \frac{\mathrm{Var}(\Delta\theta_t)}{\mathrm{Var}(\theta_{t-1})} \quad \text{for } t'' \geq t + k. \tag{14}$$

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).

[28] Lochner, Park, and Shin (2025) extend these identification results to the variances of unobserved skills, skill growth innovations, and non-skill transitory shocks when these all vary by cohort.



(a) Non-College



(b) College

Figure 10: $\text{Var}(\theta_{t'}|C) = \Omega_{C,t'}/\mu_{t'}$ implied by MD estimates using long autocovariances, 21–40 years of experience

IV estimates presented in Tables 3 and 4, as well as Appendix C.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 C.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 10 reports $\text{Var}(\theta_{t'}|C) = \Omega_{C,t'}/\mu_{t'}$ (for older men) obtained from estimates of μ_t and $\Omega_{C,t'}$ reported in Figures 7 and C-1, respectively. This figure shows that unobserved skill dispersion for the 1952–1961 birth cohorts was largely stable over the late-1970s and early-1980s. However, beginning in the early-1990s, unobserved skill dispersion grew sharply for both non-college and college men from more recent 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 3, estimated separately by education (for all ages) using MD estimation. We have already estimated this model in Section 3.5.2, 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 solid black lines in Figure 9. Figure 11 decomposes $\text{Var}(w_t)$ 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.

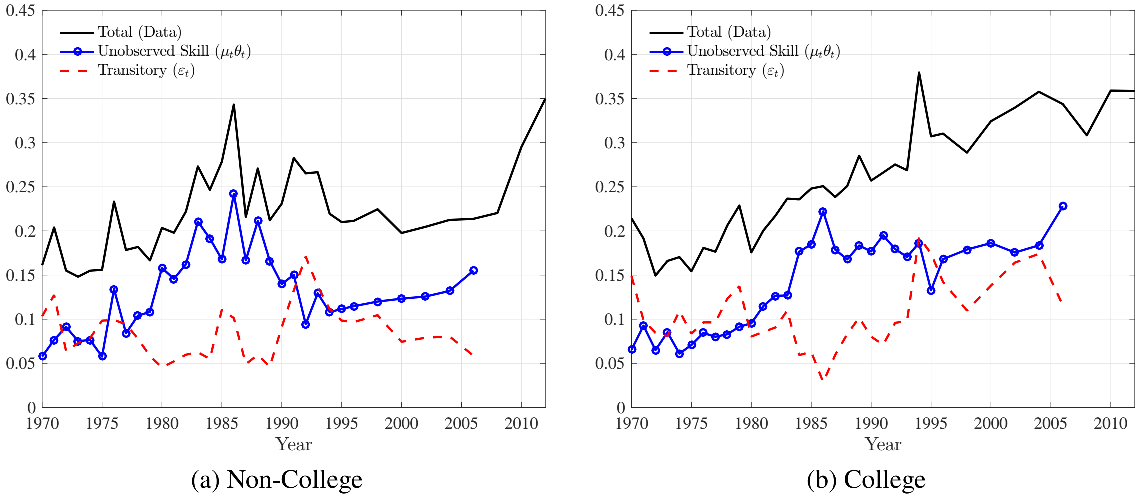


Figure 11: Log wage residual variance decomposition

Decomposing skill variation. Our estimates suggest that from 1970 to 2005 the variance of skills, Var(θ_t), rose by roughly 0.8 among non-college men and 0.4 among college men. Over that period, the variance of skills at labor market entry rose by only 0.15 among non-college men, while it declined by roughly 0.07 among college men. Thus, changes in initial skill variation contributed little to long-term growth in overall skill dispersion. Instead, growing skill inequality has been driven almost entirely by sharp increases in the variance of skill growth innovations, most of which occurred in the second half of our sample period.

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 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 by increased volatility in skills rather than growing dispersion in skill levels at labor market entry.

5 Firms, Occupations, and Multiple Unobserved Skills

Thus far, we have focused on a specification for log wage residuals that is broadly consistent with traditional assignment models of the labor market, in which each worker is assigned to a different job based on a single-dimensional ranking of worker productivity (Tinbergen, 1956;

Sattinger, 1993; Costinot and Vogel, 2010).[29] This section begins by incorporating firm-level differences in pay, estimating the evolution of skill returns using workers who remain on the same job over time. We then explore differences in the returns to skill across broad occupation classes and consider the interpretation of our IV estimator when there are multiple skills earning different returns in the market.[30] Throughout this analysis, we continue to account for the fact that skills vary over time for individual workers.

5.1 Firm-Specific Effects

Following Abowd, Kramarz, and Margolis (1999), a growing literature studies the importance of both worker- and firm-specific wage components. Song et al. (2018) show that, in the U.S., most of the variation in log earnings residuals is explained by heterogeneity across workers, which also explains most of the rise in residual dispersion. Both Song et al. (2018) and Bonhomme et al. (2023) further show that the sorting of workers across firms has become more important over time; however, this sorting still explains a relatively small share of residual dispersion.

To explore the implications of heterogeneity in pay across firms for our estimated returns to skill, we incorporate firm fixed effects into our baseline log wage residual specification (assuming a single skill):

$$w_{i,t} = \kappa_{j_{i,t}} + \mu_t \theta_{i,t} + \varepsilon_{i,t}, \tag{15}$$

where $j_{i,t}$ denotes the firm individual i works for in year t and κ_j represent firm fixed effects.[31] We invoke the standard "exogenous job mobility" assumption of the literature, assuming that the non-skill component $\varepsilon_{i,t}$ satisfies $\mathrm{E}[\varepsilon_t | j_{\underline{t}}, \ldots, j_{\bar{t}}] = 0$ for all t.

A version of equation (15), abstracting from time-variation in worker skills and their returns (i.e., $\theta_{i,t} = \theta_i$ and $\mu_t = \mu$), is typically estimated using administrative employer-employee matched data, where employer identities are directly observed and a large number of workers transition between firms.[32] Since the PSID do not contain firm identifiers, we investigate the implications of *unobserved* firm-specific heterogeneity by applying our IV estimator to job stayers, for whom firm fixed effects do not change.

Figure 12 reports the path for skill returns implied by our 2SLS estimator for job stayers

[29] 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.

[30] See Sanders and Taber (2012) for a survey of the literature on lifecycle wage dynamics in models with multiple skills and occupations.

[31] A few recent studies estimate time-varying firm-specific effects (i.e., $\kappa_{j,t}$). 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. Song et al. (2018) show that the dispersion of estimated firm effects did not change substantially over time in the U.S., providing some support for the common firm fixed effects specification.

[32] Bonhomme, Lamadon, and Manresa (2019) include interactions between firm and worker effects in their log wage specification, effectively allowing skill returns to differ across firm types. Their results suggest that these worker–firm interactions explain a small share of overall earnings inequality (in Sweden).

from $t-2$ to t (roughly two-thirds of our sample) beginning in 1985.[33] Notably, the evolution of estimated skill returns are quite similar to their counterparts for all workers as reported in Figure 6. Appendix C.7 shows that IV estimates of skill return growth for stayers are likely to be biased towards zero due to unobserved variation in κ_j. Using estimated variation in worker- and firm-fixed effects from Song et al. (2018), the appendix argues that this bias is negligible for college men but that the returns to skill for non-college men fell by as much as 50% more compared to those reported in Figure 12.

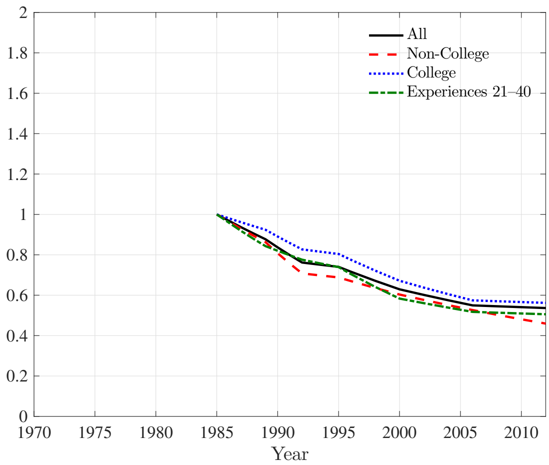


Figure 12: μ_t Implied by 2SLS for Job Stayers from $t-2$ to t

5.2 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 next modify our baseline log wage specification to allow for occupation-specific wage functions with residuals given by

$$w_{i,t} = \gamma_t^{o_{i,t}} + \mu_t^{o_{i,t}} \theta_{i,t} + \varepsilon_{i,t}, \tag{16}$$

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

[33] It is difficult to identify employer transitions prior to 1985 (Brown and Light, 1992). As discussed in Appendix C.7, we identify job stayers based on the most recent start date of the current main job.

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 (16), 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 (16) generalizes related specifications used in studies of sectoral or firm differences in pay, which typically assume fixed skills over time (e.g., Gola, 2021; Bonhomme, Lamadon, and Manresa, 2019).

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

Assumption 4. *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 4 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 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).[34]

As shown in Appendix A.2, μ_t^o can be identified for all occupations (in all but the first k years) under Assumption 4 using the following IV estimator (for $t - t' \geq k + 1$):

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'}|o_t, o_{t-1}, o_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'}|o_t, o_{t-1}, o_{t'})} = \frac{\mu_t^{o_t} - \mu_{t-1}^{o_{t-1}}}{\mu_{t-1}^{o_{t-1}}}. \tag{17}$$

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. Assumption 4 implies that estimated return growth for occupation stayers from $t-1$ to t should not depend on earlier occupation ($o_{t'}$). Results reported in Appendix C.8 confirm this prediction.

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 4, identification of γ_t^o requires $\mathrm{E}[\Delta\theta_t|o_t, o_{t-1}] = 0$ and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}] =$

[34] 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) document declining occupational mobility over the lifecycle.

$\mathrm{E}[\varepsilon_{t-1}|o_t, o_{t-1}] = 0$, which imply[35]

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

5.2.1 2SLS Estimation of Skill Return Growth for Occupation-Stayers

Lochner, Park, and Shin (2025) show that IV estimated skill returns based on the sample of all 3-digit occupation stayers, regardless of occupation, are very similar to our baseline estimated return series (for all men) reported in Figure 6. Here, 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).[36] We also estimate 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).[37] 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 (9) earlier, we use 2SLS (with lagged residuals as instruments) to estimate two-year growth rates in occupation-specific skill returns based on

$$\Delta_2 w_{i,t} = \left(\gamma_t^{o_{i,t}} - \left[\frac{\mu_t^{o_{i,t}}}{\mu_{t-1}^{o_{i,t-2}}}\right]\gamma_{t-2}^{o_{i,t-2}}\right) + \left[\frac{\mu_t^{o_{i,t}} - \mu_{t-2}^{o_{i,t-2}}}{\mu_{t-1}^{o_{i,t-2}}}\right] w_{i,t-2} + \underbrace{\left(\varepsilon_{i,t} - \left[\frac{\mu_t^{o_{i,t}}}{\mu_{t-2}^{o_{i,t-2}}}\right]\varepsilon_{i,t-2} + \mu_t^{o_{i,t}}\Delta_2\theta_{i,t}\right)}_{\equiv \xi_{i,t}}, \tag{18}$$

estimated separately for stayers with $o_{i,t} = o_{i,t-2} = o$ for cognitive, routine, or social occupation groups.[38]

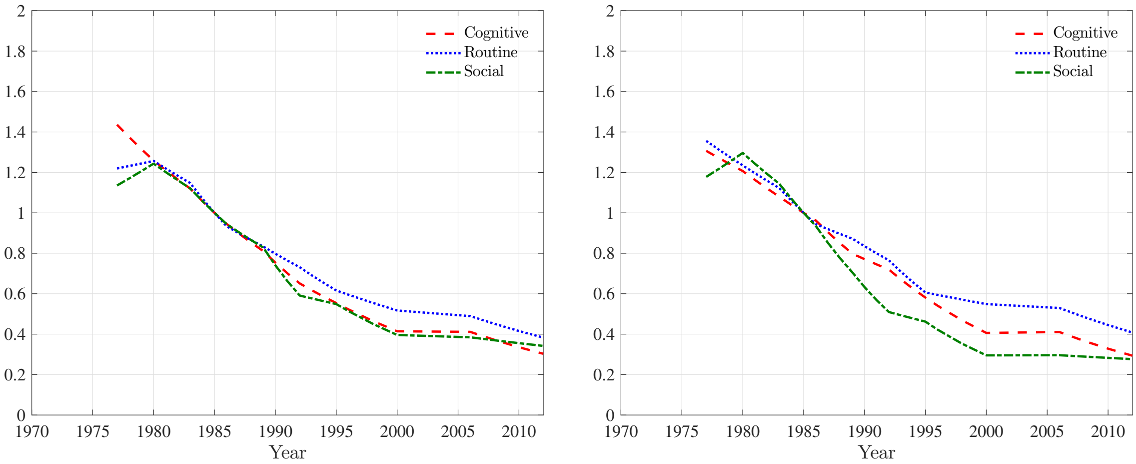
Figure 13 reports implied skill returns (relative to μ_{1985}^o) for all men who remain in cognitive, routine, or social occupations. Panel (a) reports estimates based on workers of all experience levels, while panel (b) reports estimates based on those with 21–40 years of experience. In both panels, 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

[35] 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 A.2.

[36] Cortes (2016) also considers manual occupations, but our sample contains too few observations to obtain precise results for its associated wage parameters. Appendix C.8 provides details on occupation classifications in the PSID.

[37] 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 C.8.

[38] 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 13: μ_t^o/μ_{1985}^o implied by 2SLS estimates for cognitive, routine, and social occupation-stayers between $t-2$ and t

for the full sample.

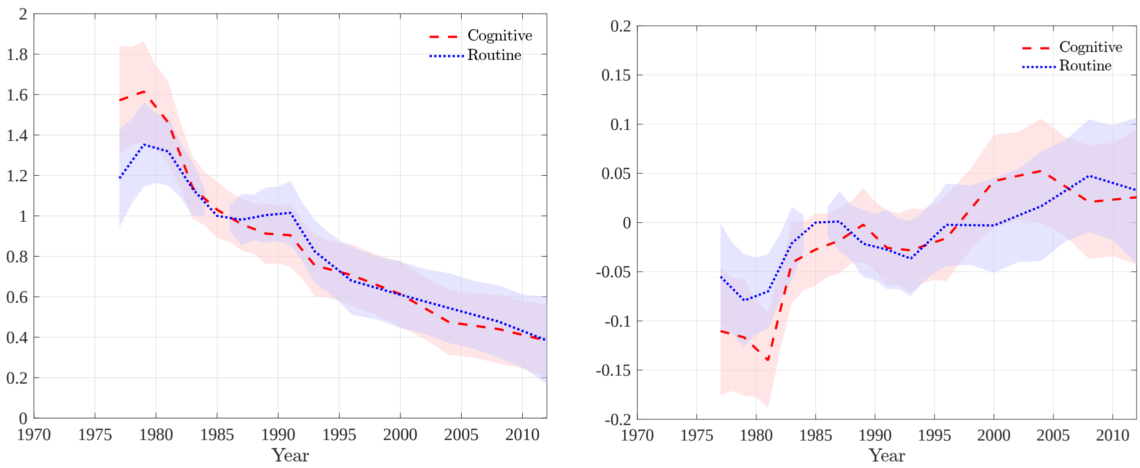
5.2.2 GMM Estimation of γ_t^o and μ_t^o for Cognitive and Routine Occupations

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 (18), we exploit the following moments in the PSID: $\mathrm{E}[\xi_t|o_t, o_{t-2}] = 0$ and $\mathrm{E}[w_{t'}\,\xi_t|o_t, o_{t-2}] = 0$, where we use lagged residuals $w_{t'}$ from periods $t-8$ and $t-9$ ($t-10$ in later years) as instruments. See Appendix C.8 for details.

Given the substantial overlap between cognitive and social occupations (and similar skill return profiles in Figure 13), 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 14(a) shows that estimated μ_t^o series both exhibit substantial declines over time, similar to the 2SLS estimates in Figure 13 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 14(b) shows sizeable increases in γ_t^o – nearly 0.20 in cognitive occupations and about 0.12 in routine occupations.[39]

Based on the estimates reported in Figure 14 and average log wage residuals by occupation, we 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 15 shows the evolution of average log wage residuals and average skills for cognitive and routine workers. During our sample period, average log wage residuals

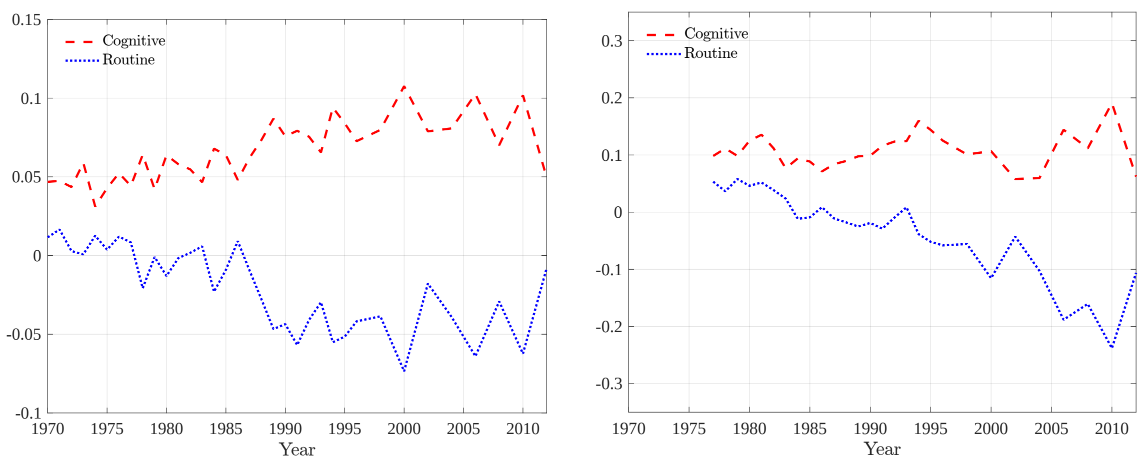
[39] Estimated time patterns for μ_t^o and γ_t^o are very similar when limiting the sample to workers with 21–40 years of experience in year t (see Lochner, Park, and Shin, 2025).



(a) μ_t^o (b) γ_t^o

Figure 14: GMM estimates of μ_t^o and γ_t^o for cognitive and routine occupations

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.



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

Figure 15: Average log wage residual and skill for cognitive and routine occupations

Appendix Figure C-8 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 estimated occupation-specific skill levels.

5.3 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 $j = 1, ..., J$:

$$w_{i,t} = \sum_{j=1}^{J} \mu_{j,t}\theta_{i,j,t} + \varepsilon_{i,t}, \tag{19}$$

where $\mu_{j,t}$ reflects the market-level value of skill j in year t.[40]

We now make use of a multi-dimensional version of Assumption 3 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.[41]

Assumption 5. *(i)* $\text{Cov}(\Delta\theta_{j,t}, \theta_{j',t'}) = 0$ *for all* j, j'*, and* $t - t' \geq 1$*; (ii)* $\text{Cov}(\theta_{j,t}, \varepsilon_{t'}) = 0$ *for all* j, t, t'*; and (iii)* $\text{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$ *for all* $t - t' \geq k$*.*

This assumption implies:

$$\text{Cov}(\Delta w_t, w_{t'}) = \sum_{j=1}^{J} \sum_{j'=1}^{J} \Delta\mu_{j,t}\, \mu_{j',t'}\, \text{Cov}(\theta_{j,t'}, \theta_{j',t'}), \quad \text{for } t - t' \geq k + 1, \tag{20}$$

where we highlight that the $\text{Cov}(\theta_{j,t'}, \theta_{j',t'})$ are within-period covariances across skills. Equation (20) shows that when all (within-period) correlations between skills are non-negative, a positive (negative) $\text{Cov}(\Delta w_t, w_{t'})$ 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 $\overline{\theta}_{i,t} \equiv \sum_{j=1}^{J} \mu_{j,t}\theta_{i,j,t}$, the total value of a worker's skill vector in period t.

[40] Multi-dimensional assignment and search/matching models of the labor market can give rise to equilibrium log wage functions of the form in equation (19) (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 (19) can be thought of as a linear approximation. Equation (19) is reminiscent of wage (rather than log wage) functions in Heckman and Scheinkman (1987) when worker characteristics can be "unbundled".

[41] 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. The stronger conditions based on Assumption 3 facilitate interpretation of the weights.

Proposition 2. *If Assumption 5 holds, then for all $t - t' \geq k + 1$ the IV estimator identifies a weighted-average growth rate across all skill returns:*

$$\frac{\text{Cov}(\Delta w_t, w_{t'})}{\text{Cov}(w_{t-1}, w_{t'})} = \sum_{j=1}^{J} \omega_{j,t',t-1} \left(\frac{\Delta \mu_{j,t}}{\mu_{j,t-1}} \right), \tag{21}$$

with weights for $j = 1, ..., J$ given by $\omega_{j,t',t-1} = \text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})\mu_{j,t-1} / \sum_{j'=1}^{J} \text{Cov}(\theta_{j',t'}, \overline{\theta}_{t'})\mu_{j',t-1}$. If $\text{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$, 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 $t - 1$).[42]

The multi-skill problem 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:

$$w_{i.t} = \sum_{j=1}^{J} \mu_{j,t} \alpha^{o_{i,t}}_{j,t} \theta_{i,j,t} + \varepsilon_{i,t}. \tag{22}$$

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, $\alpha^o_{j,t}$.[43]

Conditioning all covariances in Assumption 5 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 in occupation o, $\Delta\tilde{\mu}^o_{j,t}/\tilde{\mu}^o_{j,t-1} = (\Delta\mu_{j,t}\alpha_{j,t} + \mu_{j,t-1}\Delta\alpha^o_{j,t})/\tilde{\mu}^o_{j,t-1}$, where the returns to skills that are more important for wages in occupation o receive more weight. (See Appendix A.3 for details.) 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; 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

[42] Appendix A.3 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 3 in Appendix A.3 shows that the IV estimator also reflects growth in a weighted-average measure of skill returns.

[43] 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) consistent with the following: $w_{i,t} = \gamma^{o_{i,t}}_t + \sum_j \mu^{o_{i,t}}_t \alpha^{o_{i,t}}_{j,t} \theta_{i,j,t} + \varepsilon_{i,t}$. Lochner, Park, and Shin (2025) show that if skill intensities do not vary over time within occupations (i.e., $\alpha^o_{j,t} = \alpha^o_j$), then our IV estimator for stayers identifies occupation-specific skill return growth, $\Delta\mu^o_t/\mu^o_{t-1}$, as in Section 5.2. More generally, the IV estimator for stayers in occupation o also reflects any changes in skill intensities within that occupation. Figure 13 is 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.

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 13 are, therefore, consistent with relatively stable occupation skill intensities and similar declines in the returns to a broad range of skills.[44]

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).[45] 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 E for a detailed discussion of these data and our sample. We highlight Appendix Figures E-2 and E-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.

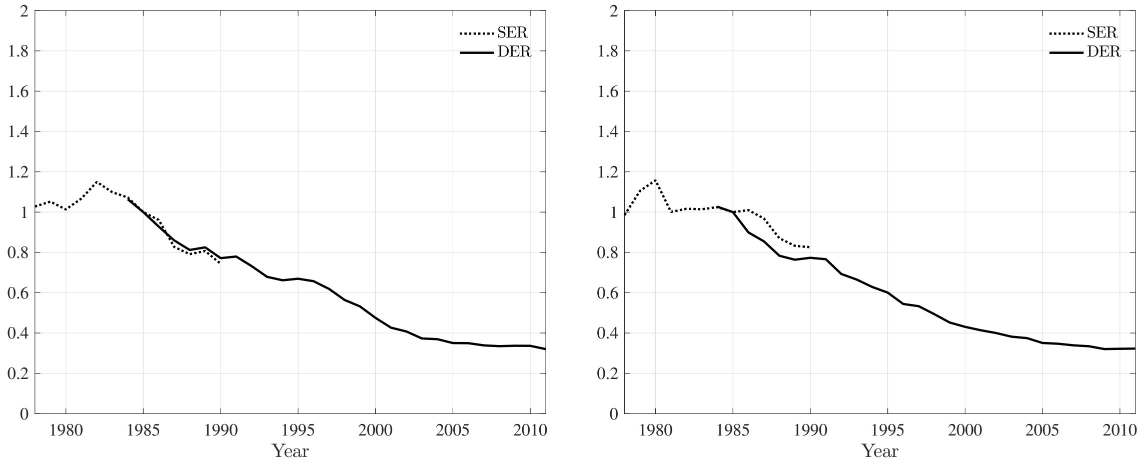
We use our IV estimator in equation (4) to estimate growth rates for skill returns using w_{t-7} as an instrument (consistent with $k = 6$). Since sample sizes are much larger than in the PSID, we limit our sample to men with 32–40 years of experience to focus on years when wage growth is generally negligible, yet before most men retire.[46] Figure 16 reports the implied 1984–2011

[44] 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 (e.g., see Figure 15) 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.

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

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

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.[47]



(a) Non-College

(b) College

Figure 16: μ_t implied by IV estimates (instrument: w_{t-7}), experience 32–40 in t (SIPP/W-2)

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 contain 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 in a routine occupation during years $t-1$ and t.[48] 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 5 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, stronger declines in routine occupations could be driven by routine-biased technical change (Autor and Dorn, 2013).

[47] See Appendix Table E-2 for the estimates, standard errors, and sample sizes when using the SER and DER earnings residuals as instruments.

[48] Occupations are based on 24 categories created by the Census Bureau. See Appendix E for details.

Table 5: 2SLS estimates of $\Delta\mu_t/\mu_{t-1}$ for occupational stayers, 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.

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 unobserved skills or early-career skill growth have remained stable across cohorts born decades 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 data on wages) 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 W-2 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 generally attribute rising residual inequality 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.

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 declines among workers in routine relative to cognitive jobs. Our estimates based on administrative W-2 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).

All of our conclusions derive directly from the time patterns for long-autocovariances for log wage residuals, notably the sharp declines over the late-1980s and 1990s (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 firm- and occupation-stayers, suggesting that they are not simply explained by shifts in firm/occupation structure or by changes in firm/occupational switching (see, e.g., Kambourov and Manovskii, 2008). Viewed through the lens of the canonical wage function for unobserved skills introduced by Juhn, Murphy, and Pierce (1993), it is difficult to reconcile these robust trends with rising returns to unobserved skill, as is so often assumed. While we do not attempt to explain why unobserved skill returns fell over a period when returns to education rose,[49] 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 greater variation in lifecycle wage growth during periods of rapid

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

innovation.[50] 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).

References

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.

[50] 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.

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.

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, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2023. "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics* 41 (2):291–322.

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

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.

Brown, James N. and Audrey Light. 1992. "Interpreting Panel Data on Job Tenure." *Journal of Labor Economics* 10 (3):219–257.

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.

Carneiro, Pedro, Karsten T Hansen, and James J Heckman. 2003. "2001 Lawrence R. Klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice." *International Economic Review* 44 (2):361–422.

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.

Cunha, Flavio, James Heckman, and Salvador Navarro. 2005. "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic papers* 57 (2):191–261.

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, Paweł. 2021. "Supply and demand in a two-sector matching model." *Journal of Political Economy* 129 (3):940–978.

Gottschalk, Peter and Robert Moffitt. 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.

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, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics* 24 (3):411–482.

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.

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.

Kline, Patrick. 2024. "Firm Wage Effects." In *Handbook of Labor Economics*, vol. 5, edited by Christian Dustmann and Thomas Lemieux, chap. 2. Elsevier, 115–181.

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.

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.

———. 2025. "The Evolution of Unobserved Skill Returns in the U.S.: A New Approach Using Panel Data." Working Paper 33407, National Bureau of Economic Research.

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.

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. 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.

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.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2018. "Firming Up Inequality." *The Quarterly Journal of Economics* 134 (1):1–50.

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.

Woessmann, Ludger. 2024. "Skills and Earnings: A Multidimensional Perspective on Human Capital." Working Paper.

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.

Online Appendix

A Identification Results

A.1 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:

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

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:

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

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 $\text{Var}(\psi) = 0$, 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 $\text{Var}(\psi) > 0$.

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

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

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 (13) imply the following orthogonality conditions in terms of the transformed variables:

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

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

$$\text{Cov}(w_{t'}, w_t | c) - \tilde{\rho}_t \text{Cov}(w_{t'}, w_{t-1} | c) = \mu_{t'} (\mu_t - \tilde{\rho}_t \mu_{t-1}) \text{Var}(\psi | c). \quad (23)$$

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 (23) 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 $t'' \le t-k-1$,

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

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

$$A\tilde{\rho}_t^2 + B\tilde{\rho}_t + C = 0, \tag{24}$$

where

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

We can investigate some cases that equation (24) has a unique solution. First, if $A = 0$ and $B \neq 0$, then the unique solution is $\tilde{\rho}_t = -C/B$. Second, if $A \neq 0$ and $B^2 - 4AC = 0$, then the unique solutions becomes $\tilde{\rho}_t = -B/2A$. Notice that there exist other sets of sufficient conditions, especially by constructing additional moment conditions using different cohorts or applying instruments from different time periods.

Identification of μ_t. From Equation (23), we have, for $t \le t' - k - 1$,

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

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

Equation (23) also implies that, for $t' \le t - k - 2$,

$$\frac{\mathrm{Cov}(w_{t'}, w_t|c) - \tilde{\rho}_t \mathrm{Cov}(w_{t'}, w_{t-1}|c)}{\mathrm{Cov}(w_{t'}, w_{t-1}|c) - \tilde{\rho}_{t-1} \mathrm{Cov}(w_{t'}, w_{t-2}|c)} = \frac{\mu_t - \tilde{\rho}_t \mu_{t-1}}{\mu_{t-1} - \tilde{\rho}_{t-1}\mu_{t-2}} = \frac{\mu_{t-1}}{\mu_{t-2}} \frac{\left(\frac{\mu_t}{\mu_{t-1}} - \tilde{\rho}_t\right)}{\left(\frac{\mu_{t-1}}{\mu_{t-2}} - \tilde{\rho}_{t-1}\right)}. \tag{26}$$

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 (25), μ_t/μ_{t-1} for $t \geq \bar{t} - k$ is also identified from Equation (26) 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 (23) 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.

A.2 Identification with Occupations

Identification of μ_t^o. With Assumption 4(iii)–(iv), the long autocovariance for log wage residuals for $t - t' \geq k + 1$ can be written as follows:

$$\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}, o_{t'}) = \mu_t^{o_t} \left[\mu_{t'}^{o_{t'}} \mathrm{Cov}(\theta_t, \theta_{t'}|o_t, o_{t-1}, o_{t'}) + \mathrm{Cov}(\theta_t, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'}) \right], \tag{27}$$

$$\mathrm{Cov}(w_{t-1}, w_{t'}|o_t, o_{t-1}, o_{t'}) = \mu_{t-1}^{o_{t-1}} \left[\mu_{t'}^{o_{t'}} \mathrm{Cov}(\theta_{t-1}, \theta_{t'}|o_t, o_{t-1}, o_{t'}) + \mathrm{Cov}(\theta_{t-1}, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'}) \right]. \tag{28}$$

Moreover, Assumption 4(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 (27) and (28) imply equation (17).

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 4. Consider the long autocovariance (27) that is not conditioned on $o_{t'}$:

$$\begin{aligned}\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}) = & \mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}, o_{t'})|o_t, o_{t-1} \right] \\ & + \mathrm{Cov}\left(\mathrm{E}[w_t|o_t, o_{t-1}, o_{t'}], \mathrm{E}[w_{t'}|o_t, o_{t-1}, o_{t'}]|o_t, o_{t-1} \right).\end{aligned} \tag{29}$$

The second term in equation (29) is

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

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

Thus, the long autocovariances (27) and (28) that are not conditioned on $o_{t'}$ are given by

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

where

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

Assumption 4(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

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

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.

A.3 Identification with Multiple Skills

Recall that we define $\overline{\theta}_{i,t} \equiv \sum_j \mu_{j,t}\theta_{i,j,t}$. Given Assumption 5, 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 5 implies the following:

$$\frac{\mathrm{Cov}(w_t - w_{t_0}, w_{t'})}{\mathrm{Cov}(w_{t_0}, w_{t'})} = \frac{\sum_{j=1}^{J} (\mu_{j,t} - \mu_{j,t_0})\,\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j=1}^{J} \mu_{j,t_0}\,\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})} = \sum_{j=1}^{J} \omega_{j,t',t_0}\left(\frac{\mu_{j,t} - \mu_{j,t_0}}{\tilde{\mu}_{j,t_0}}\right) \tag{30}$$

where the weights,

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

Notice that $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$, implies that $\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}) \geq 0, \forall j$. Since $\mu_{j,t} \geq 0, \forall j, t$, the weights $\omega_{j,t',t_0} \geq 0$ whenever $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$. 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 $L(a|b) = \mathrm{Cov}(a,b)b/\mathrm{Var}(b)$ is the linear projection of a onto b. The second equality follows from condition (i) of Assumption 5, 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 3. *If Assumption 5 holds, then for all $t - t' \geq k + 1$, the IV estimator identifies growth in the weighted-average return to skills, $m_{t,t'} = \sum_{j=1}^{J} \varphi_{j,t'} \mu_{j,t}$:*

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

with weights

$$\varphi_{j,t'} \equiv \frac{\text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j'=1}^{J} \text{Cov}(\theta_{j',t'}, \overline{\theta}_{t'})}, \quad \textit{for } j = 1, ..., J. \tag{31}$$

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

Proof of Proposition 3. Using the definitions of $m_{t,t'}$ and $\varphi_{j,t'}$, growth in the weighted-average return to skills can be written as

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

where the last equality reflects the IV estimator using $w_{t'}$ as an instrument.

Notice that $\text{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$, implies that $\text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}) \geq 0, \forall j$. 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 5 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', $\overline{\theta}_{t'}$.

A.3.1 Occupations as Bundles of Skills

We now consider the case in which log wage residuals are given by equation (22), 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 6. *(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 6, our IV estimator conditional on $o_t = o_{t-1} = o$ and $o_{t'} = o'$ identifies the following for $t - t' \geq k + 1$:[51]

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

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: $\Delta\tilde{\mu}^o_{j,t}/\tilde{\mu}^o_{j,t-1} = (\Delta\mu_{j,t}\alpha_{j,t} + \mu_{j,t-1}\Delta\alpha^o_{j,t})/\tilde{\mu}^o_{j,t-1}$. Notice that $\mathrm{Cov}(\theta_{j,t'}, \theta_{j',t'}|o, o') \geq 0, \forall j, j'$, implies that $\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o') \geq 0, \forall j$.

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 $o_{t'}$:

$$\mathrm{Cov}(w_t, w_{t'}|o) = \mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o, o_{t'})|o\right] + \mathrm{Cov}\left(\mathrm{E}[w_t|o, o_{t'}], \mathrm{E}[w_{t'}|o, o_{t'}]|o\right). \tag{32}$$

With Assumption 6, the first term in equation (32) is

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

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 (32) is

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

[51] 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 $\mathrm{E}[x|o_t = o_{t-1} = o]$.

Therefore,

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

Altogether, Assumption 6, $\mathrm{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\theta_{j,t-1} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = 0$ 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

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

where

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

If $\mathrm{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}^{o_{t'}} |o) \geq 0$ 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.

B MD Estimation and Standard Errors

B.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 $\widehat{\text{Cov}}(w_t, w_{t'}|s, C_j)$ is the sample covariance for residuals in years t and t' conditional on education group s and cohort group C_j and $\text{Cov}(w_t, w_{t'}|s, C_j, \mathbf{\Lambda})$ is the corresponding theoretical covariance given parameter vector $\mathbf{\Lambda}$. For college and non-college men (estimated separately), Table B-1 describes the cohort groups, parameters estimated, and autocovariances used in estimation of μ_t and $\Omega_{C,t'}$.

Table B-1: Cohort grouping

Cohort Group C	Range						Number		
	Cohort c		Year t'		Year t		$\Omega_{C,t'}$	μ_t	$\text{Cov}(w_t, w_{t'}\|C)$
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.

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

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

where $\Gamma = \{s, j, t, t'|1970 \le t' \le t \le 2012, t - t' \ge 6, s \in \{\text{Non-college, College}\}, j \in \{3, 4\}\}$; $\widehat{\text{Cov}}(w_t, w_{t'}|s, E_j)$ is the sample covariance for residuals in years t and t' conditional on education group s and experience group E_j; and $\text{Cov}(w_t, w_{t'}|s, E_j, \mathbf{\Lambda})$ is the corresponding theoretical covariance given parameter vector $\mathbf{\Lambda}$.

B.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 $\text{Cov}(w_t, w_{t'}|s, E_j)$. That is, both $w_{i,t}$ and $w_{i,t'}$ are non-missing and $s_{i,t} = s_{i,t'} = s$ and $e_{i,t} \in E_j$. Also let $p_m(\mathbf{\Lambda}) = \text{Cov}(w_t, w_{t'}|s, E_j, \mathbf{\Lambda})$. 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 $\mathbf{\Lambda}_0$:

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

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:

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

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

C PSID Data Details and Additional Results

C.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.[52] 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.

[52] 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).

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.

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

Table C-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 C-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.

C.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 (9) as follows:

$$\Delta_2 w_{i,t} = \left(\frac{\Delta_2 \mu_t}{\mu_{t-2}}\right) w_{i,t-2} + u_{i,t}, \tag{33}$$

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:

$$\mathrm{E}[w_{t'} u_t] = 0, \qquad \text{for } t' \leq t - 2 - k. \tag{34}$$

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

$$\mathrm{E}[w_{t''} u_t] = 0, \qquad \text{for } t'' \geq t + k. \tag{35}$$

Equation (35) 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 $t'' \geq t + k$,

$$\frac{\mathrm{Var}(\Delta_2\theta_t)}{\mathrm{Var}(\theta_{t-2})} = \left[\frac{\mathrm{Cov}(\Delta_2 w_t, w_{t''})}{\mathrm{Cov}(w_{t-2}, w_{t''})} - \frac{\mathrm{Cov}(\Delta_2 w_t, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right]\left[1 + \frac{\mathrm{Cov}(\Delta_2 w_t, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right]^{-1}. \tag{36}$$

C.3.1 Overidentification Tests

We begin by testing the moments in equation (35) 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 (34) and the first two available out of $w_{i,t+6}, w_{i,t+7}, w_{i,t+8}, w_{i,t+9}$ for (35).

Table C-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 (33). Panel A reports estimates when moments from both equations (34) and (35) are used (i.e., lags and leads), while Panel B reports estimates when only the moment condition from equation (34) is used (i.e., lags only). The sample is restricted to be the same in both panels.[53]

Comparing the J-statistics in Panels A and B in Table C-3, we can test the validity of using leads as instruments (i.e. moments in equation (35)). 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 (35) 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 (34)) 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.

Table C-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.

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

C.3.2 Inferring Relative Magnitude of Skill Shocks

Table C-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 (34) are used (i.e., 2 nearest valid lags). These results are analogous to the 2SLS estimates in Tables 1 and 2, using the same samples. Comparing estimates across the tables, we see that they are quite similar. Panel B reports GMM estimates when only the moments in equation (35) 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 (36) 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 C-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 $(w_{t+6}, \ldots, w_{t+9})$. Panel C reports estimates of skill growth shock variance relative to skill variance based on equation (36).

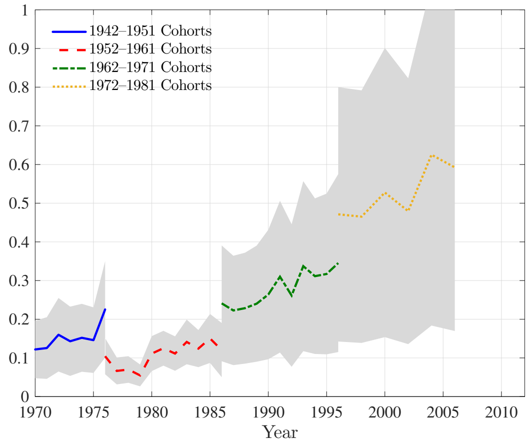
* denotes significance at 0.05 level.

C.4 MD Estimates based on Cohort Groups

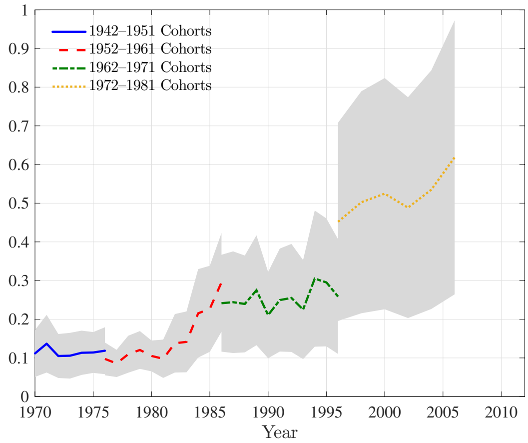
Figure C-1 reports estimated $\Omega_{C,t'}$ for non-college and college men (shaded areas reflect 95% confidence intervals) as estimated with μ_t (see Figure 7) in Section 3.4.

C.5 Testing HIP based on growth in log wage residuals

Figures C-2 to C-4 report results for $\mathrm{Cov}(\Delta(w_t/\mu_t), \Delta(w_{t'}/\mu_{t'}))$ and $\mathrm{Cov}(\Delta(w_t/\mu_t), w_{t'})$ in the PSID.

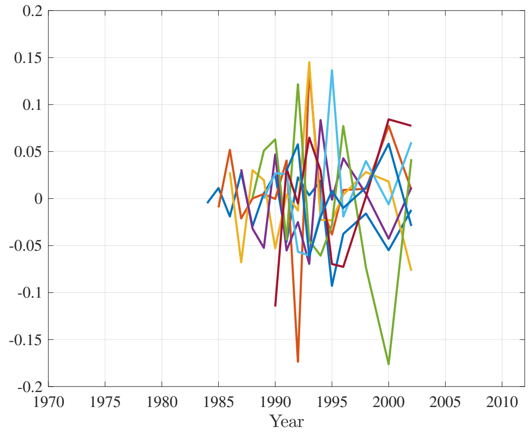


(a) Non-College

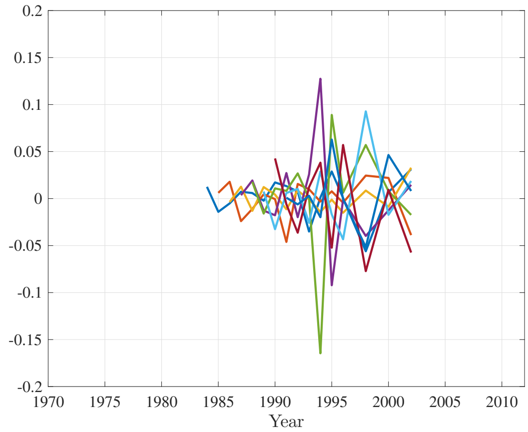


(b) College

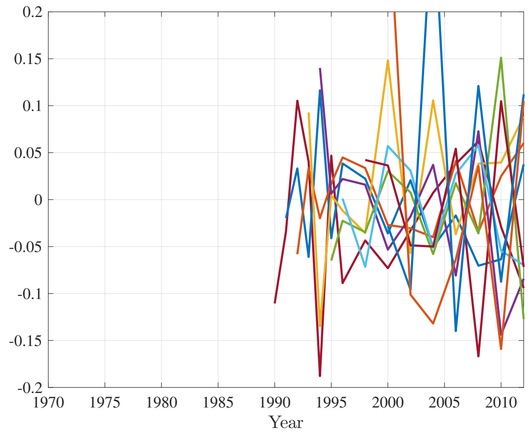
Figure C-1: $\Omega_{C,t'}$ implied by MD estimates using long autocovariances, 21–40 years of experience



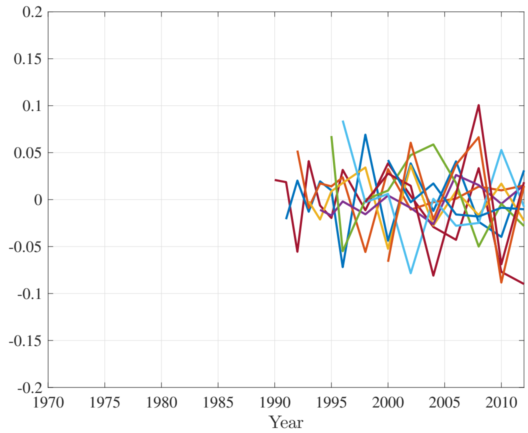
(a) 1962–1971 Cohort Group, Non-College



(b) 1962–1971 Cohort Group, College



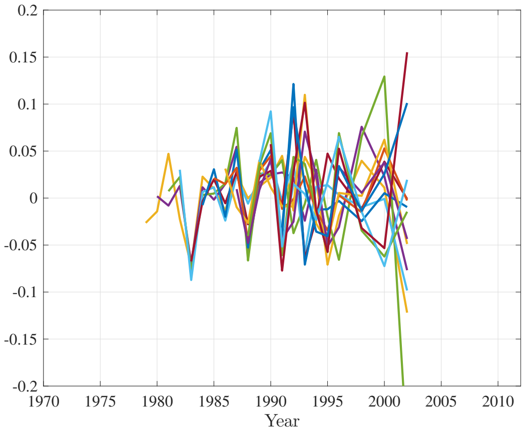
(c) 1972–1981 Cohort Group, Non-College



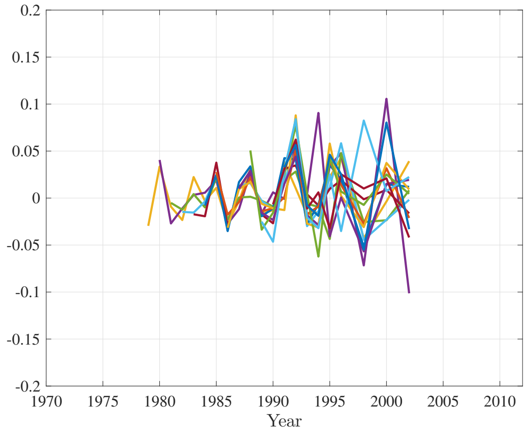
(d) 1972–1981 Cohort Group, College

Figure C-2: $\text{Cov}(\Delta(w_t/\mu_t), \Delta(w_{t'}/\mu_{t'}))$ for Men by Cohort Group

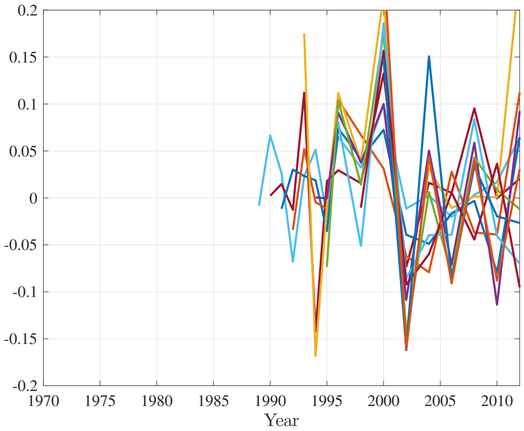
Notes: Figure reports covariances for cohort groups $C \in \{1962-1971, 1972-1981\}$ *where each line holds* t' *fixed and varies* $t \geq t' + 7$.



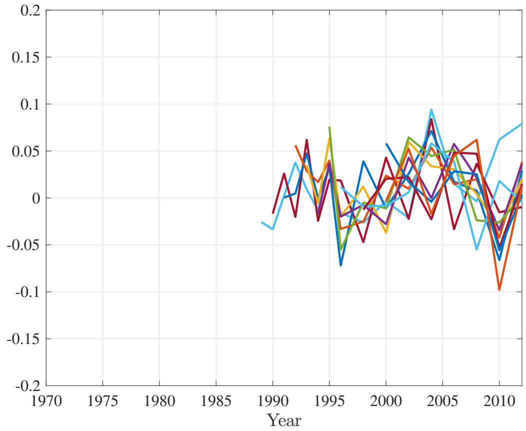
(a) 1962–1971 Cohort Group, Non-College



(b) 1962–1971 Cohort Group, College



(c) 1972–1981 Cohort Group, Non-College



(d) 1972–1981 Cohort Group, College

Figure C-3: $\mathrm{Cov}(\Delta(w_t/\mu_t), w_{t'})$ for each t, t' by Cohort Group

Notes: Figure reports covariances for cohort groups $C \in \{1962-1971, 1972-1981\}$ *where each line holds* t' *fixed and varies* $t \geq t' + 7$.

C.6 Additional estimates for model with AR(1) skill dynamics

Figures C-5 and C-6 report estimated $\mathrm{Var}(\psi|c)$ and ρ_t 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.

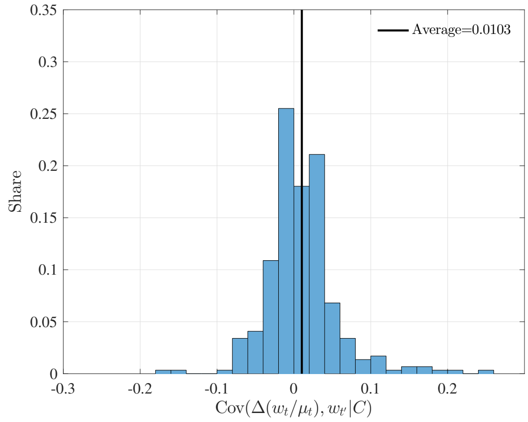
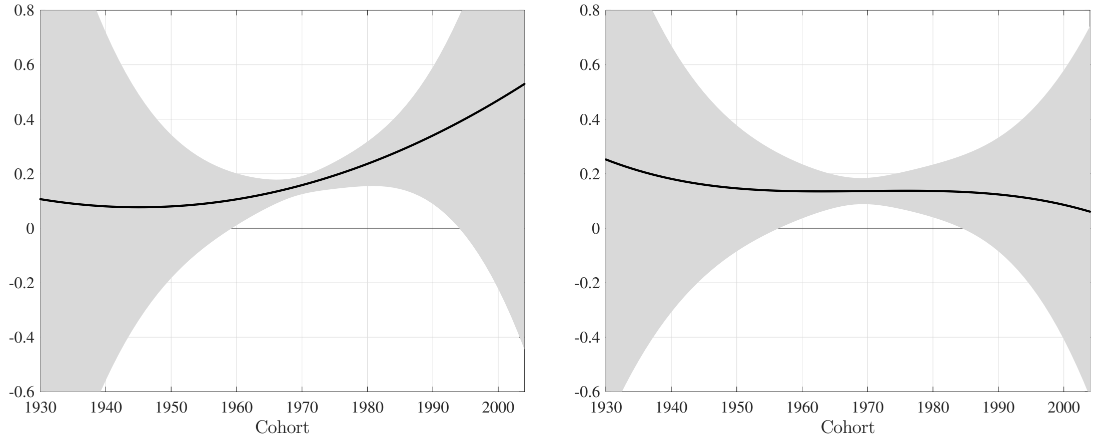


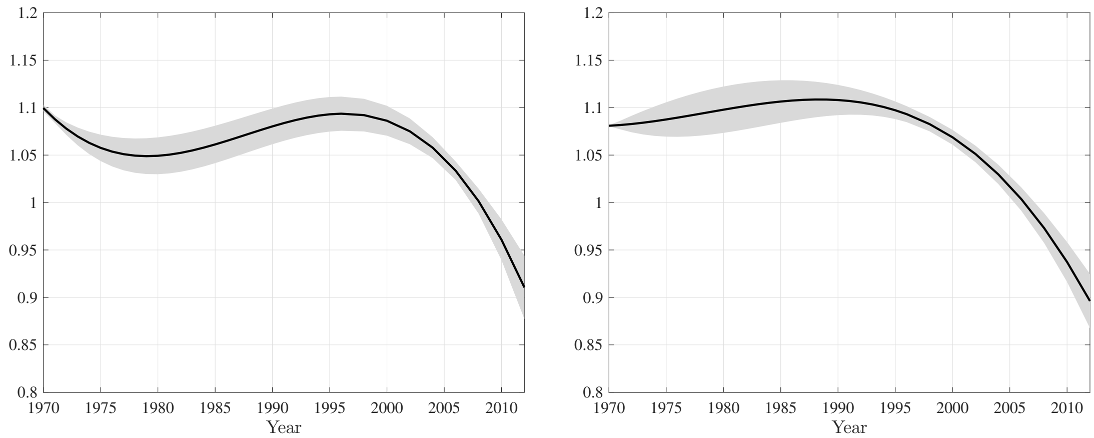
Figure C-4: Distribution of $\mathrm{Cov}(\Delta(w_t/\mu_t), w_{t'})$ for all (t, t', C) for for Low-Experience Men



(a) Non-College Men

(b) College Men

Figure C-5: $\mathrm{Var}(\psi|c)$ implied by MD estimates allowing for time-varying AR(1) skill shocks, 21–40 years of experience



(a) Non-College Men

(b) College Men

Figure C-6: ρ_t implied by MD estimates allowing for time-varying AR(1) skill shocks, 21–40 years of experience

C.7 Estimation with Job Stayers in PSID

This appendix considers log wage residuals that include firm fixed effects as in equation (15). Since the PSID do not contain firm identifiers, we investigate the implications of *unobserved* firm-specific heterogeneity for our IV estimator, focusing on job stayers for whom firm fixed effects do not change. In addition to the standard "exogenous job mobility" assumption of the literature (i.e., $\mathrm{E}[\varepsilon_t | j_{\underline{t}}, \ldots, j_{\bar{t}}] = 0, \forall t$), we assume the following.

Assumption 7. *For known* $k \geq 1$ *and for all* $t - t' \geq k + 1$*: (i)* $\mathrm{Cov}(\Delta\theta_t, \theta_{t'} | j_t = j_{t-1}) = 0$*; (ii)* $\mathrm{Cov}(\Delta\theta_t, \varepsilon_{t'} | j_t = j_{t-1}) = 0$*; (iii)* $\mathrm{Cov}(\varepsilon_t, \theta_{t'} | j_t = j_{t-1}) = \mathrm{Cov}(\varepsilon_{t-1}, \theta_{t'} | j_t = j_{t-1}) = 0$*; (iv)* $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t'} | j_t = j_{t-1}) = \mathrm{Cov}(\varepsilon_{t-1}, \varepsilon_{t'} | j_t = j_{t-1}) = 0$*; and (v)* $\mathrm{Cov}(\Delta\theta_t, \kappa_{j_{t'}} | j_t = j_{t-1}) = 0$.

Conditions (i)–(iv) strengthen Assumption 1 to condition on individuals remaining in the same job between periods $t-1$ and t, while condition (v) further assumes that recent skill changes for job stayers are orthogonal to the identity of previous employers sufficiently long ago.

Under Assumption 7, the IV estimator for job stayers, for $t - t' \geq k + 1$, is given by:

$$\begin{aligned}
&\frac{\mathrm{Cov}(\Delta w_t, w_{t'} | j_t = j_{t-1})}{\mathrm{Cov}(w_{t-1}, w_{t'} | j_t = j_{t-1})} \\
&\quad = \frac{\Delta\mu_t \,\mathrm{Cov}(\theta_{t-1}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})}{\mathrm{Cov}(\kappa_{j_{t-1}}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1}) + \mathrm{Cov}(\mu_{t-1}\theta_{t-1}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})} \\
&\quad = \frac{\Delta\mu_t}{\mu_{t-1}} \times \left(1 + \frac{\mathrm{Cov}(\kappa_{j_{t-1}}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})}{\mathrm{Cov}(\mu_{t-1}\theta_{t-1}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})}\right)^{-1}. \qquad (37)
\end{aligned}$$

If $\mathrm{Cov}(\kappa_{j_{t-1}}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1}) > 0$ and $\mathrm{Cov}(\mu_{t-1}\theta_{t-1}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1}) > 0$, then the IV estimator is biased towards zero due to unobserved variation in κ_{j_t}. With exogenous job mobility, these conditions are equivalent to assuming that both $\kappa_{j_{t-1}}$ and θ_{t-1} are positively correlated with lagged residuals $w_{t'}$ for job stayers. Importantly, equation (37) suggests that our IV estimates (reported in Figure 12) likely under-estimate the actual decline in skill returns over time.

C.7.1 Measuring Job Transitions in the PSID

We measure two-year job changes in the PSID based on the most recent start year of the current main job, which has been available since the 1988 wave.[54] Specifically, we define job stayers between earnings years $t-2$ and t as those whose job in survey year $t+1$ (interview usually takes place between March and May) started before $t-2$. The share of job stayers is substantial (over 60%) and remains relatively stable over time, with a modest decline until 2000 followed by a rebound thereafter. See Lochner, Park, and Shin (2025) for additional details.

[54] Before 1988, respondents only report the total years of experience at the current main job, which may not be continuous.

C.7.2 Accounting for the Bias in Estimated Skill Returns

As noted above, our IV estimates likely under-estimate the decline in skill returns over time. We next show that it is possible to exploit estimates from studies using worker–firm matched data to correct for the bias associated with our IV estimator.

If $\mathrm{Cov}(\Delta\theta_{t''}, \theta_{t'} | j_t = j_{t-1}) = \mathrm{Cov}(\Delta\theta_{t''}, \kappa_{j_{t'}} | j_t = j_{t-1}) = 0$ and $\mathrm{Cov}(\kappa_{j_{t''}} - \kappa_{j_{t''-1}}, \theta_{t'} | j_t = j_{t-1}) = \mathrm{Cov}(\kappa_{j_{t''}} - \kappa_{j_{t''-1}}, \kappa_{j_{t'}} | j_t = j_{t-1}) = 0$ for $t'' \geq t' + 1$, then the intertemporal covariances for job stayers in the bias term in equation (37) can be written as within-period covariances, leading to

$$\frac{\mathrm{Cov}(\Delta w_t, w_{t'} | j_t = j_{t-1})}{\mathrm{Cov}(w_{t-1}, w_{t'} | j_t = j_{t-1})} = \frac{\Delta\mu_t}{\mu_{t-1}} \times \left(1 + \frac{\mu_{t'}}{\mu_{t-1}} \underbrace{\frac{\mathrm{Var}(\kappa_{j_{t'}} | j_t = j_{t-1}) + \mathrm{Cov}(\kappa_{j_{t'}}, \mu_{t'}\theta_{t'} | j_t = j_{t-1})}{\mathrm{Var}(\mu_{t'}\theta_{t'} | j_t = j_{t-1}) + \mathrm{Cov}(\kappa_{j_{t'}}, \mu_{t'}\theta_{t'} | j_t = j_{t-1})}}_{\equiv \Upsilon_{t'}}\right)^{-1}. \tag{38}$$

With our IV estimator for stayers and estimates of $\Upsilon_{t'}$, equation (38) can be used to identify μ_t over time. Unfortunately, the PSID do not allow us to identify the within-period covariances that make up $\Upsilon_{t'}$. We, therefore, use estimates of the variance and covariance terms that compose $\Upsilon_{t'}$ from Song et al. (2018), which are based on earnings records and employer identifiers from IRS W-2 forms. Since their estimates are based on the sample of all workers, not only job stayers, this approach further assumes that $\Upsilon_{t'}$ is the same for job stayers and switchers. The variance/covariance components from Song et al. (2018) are presented in Table C-5, which reports separate estimates over "rolling windows" of 7 years. When $\Upsilon_{t'} \neq 0$, $\mu_{t'}$ is needed for earlier years to identify μ_t for later years. Since $\mu_{t'}$ prior to 1985 cannot be identified from our IV estimates, we normalize $\mu_{t'} = 1$ for early years. This is generally consistent with weak time trends for estimated returns prior to 1985 when using the full sample.

Table C-5: Within-Period Covariances (from Table III of Song et al. (2018)) and Implied Υ_t

	1980–1986	1987–1993	1994–2000	2001–2007	2007–2013
$\mathrm{Var}(\kappa_{j_t})$	0.084	0.075	0.067	0.075	0.081
$\mathrm{Cov}(\kappa_{j_t}, \mu_t\theta_t)$	0.017	0.029	0.038	0.047	0.054
$\mathrm{Var}(\mu_t\theta_t)$	0.330	0.375	0.422	0.452	0.476
Implied Υ_t	0.290	0.257	0.228	0.244	0.255

Figure C-7 reports estimates for μ_t that correct for bias associated with firm fixed effects using the values of $\Upsilon_{t'}$ reported in Table C-5. These are based on (uncorrected) IV estimates for job stayers like those reported in Figure 12; however, these estimates only use a single lag, w_{t-8}, as an instrument (rather than two lags, w_{t-8} and w_{t-10}) to align with equation (38).[55] Compared to (uncorrected) estimates reported in Figure C-7, the (corrected) μ_t series displayed

[55] Note that estimates in Figure C-7 are obtained from two-year growth rates using data from every other year starting in 1986 (instead of all available years grouped in 2 or 3 years). We construct μ_t based on equation (38) modified for two-year growth rates, the IV estimates, and $\Upsilon_{t'}$ from Table C-5, normalizing $\mu_{t'} = 1$ for $t' \leq 1986$.

in Figure C-7 implies very little bias for college men but notably stronger bias for non-college men.

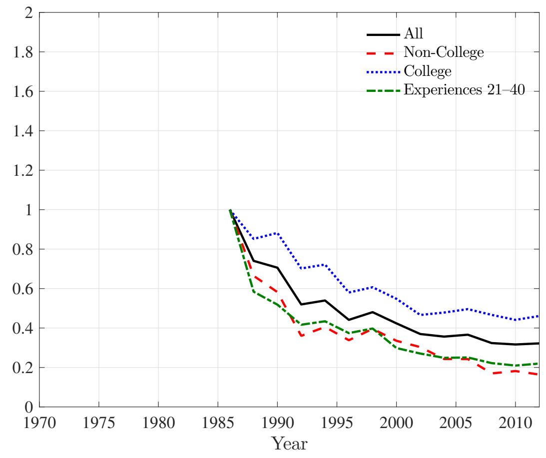


Figure C-7: μ_t Corrected for Bias due to Firm Fixed Effects

C.8 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.[56] 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.

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

[56] 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.

in social occupations. Conversely, around 76% of observations in social occupations are also in cognitive occupations.

C.8.1 GMM Estimation using Occupation Stayers and Switchers

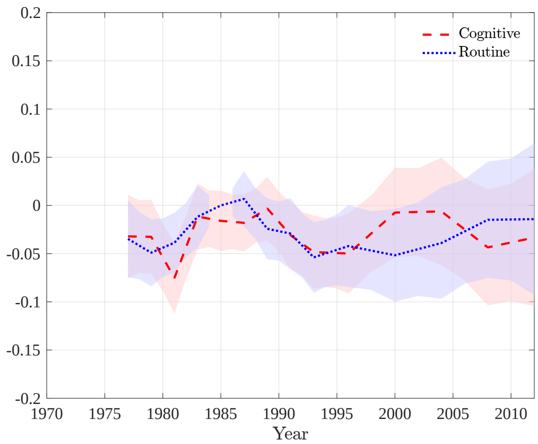
We estimate occupation-specific γ_t^o and μ_t^o for routine and cognitive occupations (normalizing $\gamma_{1985}^o = 0$ and $\mu_{1985}^o = 1$ for routine occupations) using optimal two-step GMM. Because we use the PSID data, we use the following moments based on equation (18):

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

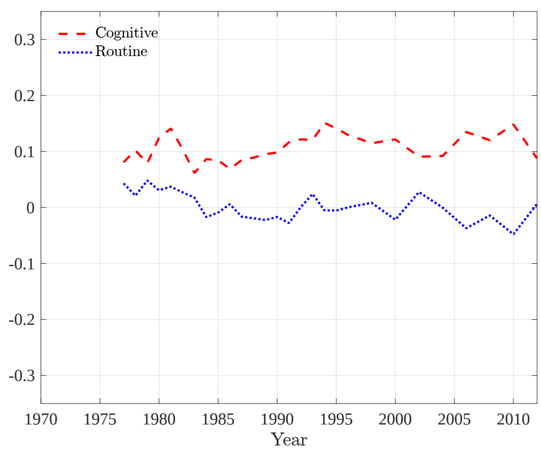
where $z_{i,t} = (1, w_{i,t-8}, w_{i,t-9})^\top$ (or $(1, w_{i,t-8}, w_{i,t-10})^\top$ in later sample years). We use linear splines for γ_t^o and μ_t^o, each with 14 knots in t, restricting to moment conditions with at least 20 observations (9 switcher moments are dropped). Altogether, there are 54 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 14, 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 C-8 reports estimates of γ_t^o and $\mathrm{E}[\theta_t|o_t]$ when imposing time-invariant $\mu_t^o = \mu^o$. Normalizing $\mu^{\text{routine}} = 1$, $\mu^{\text{cognitive}}$ is estimated to be 0.946 (SE=0.026). 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 14(b) and 15(b)).



(a) γ_t^o

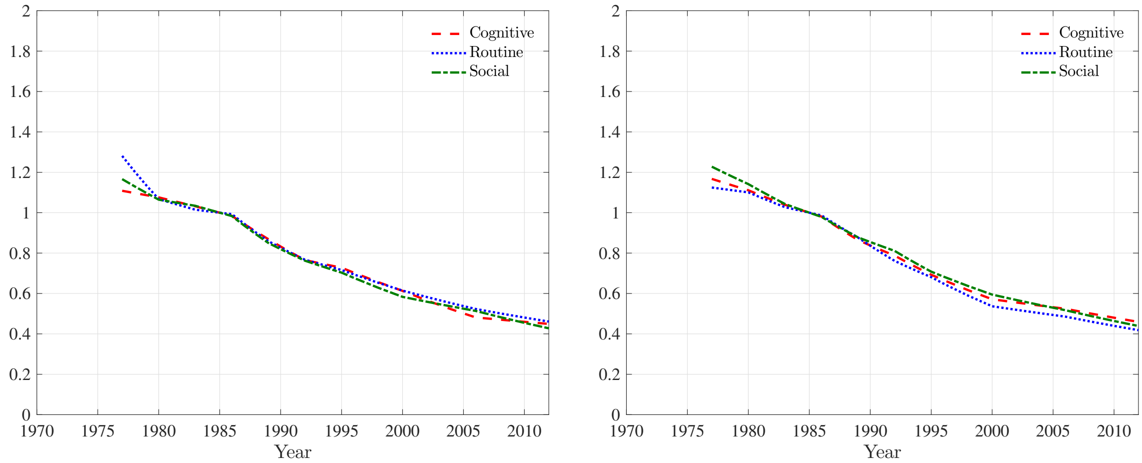


(b) $\mathrm{E}\left[\theta_t|o_t\right]$

Figure C-8: GMM estimates imposing time-invariant μ_t^o

C.8.2 2SLS Estimated Returns for Occupational Stayers

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 4. Estimates reported in Figure C-9 indicate very similar estimated skill return profiles for occupation stayers with $o_{t'} = o$ vs. $o_{t'} \neq o$.



(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 C-9: μ^o_t/μ^o_{1985} implied by 2SLS estimates for occupation-stayers: All experience levels

D HRS Data and Results

D.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. 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.) The survey has been fielded every two years since 1992, and it provides information about demographics, income, and cognition. Because the cognitive test we use changed in 1996, we use data collected from 1996 to 2018.

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.[57] The HRS also provides various cognitive functioning measures. We use word recall in our analysis. This 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 recalled in the two tasks to produce a score ranging from 0 to 20. Lochner, Park, and Shin (2025) show that this measure has correlations of about 0.3 with other cognitive measures (Serial 7s, Quantitative Reasoning) in the HRS and a correlation of 0.21 with log wages.

[57] See https://www.bls.gov/cpi/research-series/home.htm#CPI-U-RS2@Data.

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 50 years.[58] 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, and 11% Hispanic men 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.

D.2 Testing Conditions (i) and (ii) of Assumption 1

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

$$\mathrm{E}\left[(\Delta_2\tilde{T}_{t+2} - \varrho\tilde{T}_t)z_{t-\ell}\right] = 0, \quad \text{for } z_{t-\ell} \in \{\tilde{T}_{t-\ell}, w_{t-\ell}\} \text{ and } \ell \geq k, \tag{39}$$

where Δ_2 reflects the two-period time difference given our use of biennial data from the HRS.

We first test whether Assumption 1(i) holds for various k values. The first four columns of Table D-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 D-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. 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 D-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 $k = 6$.

Panel B of Table D-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 D-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.

[58] 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.

Table D-1: GMM estimates of ϱ in equation (39) 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.

E 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.[59] 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

[59]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.

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 E-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.

Table E-1: SIPP/W-2 Occupation Codes and 3-Category Grouping

Cognitive	Routine	Manual
Management (1) Business and financial operations (2) Computer and mathematical (3) Architecture and engineering (4) Life, physical, and social science (5) Community and social service (6) Legal (7) Education, training, and libraries (8) Arts, design, entertainment, sports, and media (9) Healthcare practitioner and technical (10)	Sales (16) Office and administrative support (17) Construction and extraction (19) Installation, maintenance, and repairs (20) Production (21) Transportation (22) Material moving (23)	Healthcare support (11) Protective service (12) Food prep and service (13) Building and grounds cleaning and maintenance (14) Personal care and service (15)

Notes: Farming, fishing, and forestry (18) and Military (24) are not classified.

Figure E-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 are higher than their counterparts for log wages in the PSID.

Figure E-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 E-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.

Table E-2 reports 2SLS estimates of skill return growth rates using SER- or DER-based lagged log earnings residuals (w_{t-7}) as instruments. (See Figure 16 in the paper.) Corresponding standard errors and sample sizes are also reported.

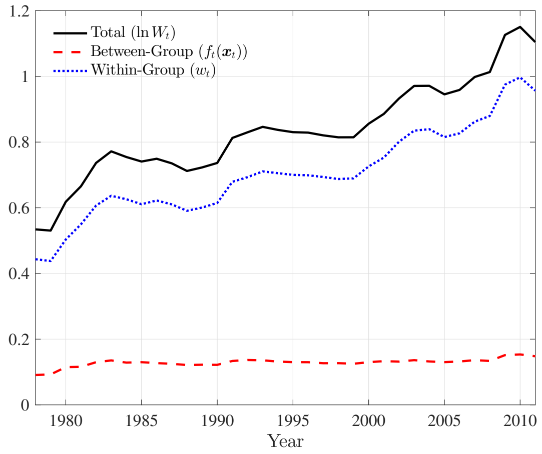


Figure E-1: Between- and within-group variances of log earnings, ages 16–64 with 5–40 years of experience (SIPP/W-2)

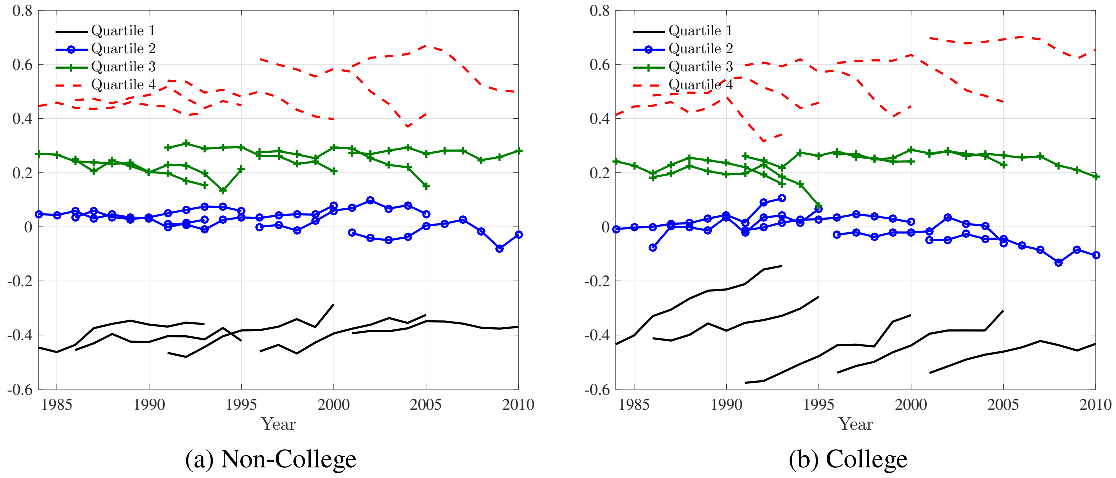
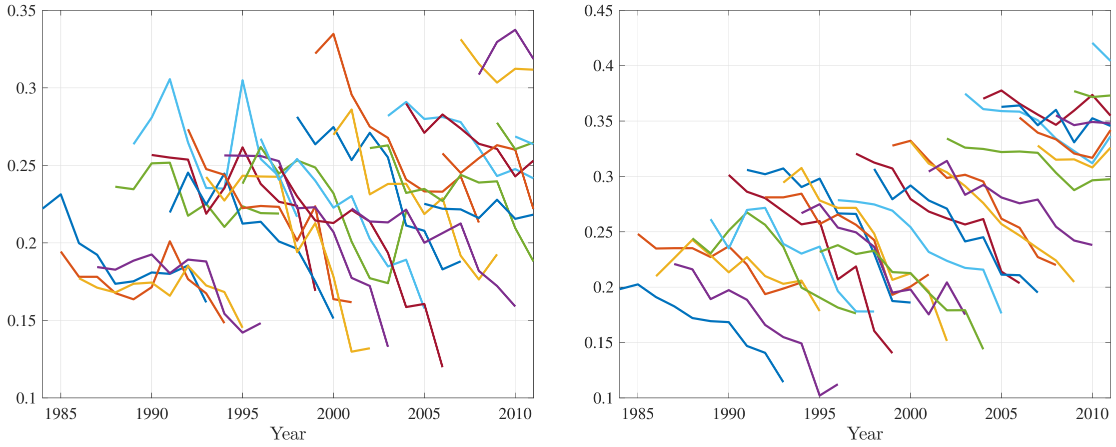


Figure E-2: Average predicted log earnings residuals by baseline residual quartile, 21–25 years of experience in base year (SIPP/W-2)



(a) Non-College

(b) College

Figure E-3: Autocovariances for log earnings residuals, 21–25 years of experience in base year (SIPP/W-2)

Table E-2: 2SLS estimates of $\Delta\mu_t/\mu_{t-1}$, 32–40 years of experience in year t (SIPP/W-2)

Year	Non-College			College		
	Estimate	Standard Error	Observations	Estimate	Standard Error	Observations
A. Using w_{t-7} from SER as instrument						
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
B. Using w_{t-7} from DER as instrument						
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.