# Online Appendix for "The Evolution of Unobserved Skill Returns in the U.S.: A New Approach Using Panel Data"

January 16, 2025

## A Prior Assumptions in the Literature

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

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

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

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

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

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

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

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

To simplify the comparison between cohort- and experienced-based growth in residual inequality, assume that shocks only depend on time and not cohort/experience: $\text{Var}(\varepsilon_t|c) = \text{Var}(\varepsilon_t)$ and $\text{Var}(\Delta\theta_t|c) =$

$\text{Var}(\Delta\theta_t)$ for all $c, t$. In this case, we have (for $c + \ell < t$):

$$\Delta_c(t,\ell) = \left(\mu_{t+\ell}^2 - \mu_t^2\right)\left[\text{Var}(\theta_c|c) + \sum_{\tau=c+1}^{t}\text{Var}(\Delta\theta_\tau)\right] + \mu_{t+\ell}^2\sum_{\tau=t+1}^{t+\ell}\text{Var}(\Delta\theta_\tau) + \left[\text{Var}(\varepsilon_{t+\ell}) - \text{Var}(\varepsilon_t)\right], \tag{27}$$

and

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

As discussed in Juhn, Murphy, and Pierce (1993), equation (27) shows that the change in variances over time for a given cohort incorporates both time effects and experience effects. The experience effects are reflected in the accumulation of permanent skill shocks from $t+1$ to $t+\ell$ (second term), while the time effects reflect changes in skill returns (first term) *and* in non-skill transitory shocks (third term). The evolution of variances over time for a given experience group includes the same three effects plus a fourth reflecting the difference the variance of skill levels between the cohorts as of the later time period. This is important, since it suggests that similar time patterns for residual variances obtained by following cohorts or experience groupings (i.e., $\Delta_e(t,\ell) \approx \Delta_c(t,\ell)$) implies that there is little variation across cohorts in early skill levels (i.e., $\text{Var}(\theta_{c+\ell}|c+\ell) \approx \text{Var}(\theta_{c+\ell}|c)$). This would be the case if the variance of initial skill levels were identical across cohorts ($\text{Var}(\theta_{c+\ell}|c+\ell) = \text{Var}(\theta_c|c)$) and if there were no early skill shocks over the first $\ell+1$ years of working careers. Alternatively, growth in the variance of initial skills across cohorts could offset growth in the variance of skills accumulated via labor market experience.

In the absence of initial cohort differences and early career skill shocks, changes in the variance of residuals should be the same over time whether we follow cohorts or experience groups: $\Delta_e(t,\ell) = \Delta_c(t,\ell)$. Put another way, we should observe similar growth over time in the variance when following cohorts or experience groups regardless of whether that growth is due to an increase in skill returns (i.e., the first term in equation (27)), the existence of skill growth shocks (i.e., the second term in equation (27)), or growth in transitory non-skill shocks (i.e., the third term in equation (27)). Thus, comparing growth in the variance of residuals for given cohorts or experience groups (as in Juhn, Murphy, and Pierce (1993)) is not directly informative about changes in the returns to skill unless there are no skill shocks and

the variance of non-skill shocks is time-invariant. Stated differently, $\Delta_e(t,\ell) = \Delta_c(t,\ell) > 0$ is consistent with growth in skill returns, permanent skill shocks, or growth in the variance of non-skill shocks.

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

## A.2 Lemieux (2006)

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

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

Table A-1: Wald test $p$-values

| Variables | All Men | Non-College | College |
|---|---|---|---|
| Year | 0.0000 | 0.0003 | 0.0000 |

## A.3 Castex and Dechter (2014)

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

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

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

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

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

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

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

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

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

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

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

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

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

which implies

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

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

# B Early Occupational Experiences in NLSY

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

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

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

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

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

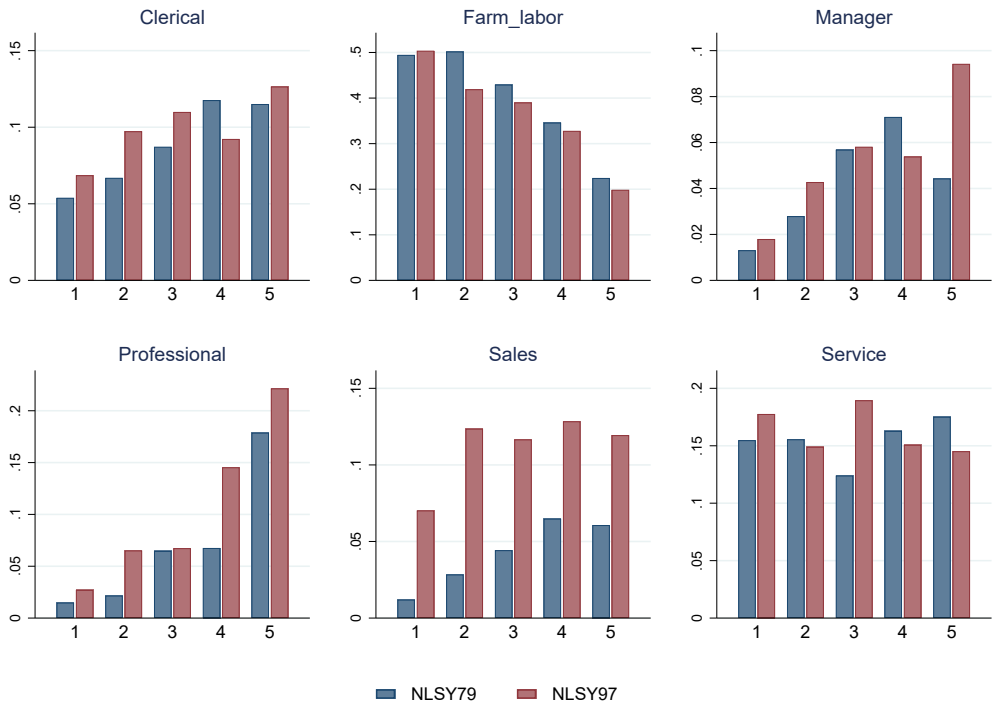



Figure B-1: Average fraction of years (over ages 17–26) spent working in occupations by AFQT quintile

# C Identification Results

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

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

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

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

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

$$\begin{aligned}
\text{Cov}(w_{t^*}, w_{t'}|c) &= \mu_{t^*}\mu_{t'}\,\text{Cov}(\theta_{t^*}, \theta_{t'}|c) && \text{[Assum 4′ (ii)–(iii)]} \\
&= \mu_{t^*}\mu_{t'}\,\text{Cov}(\theta_{t'} + \Delta\theta_{t'+1} + \cdots + \Delta\theta_{t^*}, \theta_{t'}|c) \\
&= \mu_{t^*}\mu_{t'}\,\text{Var}(\theta_{t'}|c) && \text{[Assum 4′ (i)].} \qquad (28)
\end{aligned}$$

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

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

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

**(ii) Identification of $\text{Var}(\theta_t|c)$ and $\text{Var}(\varepsilon_t|c)$.** For any $(t, t')$ such that $t' - t \geq k$, we can rearrange

equation (28) and get the following expression:

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

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

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

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

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

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

□

## C.2 Identifying early skill returns, $\mu_t$

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

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

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

$$\frac{\text{Cov}(w_t, w_{t'}|c) - \text{Cov}(w_t, w_{t'}|\tilde{c})}{\text{Cov}(w_{t-1}, w_{t'}|c) - \text{Cov}(w_{t-1}, w_{t'}|\tilde{c})} = \frac{\mu_t\mu_{t'}\left[\text{Var}(\theta_{t-1}|c) - \text{Var}(\theta_{t-1}|\tilde{c})\right]}{\mu_{t-1}\mu_{t'}\left[\text{Var}(\theta_{t-1}|c) - \text{Var}(\theta_{t-1}|\tilde{c})\right]} = \frac{\mu_t}{\mu_{t-1}}, \tag{29}$$

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

Condition (a) on cohorts is likely to hold quite generally. For example, differences in the variance of initial skill levels would contribute to different variances later in life. Even if initial skill distributions were identical across cohorts, the older cohort is likely to have accumulated more skill growth innovations over its longer career. Condition (b) holds when the skill growth variance depends only on time (and not experience) or when there is a non-monotonic experience trend in the variance of skill changes. For example, young workers may experience greater variation in skill growth than middle age workers due to differences in training or learning opportunities, while older workers may have a greater variance in skill changes due to differences in health shocks or skill obsolescence. Indeed, Baker and Solon (2003) and Blundell, Graber, and Mogstad (2015) estimate a U-shaped age profile for the variance of earnings shocks.

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

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

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

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

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

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

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:

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

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

Notice that Assumption 2 and the AR(1) process modeled in Equation (14) imply the following

orthogonality conditions in terms of the transformed variables:

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

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

$$\mathrm{Cov}(w_{t'}, w_t|c) - \tilde{\rho}_t \mathrm{Cov}(w_{t'}, w_{t-1}|c) = \mu_{t'}(\mu_t - \tilde{\rho}_t \mu_{t-1}) \mathrm{Var}(\psi|c). \tag{30}$$

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 (30) for cohort $c$ relative to $\tilde{c}$ yields

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

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

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

Combining these two equations yields

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

which becomes

$$A\tilde{\rho}_t^2 + B\tilde{\rho}_t + C = 0, \tag{31}$$

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 (31) has a unique solution. First, if $A = 0$ and $B \neq 0$, then the unique solution is

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

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

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

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

**Identification of $\mu_t$.** From Equation (30), we have, for $t \leq t' - k - 1$,

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

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

Equation (30) also implies that, for $t' \leq t - k - 2$,

$$\frac{\text{Cov}(w_{t'}, w_t|c) - \tilde{\rho}_t \text{Cov}(w_{t'}, w_{t-1}|c)}{\text{Cov}(w_{t'}, w_{t-1}|c) - \tilde{\rho}_{t-1} \text{Cov}(w_{t'}, w_{t-2}|c)} = \frac{\mu_t - \tilde{\rho}_t \mu_{t-1}}{\mu_{t-1} - \tilde{\rho}_{t-1}\mu_{t-2}} = \frac{\mu_{t-1}}{\mu_{t-2}} \frac{\left(\frac{\mu_t}{\mu_{t-1}} - \tilde{\rho}_t\right)}{\left(\frac{\mu_{t-1}}{\mu_{t-2}} - \tilde{\rho}_{t-1}\right)}. \tag{33}$$

Because $\tilde{\rho}_t$ is identified for all $t \geq \underline{t} + k + 1$ and $\mu_t/\mu_{t-1}$ is identified for all $t \leq \bar{t} - k - 1$ based on Equation (32), $\mu_t/\mu_{t-1}$ for $t \geq \bar{t} - k$ is also identified from Equation (33) as long as $\bar{t} - k - 1 \geq \underline{t} + k + 1$. Therefore, $\mu_t$ is identified for all $t$ (up to a normalization $\mu_{t^*} = 1$) if $\bar{t} - \underline{t} \geq 2(k + 1)$.

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

**Identification of $\text{Var}(\psi|c)$.** Equation (30) implies $\text{Var}(\psi|c)$ is identified for all cohorts observed in $t'$, $t - 1$, and $t$ such that $t' \leq t - k - 1$.

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

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

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

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

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

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

$$\varepsilon_{i,t} = \varphi_{i,t} + \tilde{\varepsilon}_{i,t} \tag{34}$$

$$\varphi_{i,t} = \rho_t \varphi_{i,t-1} + \sum_{j=0}^{\min\{q, t-c_i-1\}} \beta_j \nu_{i,t-j}, \tag{35}$$

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

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

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

and therefore

$$\text{Cov}(w_t, w_{t'}|c) - \rho_{t'} \text{Cov}(w_t, w_{t'-1}|c) = \mu_t (\mu_{t'} - \rho_{t'} \mu_{t'-1}) \text{Var}(\theta_t|c). \tag{36}$$

Taking the ratio of (36) for cohort $c$ relative to $\tilde{c}$ yields

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

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

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

Combining these two equations yields

$$\frac{\mathrm{Cov}(w_t, w_{t'}|c) - \rho_{t'}\,\mathrm{Cov}(w_t, w_{t'-1}|c)}{\mathrm{Cov}(w_t, w_{t'}|\tilde{c}) - \rho_{t'}\,\mathrm{Cov}(w_t, w_{t'-1}|\tilde{c})} = \frac{\mathrm{Cov}(w_t, w_{t''}|c) - \rho_{t''}\,\mathrm{Cov}(w_t, w_{t''-1}|c)}{\mathrm{Cov}(w_t, w_{t''}|\tilde{c}) - \rho_{t''}\,\mathrm{Cov}(w_t, w_{t''-1}|\tilde{c})}. \tag{37}$$

Equation (37) can be written as

$$A_1\rho_{t'}\rho_{t''} + B_1\rho_{t'} + C_1\rho_{t''} + D_1 = 0, \tag{38}$$

where

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

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

$$A_2\rho_{t'}\rho_{t''} + B_2\rho_{t'} + C_2\rho_{t''} + D_2 = 0, \tag{39}$$

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

We investigate some cases that this system of equations (38)–(39) has a unique solution. When $A_1 = A_2 = 0$, it becomes a system of linear equations. If $B_1C_2 - B_2C_1 \neq 0$, it has a unique solution

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

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

$$\begin{aligned}
\left(\rho_{t'} + \frac{C_1}{A_1}\right)\left(\rho_{t''} + \frac{B_1}{A_1}\right) &= \frac{B_1C_1 - A_1D_1}{A_1^2}\\
\left(\rho_{t'} + \frac{C_2}{A_2}\right)\left(\rho_{t''} + \frac{B_2}{A_2}\right) &= \frac{B_2C_2 - A_2D_2}{A_2^2}.
\end{aligned}$$

If the constants on the right hand side have different signs and the graphs share only one asymptote, it always has a unique solution. Therefore, we conclude that $(\rho_{t'}, \rho_{t''})$ are jointly identified when (i) $(B_1C_1 - A_1D_1)(B_2C_2 - A_2D_2) < 0$; and (ii-1) $C_1/A_1 = C_2/A_2$ and $(B_1/A_1) \neq (B_2/A_2)$ or (ii-2) $C_1/A_1 \neq C_2/A_2$ and $(B_1/A_1) = (B_2/A_2)$. Once we identify $\rho_{t'}$ for some $t'$, we can recover $\rho_{t^\dagger}$ recursively by constructing equation (38) with $t^\dagger$ instead of $t''$ or with different cohort if necessary. Then, we can identify $\rho_t$ for all $t \geq \underline{t} + k + 1$. Notice that these conditions are sufficient for identification. Additional equations generated by varying $t$, $c$, or $\tilde{c}$ can provide a different set of sufficient conditions for identification.

**Identification of $\mu_t$.** For $t' \geq t+k+1$, suppose that there exists $(c, \tilde{c})$ such that $\text{Var}(\theta_{t-1}|c) \neq \text{Var}(\theta_{t-1}|\tilde{c})$ and $\text{Var}(\Delta\theta_t|c) = \text{Var}(\Delta\theta_t|\tilde{c})$. Then, from Equation (36), we have

$$\begin{aligned}
&\frac{\big[\text{Cov}(w_t, w_{t'}|c) - \rho_{t'}\,\text{Cov}(w_t, w_{t'-1}|c)\big] - \big[\text{Cov}(w_t, w_{t'}|\tilde{c}) - \rho_{t'}\,\text{Cov}(w_t, w_{t'-1}|\tilde{c})\big]}{\big[\text{Cov}(w_{t-1}, w_{t'}|c) - \rho_{t'}\,\text{Cov}(w_{t-1}, w_{t'-1}|c)\big] - \big[\text{Cov}(w_{t-1}, w_{t'}|\tilde{c}) - \rho_{t'}\,\text{Cov}(w_{t-1}, w_{t'-1}|\tilde{c})\big]} \\
=&\frac{\mu_t(\mu_{t'} - \rho_{t'}\mu_{t'-1})\big[\text{Var}(\theta_{t-1}|c) - \text{Var}(\theta_{t-1}|\tilde{c})\big]}{\mu_{t-1}(\mu_{t'} - \rho_{t'}\mu_{t'-1})\big[\text{Var}(\theta_{t-1}|c) - \text{Var}(\theta_{t-1}|\tilde{c})\big]} \\
=&\frac{\mu_t}{\mu_{t-1}}.
\end{aligned} \tag{40}$$

Since $\rho_t$ is identified for all $t \geq \underline{t} + k + 1$, $\mu_t/\mu_{t-1}$ is identified for all $t \leq \bar{t} - k - 1$.

Equation (36) also implies that, for $t' \leq t - k - 2$,

$$\frac{\text{Cov}(w_{t'}, w_t|c) - \rho_t\,\text{Cov}(w_{t'}, w_{t-1}|c)}{\text{Cov}(w_{t'}, w_{t-1}|c) - \rho_{t-1}\,\text{Cov}(w_{t'}, w_{t-2}|c)} = \frac{\mu_t - \rho_t\mu_{t-1}}{\mu_{t-1} - \rho_{t-1}\mu_{t-2}} = \frac{\mu_{t-1}}{\mu_{t-2}}\frac{\left(\frac{\mu_t}{\mu_{t-1}} - \rho_t\right)}{\left(\frac{\mu_{t-1}}{\mu_{t-2}} - \rho_{t-1}\right)} \tag{41}$$

Because $\rho_t$ is identified for all $t \geq \underline{t} + k + 1$ and $\mu_t/\mu_{t-1}$ is identified for all $t \leq \bar{t} - k - 1$ based on Equation (40), $\mu_t/\mu_{t-1}$ for $t \geq \bar{t} - k$ is also identified from Equation (41) as long as $\bar{t} - k - 1 \geq \underline{t} + k + 1$. Therefore, $\mu_t$ is identified for all $t$ (up to a normalization $\mu_{t^*} = 1$) if $\bar{t} - \underline{t} \geq 2(k+1)$.

**Identification of $\text{Var}(\theta_t|c)$.** For $t' \geq t + k + 1$, Equation (36) implies

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

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

Finally, we note that it is straightforward to identify $\{\beta_j\}$ and $\text{Var}(\nu_t|c)$ from "close" autocovariances

given identification of everything else.

## C.5 Identification with Heterogeneous Skill Growth Rates

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

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

$$\theta_{i,t} = \psi_i + \Lambda_t(c_i)\delta_i + \sum_{j=0}^{t-c_i-1} \nu_{i,t-j}, \tag{42}$$

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

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

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

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

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

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

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

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

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

and its first difference is

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

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

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

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

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

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

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

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

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

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

where

$$\text{Cov}(\theta_t, \delta|c) = \text{Cov}(\psi, \delta|c) + \text{Var}(\delta|c)\sum_{j=0}^{t-c-1}\lambda_{t-j}(c). \tag{43}$$

Therefore,

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

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

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

Then,

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

Therefore,

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

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

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

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

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

## C.6 Identification with Occupations

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

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

$$\mathrm{Cov}(w_{t-1}, w_{t'}|o_t, o_{t-1}, o_{t'}) = \mu_{t-1}^{o_{t-1}}\left[\mu_{t'}^{o_{t'}}\,\mathrm{Cov}(\theta_{t-1}, \theta_{t'}|o_t, o_{t-1}, o_{t'}) + \mathrm{Cov}(\theta_{t-1}, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'})\right]. \tag{45}$$

Moreover, Assumption 5(i)–(ii) imply $\mathrm{Cov}(\theta_t, \theta_{t'}|o_t, o_{t-1}, o_{t'}) = \mathrm{Cov}(\theta_{t-1}, \theta_{t'}|o_t, o_{t-1}, o_{t'})$ and $\mathrm{Cov}(\theta_t, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'} = \mathrm{Cov}(\theta_{t-1}, \varepsilon_{t'}|o_t, o_{t-1}, o_{t'})$, so equations (44) and (45) imply equation (19).

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

0 in addition to Assumption 5. Consider the long autocovariance (44) that is not conditioned on $o_{t'}$:

$$\begin{aligned}\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}) =& \mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o_t, o_{t-1}, o_{t'})|o_t, o_{t-1}\right] \\ &+ \mathrm{Cov}\left(\mathrm{E}[w_t|o_t, o_{t-1}, o_{t'}], \mathrm{E}[w_{t'}|o_t, o_{t-1}, o_{t'}]|o_t, o_{t-1}\right). \end{aligned} \tag{46}$$

The second term in equation (46) is

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

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

Thus, the long autocovariances (44) and (45) that are not conditioned on $o_{t'}$ are given by

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

where

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

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

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

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

**Identification of $\gamma_t^o$.** Given $\mu_t^o$, we show that $\gamma_t^o$ 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 $\mu_t^o$ 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, $\gamma_t^o$ for $o \neq o^*$ is identified from occupation switchers.

## C.7 Identification with Multiple Skills

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

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

which implies Proposition 2.

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

Assumption 6 implies the following:

$$\begin{aligned}\frac{\text{Cov}(w_t - w_{t_0}, w_{t'})}{\text{Cov}(w_{t_0}, w_{t'})} &= \frac{\sum_{j=1}^{J} (\mu_{j,t} - \mu_{j,t_0}) \text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j=1}^{J} \mu_{j,t_0} \text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})} \\ &= \frac{\sum_{j=1}^{J} \left(\frac{\mu_{j,t} - \mu_{j,t_0}}{\mu_{j,t_0}}\right) \mu_{j,t_0} \text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j=1}^{J} \mu_{j,t_0} \text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})} \\ &= \sum_{j=1}^{J} \omega_{j,t',t_0} \left(\frac{\mu_{j,t} - \mu_{j,t_0}}{\tilde{\mu}_{j,t_0}}\right)\end{aligned} \tag{47}$$

where the weights,

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

Notice that $\text{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$, implies that $\text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'}) \geq 0, \forall j$. Since $\mu_{j,t} \geq 0, \forall j, t$, the weights $\omega_{j,t',t_0} \geq 0$ whenever $\text{Cov}(\theta_{j,t'}, \theta_{j',t'}) \geq 0, \forall j, j'$. 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 $L(a|b) = \text{Cov}(a, b)b/\text{Var}(b)$ is the linear projection of $a$ onto $b$. The second equality follows from condition (i) of Assumption 6, which implies that $L(\theta_{j,t}|\overline{\theta}_{t'}) = L(\theta_{j,t'}|\overline{\theta}_{t'})$ for all $t \geq t'$. Thus, the weight on growth in returns to skill $j$ depends on the (linearly) predicted rewards from skill $j$ in period $t_0$, $\mu_{j,t_0}\theta_{j,t_0}$, given total worker productivity in period $t'$, $\overline{\theta}_{t'}$.

**Proposition 4.** *If Assumption 6 holds, then for all $t - t' \geq k + 1$, the IV estimator identifies growth in the*

*weighted-average return to skills,* $m_{t,t'} = \sum_{j=1}^{J} \varphi_{j,t'} \mu_{j,t}$*:*

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

*with weights*

$$\varphi_{j,t'} \equiv \frac{\text{Cov}(\theta_{j,t'}, \overline{\theta}_{t'})}{\sum_{j'=1}^{J} \text{Cov}(\theta_{j',t'}, \overline{\theta}_{t'})}, \quad \textit{for } j = 1, ..., J. \tag{48}$$

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

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

where the last equality reflects equation (47).

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

### C.7.1 Occupations as Bundles of Skills

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

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

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

**IV Estimator Conditional on $o_t = o_{t-1} = o$ and $o_{t'} = o'$.** Given Assumption 8, our IV estimator conditional on $o_t = o_{t-1} = o$ and $o_{t'} = o'$ identifies the following for $t - t' \geq k + 1$:[68]

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

where

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

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

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

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

$$\mathrm{Cov}(w_t, w_{t'}|o) = \mathrm{E}\left[\mathrm{Cov}(w_t, w_{t'}|o, o_{t'})|o\right] + \mathrm{Cov}\left(\mathrm{E}[w_t|o, o_{t'}], \mathrm{E}[w_{t'}|o, o_{t'}]|o\right). \tag{49}$$

With Assumption 8, the first term in equation (49) is

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

With additional assumptions $\mathrm{E}[\theta_{j,t} - \theta_{j,t'}|o_t, o_{t-1}, o_{t'}] = 0$ for all $j$ and $\mathrm{E}[\varepsilon_t|o_t, o_{t-1}, o_{t'}] = \mathrm{E}[\varepsilon_{t'}|o_t, o_{t-1}, o_{t'}] = 0$, the second term in equation (49) is

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

Therefore,

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

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

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

where

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

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

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

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

The first term is

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

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

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

Altogether,

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

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

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

where

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

The weights are positive if and only if

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

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

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

We now consider the case in which log wage residuals are given by equation (25), where we now define $\tilde{\mu}^{o}_{j,t} \equiv \mu^{o}_{t}\alpha^{o}_{j,t}$ and $\overline{\theta}^{o_{i,t}}_{i,t} \equiv \sum_{j=1}^{J} \tilde{\mu}^{o_{i,t}}_{j,t}\theta_{i,j,t}$.

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

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

$$\text{Cov}(w_{t-1}, w_{t'}|o, o') = \mu^{o}_{t-1} \sum_{j=1}^{J} \alpha^{o}_{j,t-1} \text{Cov}(\theta_{j,t'}, \overline{\theta}^{o'}_{t'}|o, o'). \tag{51}$$

Together, equations (50) and (51) imply

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

where

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

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

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

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

and equation (50), the first term in equation (49) is

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

The second term in equation (49) is

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

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

Therefore, the long autocovariances in equations (50) and (51) that are not conditioned on $o_{t'}$ are given by

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

where we use the law of total covariance:

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

Combining the two equations gives

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

where

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

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

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

The log wage residual autocovariance for occupational stayers is

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

Under Assumption 8, the first term is

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

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

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

Therefore, the long autocovariances for occupational stayers are

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

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

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

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

The two long residual autocovariances imply

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

where

$$\zeta_{t',t-1}^o \equiv \frac{\Pr(o_t = o_{t-1} = o | o_t = o_{t-1}) \Psi_{t',t-1}^o \mu_{t-1}^o}{\sum_{o'} \Pr(o_t = o_{t-1} = o' | o_t = o_{t-1}) \Psi_{t',t-1}^{o'} \mu_{t-1}^{o'}}.$$

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

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

# D MD Estimation and Standard Errors

## D.1 MD Estimation

For a given parameter vector $\mathbf{\Lambda}$, we can compute theoretical counterparts for $\text{Cov}(w_t, w_{t'} | s, c)$, 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

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

where $\Gamma$ is described in Table 4; $\widehat{\text{Cov}}(w_t, w_{t'} | s, C_j)$ 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 $\mathbf{\Lambda}$.

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 $\Gamma = \{s, j, t, t' | 1970 \leq t' \leq t \leq 2012, t - t' \geq 6, s \in \{\text{Non-college, College}\}, j \in \{3, 4\}\}$; $\widehat{\text{Cov}}(w_t, w_{t'}|s, E_j)$ 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 $\mathbf{\Lambda}$. In Sections 3.6 and 4, we also include covariance moments for less-experienced workers, $E_1$ and $E_2$, covering the same time periods.

## D.2 Standard Errors

Consider the case of experience-based moments, and let $m = 1, 2, \ldots, M$ be the index of all moments. Let $d_{i,m}$ be the indicator of whether individual $i$ contributes to the $m^{th}$ moment $\text{Cov}(w_t, w_{t'}|s, E_j)$. 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 $\boldsymbol{W} = \text{diag}(\frac{N^2}{N_1^2}, \frac{N^2}{N_2^2}, \ldots, \frac{N^2}{N_M^2})$ and $N_m = \sum_{i=1}^{N} d_{i,m}$.

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

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

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

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

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

$$\hat{\boldsymbol{\Omega}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}(\boldsymbol{z}_i, \hat{\boldsymbol{\Lambda}})\boldsymbol{h}(\boldsymbol{z}_i, \hat{\boldsymbol{\Lambda}})^\top,$$

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

# E PSID Data Details and Additional Results

## E.1 Data Description

The PSID is a longitudinal survey of a representative sample of individuals and families in the U.S. beginning in 1968. The survey was conducted annually through 1997 and biennially since. We use data collected from 1971 through 2013. Since earnings were collected for the year prior to each survey, our analysis studies hourly wages from 1970 to 2012.

Our sample is restricted to male heads of households from the core (SRC) sample and excludes those from any PSID oversamples (SEO, Latino) as well as those with zero individual weights.[69] We use earnings (total wage and salary earnings, excluding farm and business income) from any year these men were ages 16–64, had potential experience of 1–40 years, had positive wage and salary income, had positive hours worked, and were not enrolled as a student.

Our sample is composed of 92% whites, 6% blacks, and 1% hispanics with an average age of 39 years old. We create seven education categories based on current years of completed schooling: 1-5 years, 6-8 years, 9-11 years, 12 years, 13-15 years, 16 years, and 17 or more years. College workers are defined as those with more than 12 years of schooling. In our sample, 13% of respondents finished less than 12 years of schooling, 35% had exactly 12 years of completed schooling, 21% completed some college (13-15 years), 21% completed college (16 years), and 10% had more than 16 years of schooling.

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

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

To examine whether attrition affects the residual autocovariances reported in Figure 3, Figure E-2 shows the autocovariances, $\text{Cov}(w_b, w_t)$ for $6 \leq t - b \leq 16$, where the samples for each line (representing different base years, $b$) are restricted to those individuals observed in the base year as well as at least one of the last two years used for that line (i.e. $t - b = 15$ or 16 in early years or $t - b = 14$ or 16 in later years with biannual surveys). Comparing Figures 3 and E-2, the autocovariance patterns are quite similar, indicating little effect of sample attrition (due to non-response or retirement) on the key moments used in our analysis.

Figure E-3 shows the residual autocovariances for individuals with 1–14 years of experience in the

[69] The earnings questions we use are asked only of household heads. We also restrict our sample to those who were heads of household and not students during the survey year of the observation of interest as well as two years earlier. Our sampling scheme is very similar to that of Moffitt and Gottschalk (2012).

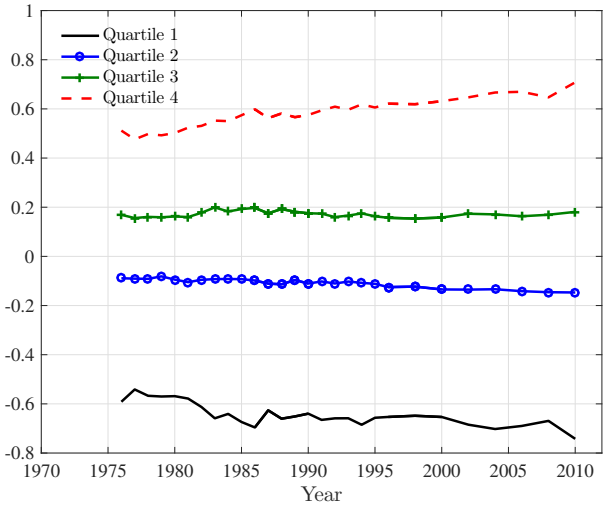



Figure E-1: Average Log Wage Residuals by Quartile

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

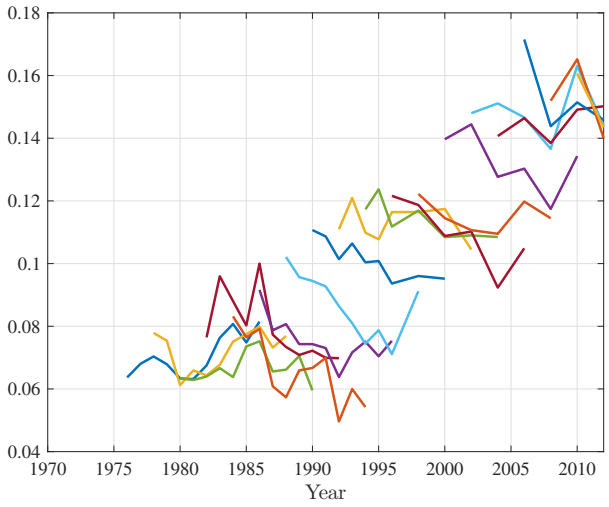



Figure E-2: Log Wage Residual Autocovariances ('Balanced' Sample)

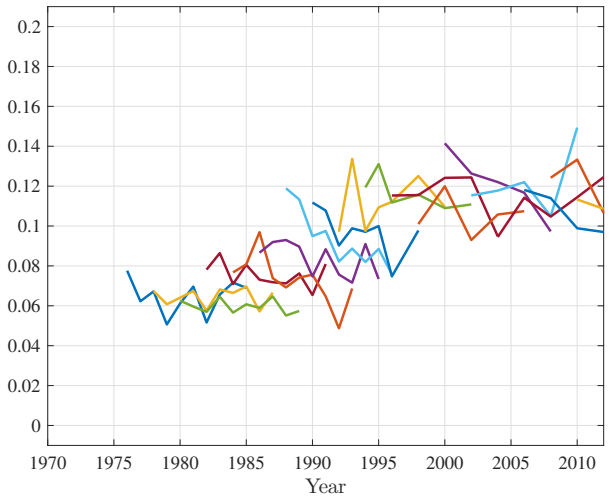



Figure E-3: Autocovariances for Log Wage Residuals (1–14 Years of Experience)

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

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

| | 1979–1980 | 1981–1983 | 1984–1986 | 1987–1989 | 1990–1992 | 1993–1995 |
|---|---|---|---|---|---|---|
| | | | A. All men | | | |
| $\Delta_2\mu_t/\mu_{t-2}$ | -0.052 | -0.088* | -0.031 | -0.100* | -0.036 | -0.104* |
| | (0.050) | (0.043) | (0.050) | (0.046) | (0.044) | (0.045) |
| Observations | 928 | 1,323 | 1,244 | 1,211 | 1,244 | 1,300 |
| 1st stage $F$-Statistic | 117.23 | 132.19 | 66.26 | 130.53 | 132.83 | 201.62 |
| | | | B. Non-college men | | | |
| $\Delta_2\mu_t/\mu_{t-2}$ | -0.108 | 0.009 | -0.019 | -0.101 | -0.051 | -0.105 |
| | (0.061) | (0.062) | (0.072) | (0.070) | (0.066) | (0.065) |
| Observations | 552 | 777 | 678 | 609 | 555 | 542 |
| 1st stage $F$-Statistic | 66.06 | 59.12 | 24.04 | 55.22 | 65.32 | 72.32 |
| | | | C. College men | | | |
| $\Delta_2\mu_t/\mu_{t-2}$ | -0.031 | -0.166** | -0.003 | -0.088 | -0.024 | -0.104 |
| | (0.068) | (0.053) | (0.074) | (0.060) | (0.059) | (0.060) |
| Observations | 314 | 491 | 509 | 524 | 594 | 758 |
| 1st stage $F$-Statistic | 73.87 | 90.56 | 99.30 | 71.46 | 66.14 | 142.24 |

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

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

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

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

## E.3 GMM Estimates of Skill Returns, Over-Identification Tests, and Variance of Skill Growth

In this appendix, we report GMM estimates for the returns to skill using the same model and moments (i.e. lagged residuals serve as instruments) as with our 2SLS approach in Section 3.3 along with $J$-statistics to test for overidentification. We also report analogous GMM estimates that use both past and future wage residuals as instruments, reporting $J$-statistics to test the validity of the latter. Finally, we combine estimates using past vs. future residuals as instruments to estimate the variance of skill growth relative to lagged skill levels.

To begin, rewrite the two-period wage growth equation (10) as follows:

$$\Delta_2 w_{i,t} = \left(\frac{\Delta_2 \mu_t}{\mu_{t-2}}\right) w_{i,t-2} + u_{i,t}, \tag{52}$$

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:

$$\mathrm{E}[w_{t'} u_t] = 0, \qquad \text{for } t' \leq t-2-k. \tag{53}$$

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

$$\mathrm{E}[w_{t''} u_t] = 0, \qquad \text{for } t'' \geq t+k. \tag{54}$$

Equation (54) 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 $t'' \geq t+k$,

$$\frac{\mathrm{Var}(\Delta_2\theta_t)}{\mathrm{Var}(\theta_{t-2})} = \left[\frac{\mathrm{Cov}(\Delta_2 w_t, w_{t''})}{\mathrm{Cov}(w_{t-2}, w_{t''})} - \frac{\mathrm{Cov}(\Delta_2 w_t, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right]\left[1 + \frac{\mathrm{Cov}(\Delta_2 w_t, w_{t'})}{\mathrm{Cov}(w_{t-2}, w_{t'})}\right]^{-1}. \tag{55}$$

### E.3.1 Overidentification Tests

We begin by testing the moments in equation (54) 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 (53) and

the first two available out of $w_{i,t+6}, w_{i,t+7}, w_{i,t+8}, w_{i,t+9}$ for (54).

Table E-3 reports the two-step optimal GMM estimates (allowing for heteroskedasticity and serial correlation within individual) for the coefficient on $w_{i,t-2}$ along with Hansen's $J$-statistics when estimating the wage growth equation (52). Panel A reports estimates when moments from both equations (53) and (54) are used (i.e., lags and leads), while Panel B reports estimates when only the moment condition from equation (53) is used (i.e., lags only). The sample is restricted to be the same in both panels.[70]

Comparing the $J$-statistics in Panels A and B in Table E-3, we can test the validity of using leads as instruments (i.e. moments in equation (54)). Since the differences are greater than 5.991 (critical value for $\chi^2_2$ at significance level 0.05) except for 1979–1980 and 2002–2004, we reject the 'leads' moments in equation (54) 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 $\chi^2_1$ at significance level 0.05), implying that we cannot reject the lags as instruments (i.e. moments in equation (53)) 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.

[70] Because use of both leads and lags requires observations that are as many as 19 years apart, this restriction reduces the sample size substantially relative to that used in our baseline 2SLS analysis (see Tables 2 and 3). Panel A of Table E-4 below reports GMM estimates when this sample selection is not imposed. Those results are directly comparable and quite similar to those in Tables 2 and 3.

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

| | 1979–80 | 1981–83 | 1984–86 | 1987–89 | 1990–92 | 1993–95 | 1996–2000 | 2002–04 |
|---|---|---|---|---|---|---|---|---|
| A. 2 Nearest Valid Lags and 2 Nearest (Potentially Valid) Leads as Instruments | | | | | | | | |
| Coeff. on $w_{i,t-2}$ | -0.019 | 0.088* | 0.053 | 0.007 | -0.030 | 0.026 | 0.008 | 0.022 |
| | (0.053) | (0.044) | (0.046) | (0.034) | (0.038) | (0.035) | (0.0235) | (0.035) |
| Observations | 818 | 1,251 | 1,325 | 1,356 | 1,313 | 1,311 | 1,375 | 777 |
| $J$-Statistic | 4.400 | 10.392 | 11.743 | 9.579 | 9.461 | 6.991 | 8.922 | 1.646 |
| B. 2 Nearest Valid Lags as Instruments | | | | | | | | |
| Coeff. on $w_{i,t-2}$ | -0.070 | -0.010 | -0.065 | -0.057 | -0.103* | -0.025 | -0.041 | -0.003 |
| | (0.056) | (0.053) | (0.055) | (0.040) | (0.046) | (0.039) | (0.029) | (0.0389) |
| Observations | 818 | 1,251 | 1,325 | 1,356 | 1,313 | 1,311 | 1,375 | 777 |
| $J$-Statistic | 0.009 | 0.187 | 0.632 | 0.869 | 0.064 | 0.238 | 0.107 | 0.016 |
| C. $p$-Values for $J$-Tests of the Validity of Leads as Instruments | | | | | | | | |
| Leads | 0.111 | 0.006 | 0.004 | 0.013 | 0.009 | 0.034 | 0.012 | 0.443 |
| Lags | 0.924 | 0.665 | 0.427 | 0.351 | 0.800 | 0.626 | 0.744 | 0.899 |

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

### E.3.2 Inferring Relative Magnitude of Skill Shocks

Table E-4 reports GMM estimates using only lags or leads as instruments where all available observations are used (i.e., samples are not restricted to be the same across specifications). Panel A reports estimates when only the moments in equation (53) are used (i.e., 2 nearest valid lags). These results are analogous to the 2SLS estimates in Tables 2 and 3, using the same samples. Comparing estimates across the tables, we see that they are quite similar. Panel B reports GMM estimates when only the moments in equation (54) 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 (55) to estimate the relative importance of skill growth shocks. These estimates are reported in Panel C. The variance of (two-year) skill growth relative to the variance of prior skill levels ranges from 0.16 to 0.29 over our entire sample period.

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

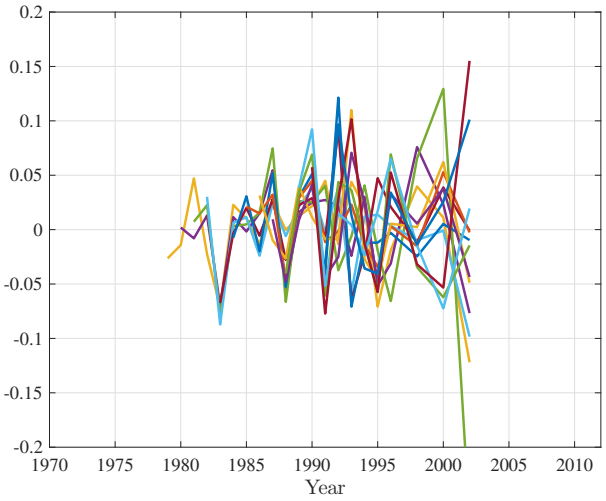
| | 1979–80 | 1981–83 | 1984–86 | 1987–89 | 1990–92 | 1993–95 | 1996–2000 | 2002–04 |
|---|---|---|---|---|---|---|---|---|
| A. 2 Nearest Valid Lags as Instruments | | | | | | | | |
| Coeff. on $w_{i,t-2}$ | -0.033 | -0.045 | -0.044 | -0.084* | -0.083* | -0.067 | -0.076* | -0.090* |
| | (0.045) | (0.038) | (0.038) | (0.033) | (0.035) | (0.035) | (0.025) | (0.035) |
| Observations | 1,349 | 2,077 | 2,188 | 2,245 | 2,189 | 2,095 | 2,122 | 1,377 |
| B. 2 Nearest (Potentially Valid) Leads as Instruments | | | | | | | | |
| Coeff. on $w_{i,t-2}$ | 0.165* | 0.229* | 0.193* | 0.099* | 0.067 | 0.087* | 0.073* | 0.115* |
| | (0.059) | (0.053) | (0.047) | (0.042) | (0.043) | (0.038) | (0.028) | (0.039) |
| Observations | 1,500 | 2,229 | 2,159 | 2,100 | 2,042 | 1,994 | 2,178 | 1,249 |
| C. Estimated Shock Variances Relative to Skill Variances | | | | | | | | |
| $\text{Var}(\Delta_2\theta_t)/\text{Var}(\theta_{t-2})$ | .204 | 0.287 | 0.248 | 0.200 | 0.163 | 0.166 | 0.161 | 0.225 |

Notes: GMM estimates for a regression of $(w_{i,t} - w_{i,t-2})$ on $w_{i,t-2}$. Panel A uses 2 nearest available lags as instruments from $(w_{t-8}, w_{t-9}, w_{t-10})$. Panel B uses 2 nearest available leads as instruments from $(w_{t+6}, ..., w_{t+9})$. Panel C reports estimates of skill growth shock variance relative to skill variance based on equation (55).

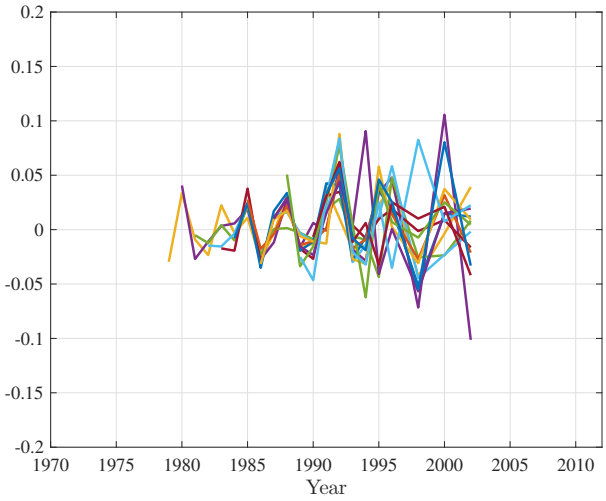
* denotes significance at 0.05 level.

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

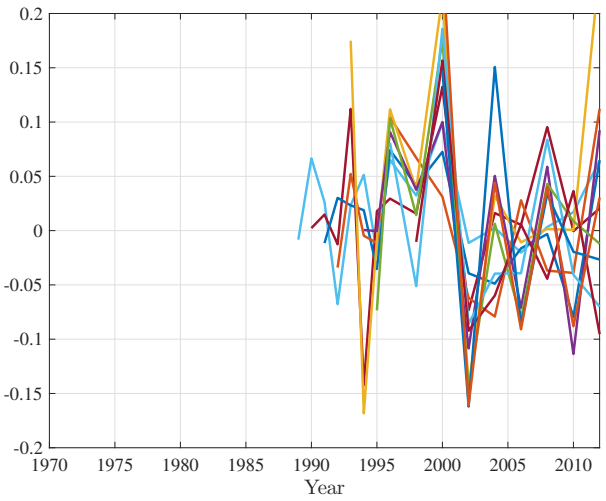
This appendix shows results for $\text{Cov}(\Delta(w_t/\mu_t), w_{t'})$ in PSID.



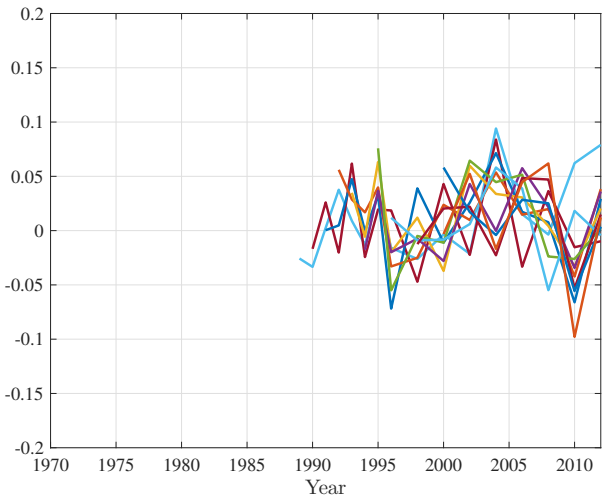

(a) Cohort Group 3, Non-College




(b) Cohort Group 3, College




(c) Cohort Group 4, Non-College




(d) Cohort Group 4, College

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

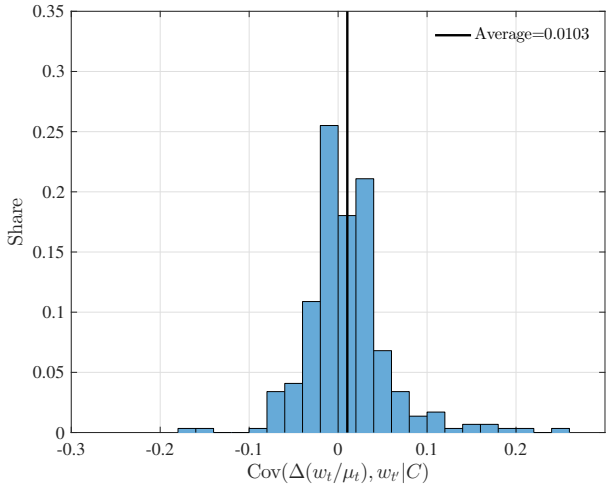
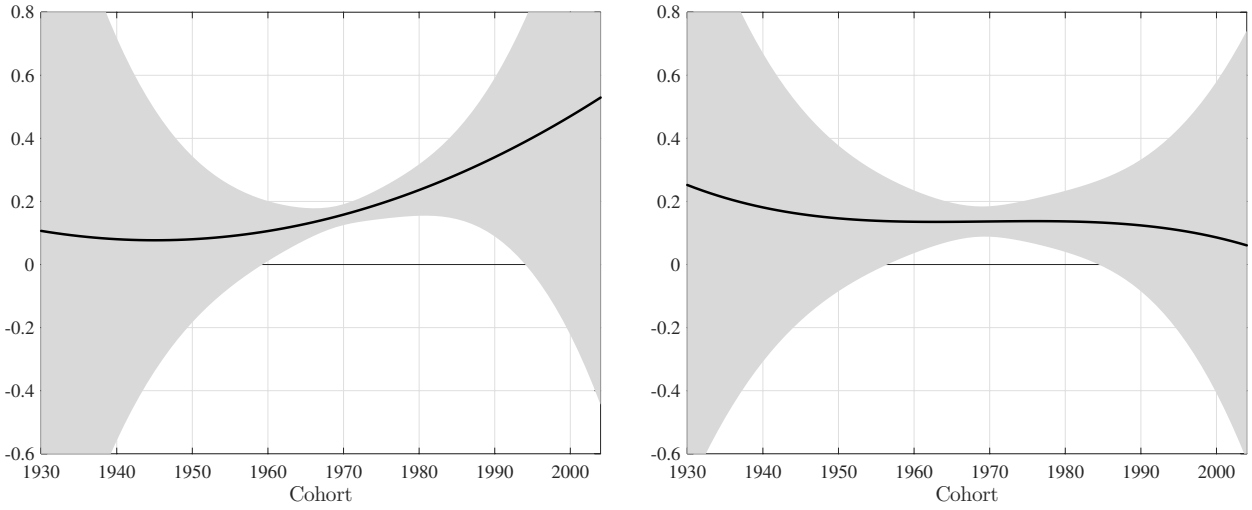



Figure E-5: Distribution of $\text{Cov}(\Delta(w_t/\mu_t), w_{t'})$ for all $(t, t', C)$ for for Low-Experience Men

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

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




(a) Non-College Men

(b) College Men

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

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

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

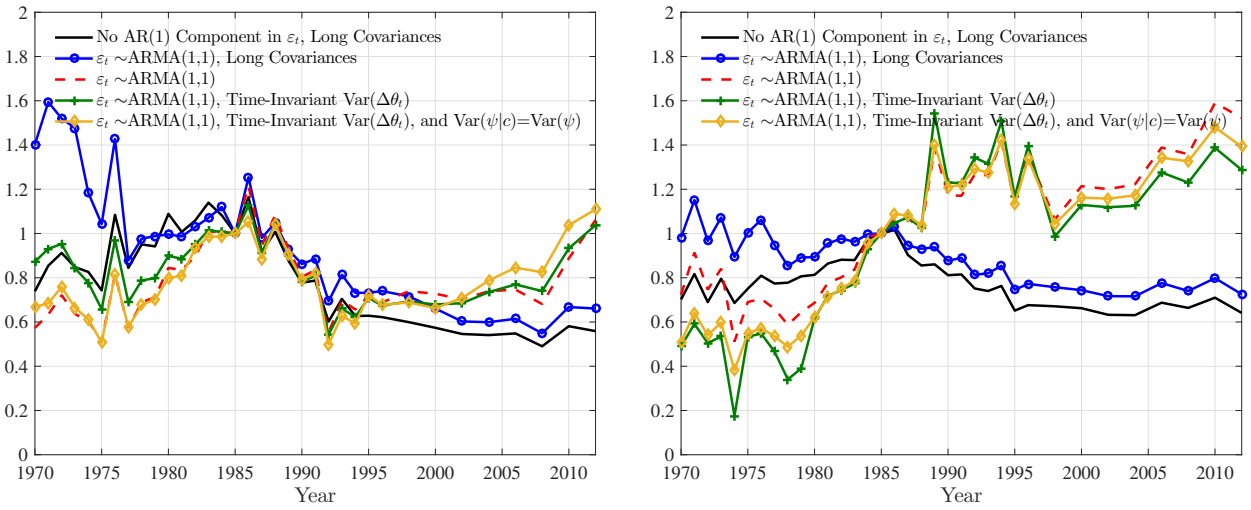
As with most of the earnings dynamics literature, Haider (2001) and Moffitt and Gottschalk (2012) assume $\varepsilon_t \sim \text{ARMA}(1,1)$, which is very similar, though not identical, to our specification in equation (15) with $k = 2$. We prefer the latter given our desire to maintain a completely flexible (time-varying) autocorrelation structure for transitory shocks, $\tilde{\varepsilon}_t$. Regardless, the two specifications yield very similar estimated $\mu_t$ series when using the same long autocovariances (i.e., $\text{Cov}(w_t, w_{t'})$ for $|t - t'| \geq 6$) in MD estimation, as can be seen from comparing the estimated returns given by the blue lines with circles in Figure E-7 and the returns reported in Figure 13. The former assumes $\varepsilon_t \sim \text{ARMA}(1,1)$, while the latter assumes $\varepsilon_t$ contains an AR(1) component, $\varphi_t$, and transitory component (with $k = 6$), $\tilde{\varepsilon}_t$, as described by

equation (15).[71] Figure E-7 also reproduces estimates from our baseline specification, which assumes that $\varepsilon_t$ only contains a transitory component (with $k = 6$). Altogether, these estimates indicate that accounting for persistent non-skill shocks has little effect on estimated returns to skills over time.

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

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

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




(a) Non-College (b) College

Figure E-7: Estimated $\mu_t$ under Different Restrictions: Non-College and College

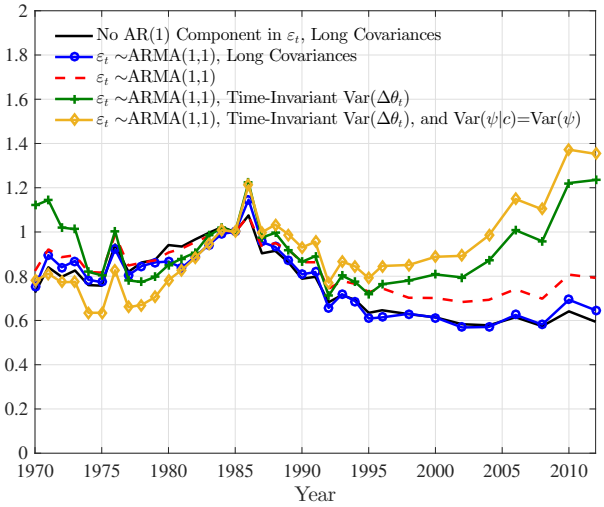



Figure E-8: Estimated $\mu_t$ under Different Restrictions: All Men

## E.7 Estimation with Job Stayers in PSID

This appendix considers log wage residuals that include firm fixed effects as in equation (17). 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 9.** *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 9, the IV estimator for job stayers, for $t - t' \geq k + 1$, is given by:

$$\begin{aligned} \frac{\mathrm{Cov}(\Delta w_t, w_{t'} | j_t = j_{t-1})}{\mathrm{Cov}(w_{t-1}, w_{t'} | j_t = j_{t-1})} &= \frac{\Delta\mu_t \, \mathrm{Cov}(\theta_{t-1}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})}{\mathrm{Cov}(\kappa_{j_{t-1}}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1}) + \mathrm{Cov}(\mu_{t-1}\theta_{t-1}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})} \\ &= \frac{\Delta\mu_t}{\mu_{t-1}} \times \left(1 + \frac{\mathrm{Cov}(\kappa_{j_{t-1}}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})}{\mathrm{Cov}(\mu_{t-1}\theta_{t-1}, \kappa_{j_{t'}} + \mu_{t'}\theta_{t'} | j_t = j_{t-1})}\right)^{-1}. \end{aligned} \tag{56}$$

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 $\kappa_{j_t}$. With exogenous job mobility, these conditions are equivalent to assuming that both $\kappa_{j_{t-1}}$ and $\theta_{t-1}$ are positively correlated with lagged residuals $w_{t'}$ for job stayers. Importantly, equation (56) suggests that our IV estimates (reported in Figure 17) likely under-estimate the actual decline in skill returns over time.

### E.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.[72] 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 $t - 2$. Figure E-9 shows the shares of job stayers, job switchers, and missing job start years for individuals with non-missing log wage residuals in years $t$ and $t - 2$. 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. The drop in the share with missing job tenure between 2000 and 2002 is likely

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

due to a modest change in survey wording about start date(s) on the current job (asking about all start/stop dates if multiple spells rather than only the most recent start date).

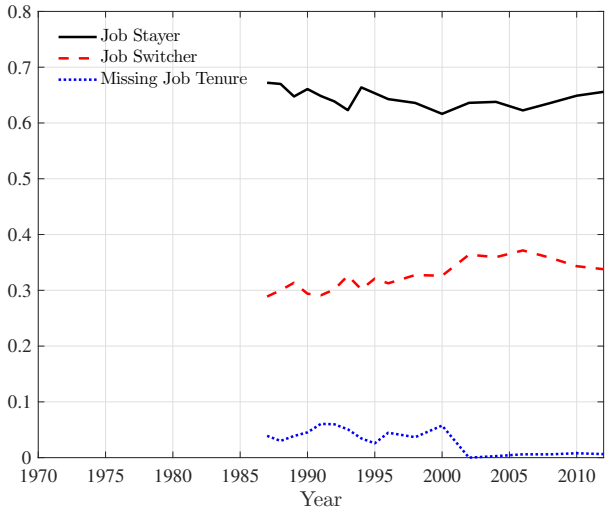



Figure E-9: Job Transition Rates from $t-2$ to $t$ (PSID)

### E.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 $\text{Cov}(\Delta\theta_{t''}, \theta_{t'}|j_t = j_{t-1}) = \text{Cov}(\Delta\theta_{t''}, \kappa_{j_{t'}}|j_t = j_{t-1}) = 0$ and $\text{Cov}(\kappa_{j_{t''}} - \kappa_{j_{t''-1}}, \theta_{t'}|j_t = j_{t-1}) = \text{Cov}(\kappa_{j_{t''}} - \kappa_{j_{t''-1}}, \kappa_{j_{t'}}|j_t = j_{t-1}) = 0$ for $t'' \geq t'+1$, then the intertemporal covariances for job stayers in the bias term in equation (56) can be written as within-period covariances, leading to

$$\frac{\text{Cov}(\Delta w_t, w_{t'}|j_t = j_{t-1})}{\text{Cov}(w_{t-1}, w_{t'}|j_t = j_{t-1})} = \frac{\Delta\mu_t}{\mu_{t-1}} \times \left(1 + \frac{\mu_{t'}}{\mu_{t-1}} \underbrace{\frac{\text{Var}(\kappa_{j_{t'}}|j_t = j_{t-1}) + \text{Cov}(\kappa_{j_{t'}}, \mu_{t'}\theta_{t'}|j_t = j_{t-1})}{\text{Var}(\mu_{t'}\theta_{t'}|j_t = j_{t-1}) + \text{Cov}(\kappa_{j_{t'}}, \mu_{t'}\theta_{t'}|j_t = j_{t-1})}}_{\equiv \Upsilon_{t'}}\right)^{-1}. \quad (57)$$

With our IV estimator for stayers and estimates of $\Upsilon_{t'}$, equation (57) can be used to identify $\mu_t$ 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 E-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 $\mu_t$ 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 E-5: Within-Period Covariances (from Table III of Song et al. (2018)) and Implied $\Upsilon_t$

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

Figure E-10 reports estimates for $\mu_t$ that correct for bias associated with firm fixed effects using the values of $\Upsilon_{t'}$ reported in Table E-5. These are based on (uncorrected) IV estimates for job stayers like those reported in Figure 17; 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 (57).[73] Compared to (uncorrected) estimates reported in Figure E-10, the (corrected) $\mu_t$ series displayed in Figure E-10 implies very little bias for college men but notably stronger bias for non-college men.

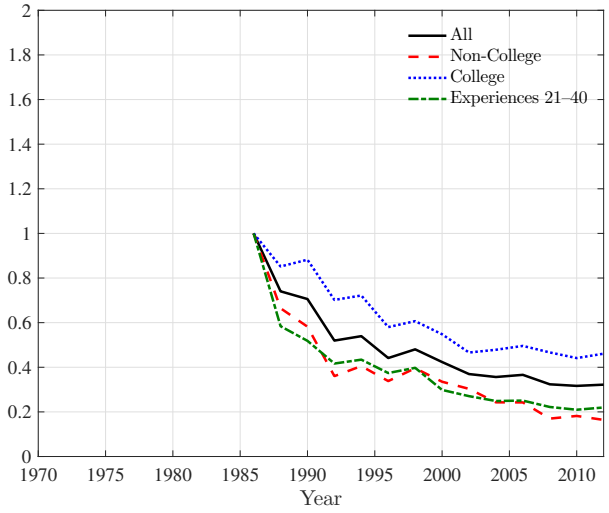



Figure E-10: $\mu_t$ Corrected for Bias due to Firm Fixed Effects

[73] Note that estimates in Figure E-10 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 $\mu_t$ based on equation (57) modified for two-year growth rates, the IV estimates, and $\Upsilon_{t'}$ from Table E-5, normalizing $\mu_{t'} = 1$ for $t' \leq 1986$.

## E.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.[74] 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 in social occupations. Conversely, around 76% of observations in social occupations are also in cognitive occupations.

### E.8.1 GMM Estimation using Occupation Stayers and Switchers

We estimate occupation-specific $\gamma_t^o$ and $\mu_t^o$ 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 (20):

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

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 $\gamma_t^o$ and $\mu_t^o$, each with 14 knots in $t$, restricting to moment conditions with at least 20 observations (9 switcher

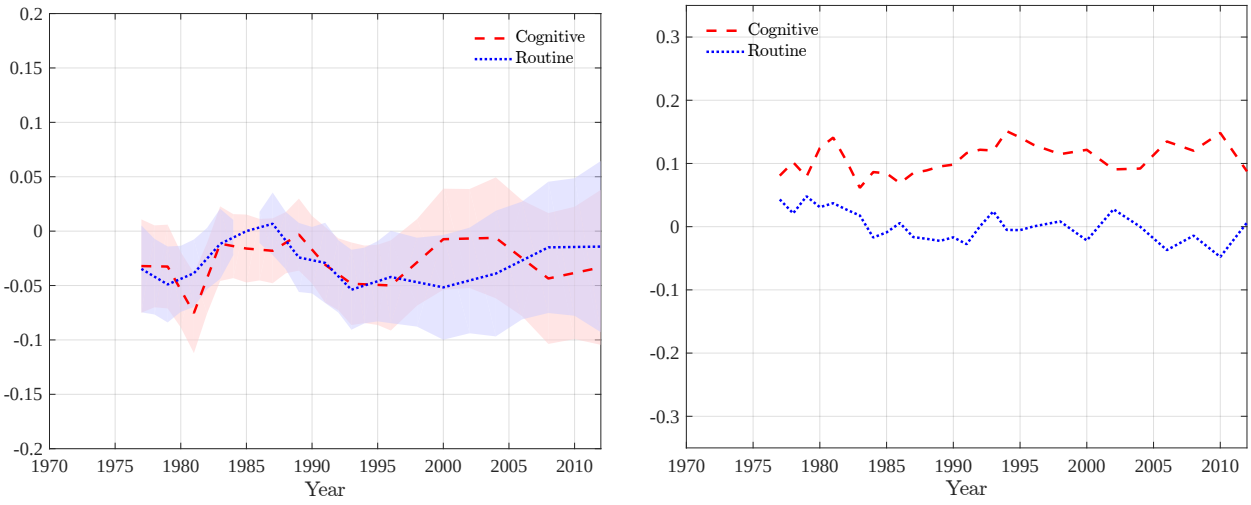
[74] Since we pool observations across several years (assuming constant growth of skill returns within each pooled sample) for 2SLS estimation, the change in skill return between 2000 and 2002 reflects an extrapolation from adjacent years.

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 $\mu_t$ and $\gamma_t^o$ series are nearly identical to their counterparts reported in Figure 19, while the $J$-statistic comparing the unrestricted and restricted criterion functions equals 20.08 and is distributed $\chi^2_{14}$ under the null hypothesis of equal skill returns. Thus, we cannot reject the null of identical returns at conventional levels ($p$-value = 0.128).

For comparison with Cortes (2016), Figure E-11 reports estimates of $\gamma_t^o$ and $\mathrm{E}[\theta_t|o_t]$ when imposing time-invariant $\mu_t^o = \mu^o$. Estimated time profiles for $\gamma_t^o$ and $\mathrm{E}[\theta_t|o_t]$ are notably flatter than their counterparts allowing for variation in skill returns (see Figures 19(b) and 20(b)).

Finally, we use the same estimation strategy, restricting the sample to men who had 21–40 years of experience in year $t$. Due to fewer observations, this reduces the number of moments used in estimation to 261. These results are reported in Figures E-12 and E-13.




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

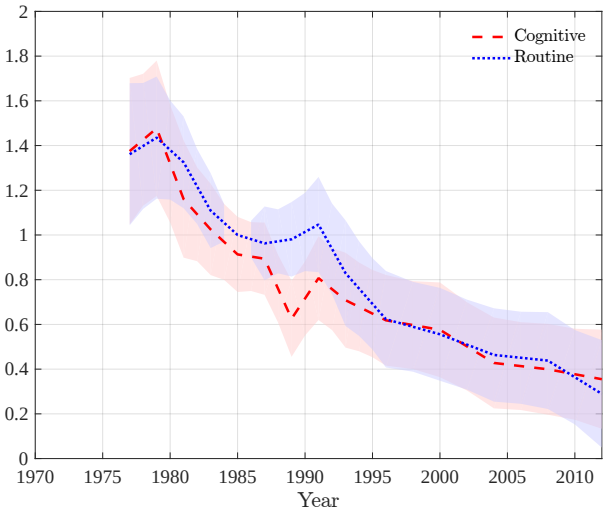
Figure E-11: GMM estimates imposing time-invariant $\mu_t^o$

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

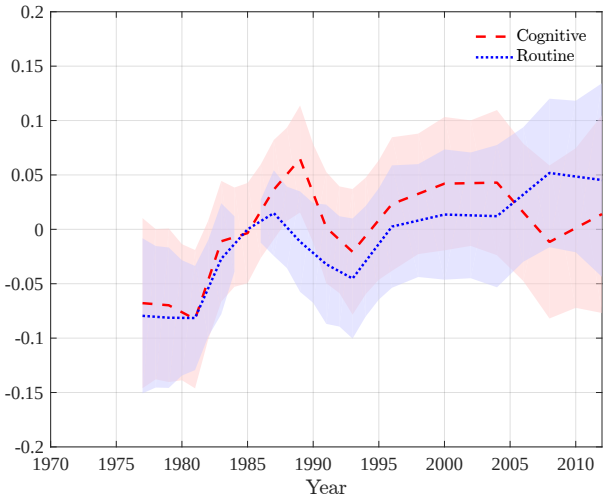
### E.8.2 2SLS Estimated Returns for Occupational Stayers

Tables E-6 and E-7 report 2SLS estimates (and standard errors) for skill return growth underlying Figures 18 and 21 in the main text. First-stage $F$-statistics for the instruments are also reported.

Figure E-14 reports 2SLS estimates based on occupation-stayers in years $t-2$ to $t$ based on different

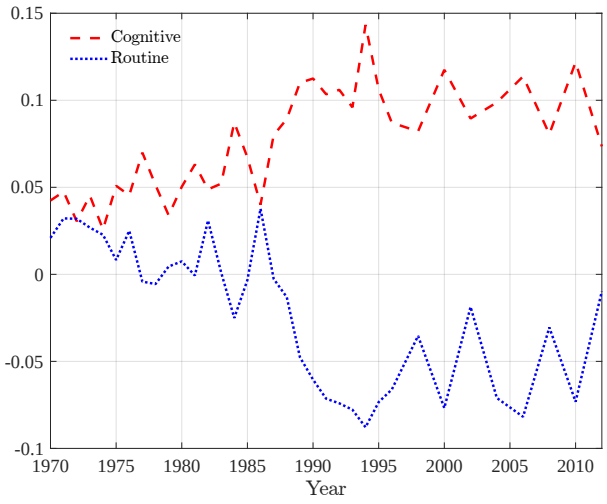



(a) $\mu_t^o$

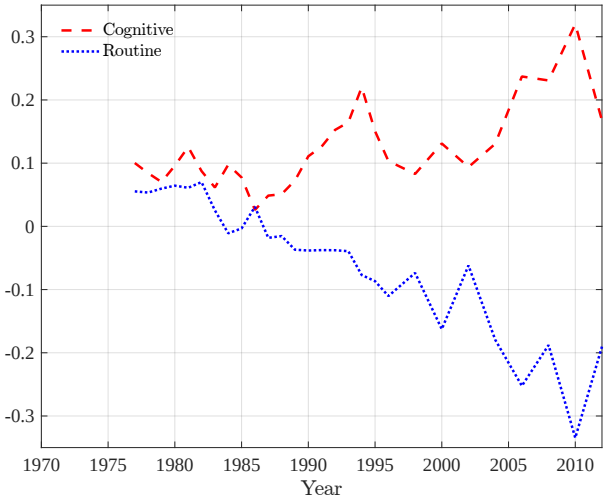



(b) $\gamma_t^o$

Figure E-12: GMM estimates of $\mu_t^o$ and $\gamma_t^o$, 21–40 years of experience in $t$




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

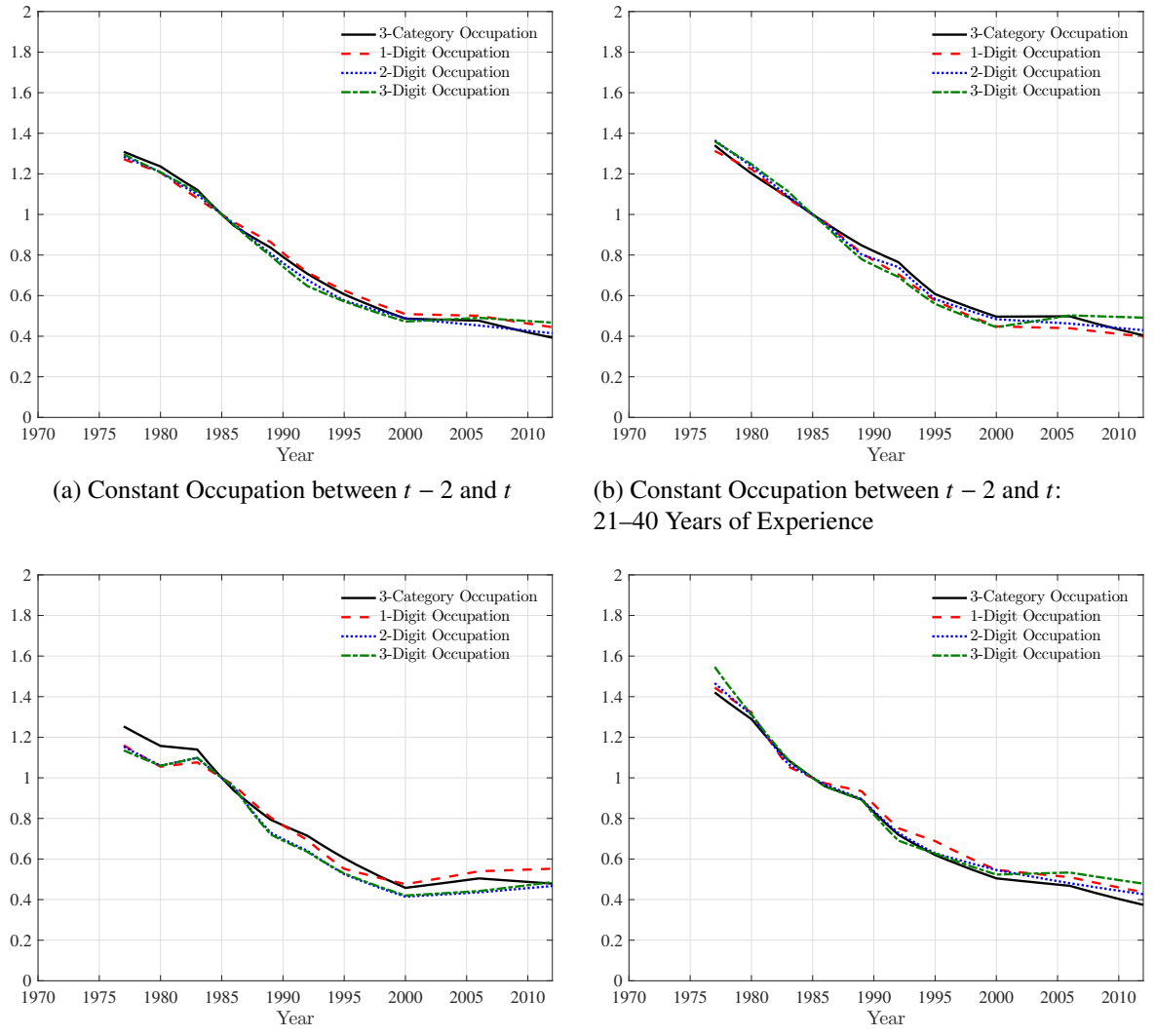



(b) $\mathrm{E}\left[\theta_t|o_t\right]$

Figure E-13: Average log wage residual and skill by occupation, 21–40 years of experience in $t$

occupation classifications.[75] In all cases, $w_{i,t-8}$ and $w_{i,t-9}$ (or $w_{i,t-10}$) residuals are used as instruments. Estimated skill return patterns are very similar regardless of how narrowly we define occupations.

As noted in the text, estimated return growth for stayers in occupation $o_t = o_{t-1} = o$ should not depend on earlier occupation ($o_{t'}$) under Assumption 5. Estimates reported in Figures E-15 and E-16 indicate very similar estimated skill return profiles for occupation stayers with $o_{t'} = o$ vs. $o_{t'} \neq o$.




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

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

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

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

Figure E-14: $\mu_t$ Implied by 2SLS Estimates for Occupational Stayers from $t-2$ to $t$

[75] These estimates are based on the same subperiods as reported in Tables E-6 and E-7. The 3-category occupation estimates reflect those remaining within occupations classified as either cognitive, routine, or manual based on Cortes (2016).

Table E-6: 2SLS Estimates for Occupational Stayers, 1979–1995

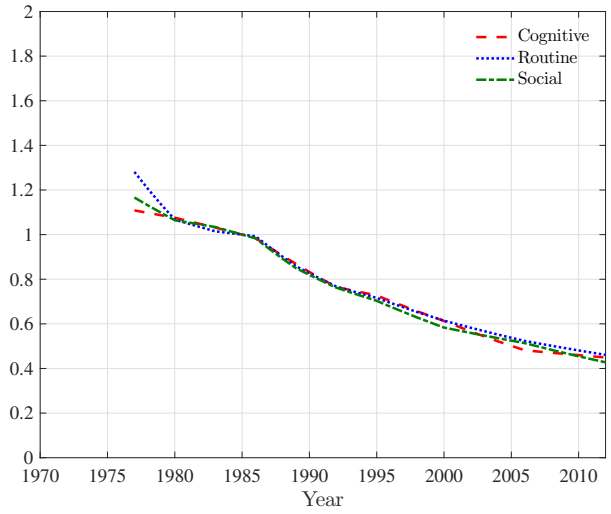
| | 1979–1980 | 1981–1983 | 1984–1986 | 1987–1989 | 1990–1992 | 1993–1995 |
|---|---|---|---|---|---|---|
| | | | A. Non-College Men | | | |
| $\Delta_2\mu_t/\mu_{t-2}$ | -0.045 | 0.024 | -0.090 | -0.171* | -0.081 | -0.115 |
| | (0.060) | (0.064) | (0.060) | (0.053) | (0.059) | (0.075) |
| Observations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

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

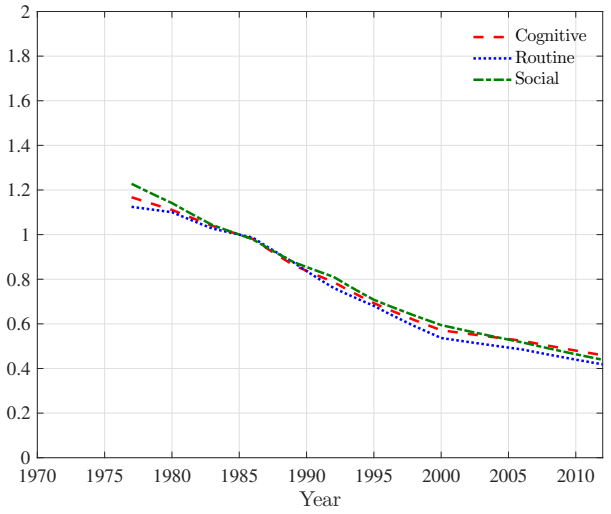
Table E-7: 2SLS Estimates for Occupational Stayers, 1996–2012

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

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

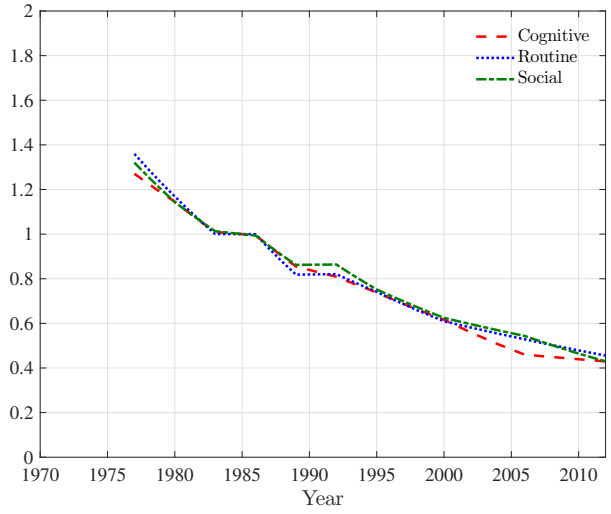



(a) Same Occupation in $t'$ ($o_t = o_{t-2} = o_{t'}$)

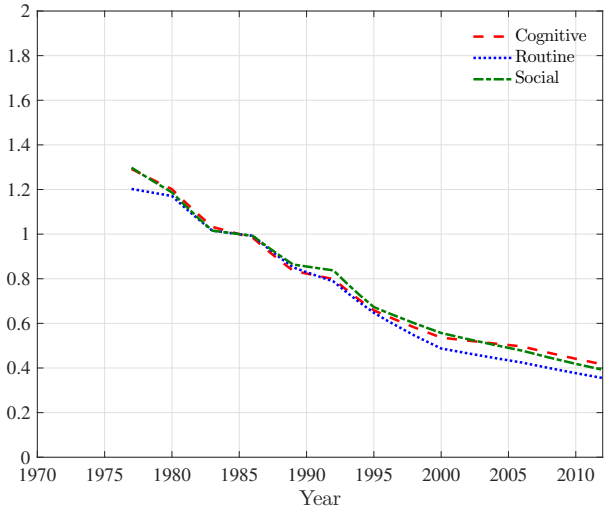



(b) Different Occupation in $t'$ ($o_t = o_{t-2} \neq o_{t'}$)

Figure E-15: $\mu_t^o / \mu_{1985}^o$ implied by 2SLS estimates for occupation-stayers: All experience levels




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




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

Figure E-16: $\mu_t^o / \mu_{1985}^o$ implied by 2SLS estimates for occupation-stayers: 21–40 years of experience

# F HRS Data and Results

## F.1 Data Description

We use data from the Health and Retirement Study (HRS), a national U.S. panel survey of individuals over age 50 and their spouses.[76] 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.)[77] The survey has been fielded every two years since 1992, and it provides information about demographics, income, and cognition, making it ideal data for the purpose of our study. Because one of the cognitive tests (word recall) in 1992 and 1994 differs from that of later years, we use data collected from 1996 to 2018.[78]

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

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

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

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

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

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

Table F-1: Summary of Cognitive Measures

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

50 years.[80] In estimation, we use non-imputed wages and cognitive measures only. The sample contains 10,151 individuals and 43,096 person-year observations.

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

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

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

## F.2 Detailed Description of Cognitive Measures

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

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

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

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

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

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

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

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

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

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

This appendix describes data from Internal Revenue Service (IRS)/Social Security Administration (SSA) W-2 Forms linked with the Survey of Income and Program Participation (SIPP), referred to as the Gold Standard File (GSF) by the Census Bureau (U.S. Census Bureau, 2018). These data include the full SSA history of annual earnings (i.e., wage and salary) for all linked respondents from 1951 to 2011.[81] Because we use annual earnings from administrative records, annual hours of work and hourly wages are not available.

Our analysis is based on 16–69 year-old, US-born white men who could be linked to any of nine SIPP panels (1984, 1990, 1991, 1992, 1993, 1996, 2001, 2004, and 2008). The highest level of education achieved at the time of survey (asked only once in each panel) is available in 5 categories: no high school degree, high school degree, some college, college degree, and graduate degree. We map these categories to 10, 12, 14, 16, and 18 years of completed schooling in order to calculate potential experience (age - years of education - 6). Since some individuals were still young and unlikely to have completed their schooling at the time of survey, we exclude those who were under 30 years old or were enrolled in school when their education level was measured.

We focus mainly on results using Detailed Earnings Records (DER), which are uncapped and available from 1978 onward; however, we also take advantage of Summary Earnings Records (SER) available since 1951, which report earnings capped at the FICA taxable maximum. We work with log earnings residuals constructed as with the PSID and restrict observations to years when individuals were no longer enrolled in school. We trim the top and bottom 1% of DER-based earnings within year and college/non-college status by five-year experience cells, and residualize log DER-based earnings by regressing on experience indicators and interactions between education indicators and a third order polynomial in experience, separately by year and college/non-college status. Log SER-based earnings – used only as instruments in our analysis – are residualized by subtracting median values conditional on year, education, and five-year experience cells.

Based on a worker's primary job (i.e., the job with the highest earnings), the Census Bureau classified workers into 24 occupation categories each survey wave. Table G-1 reports these occupation codes, along with our 3-category grouping of occupations (cognitive, manual, and routine). Since respondents can report different occupations in each of 3 survey waves each year, we define occupation stayers between two years as those who reported any occupation in both years.

[81] This analysis was first performed using the SIPP Synthetic Beta (SSB), while final results were obtained by Census Bureau staff using the SIPP Completed Gold Standard Files. See Reeder, Stanley, and Vilhuber (2018) and Benedetto, Stanley, and Totty (2018) for additional details on the data.

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

| Code | Occupation | 3-Category Grouping |
|---|---|---|
| 1 | Management | Cognitive |
| 2 | Business and financial operations | |
| 3 | Computer and mathematical | |
| 4 | Architecture and engineering | |
| 5 | Life, physical, and social science | |
| 6 | Community and social service | |
| 7 | Legal | |
| 8 | Education, training, and libraries | |
| 9 | Arts, design, entertainment, sports, and media | |
| 10 | Healthcare practitioner and technical | |
| 11 | Healthcare support | Manual |
| 12 | Protective service | |
| 13 | Food prep and service | |
| 14 | Building and grounds cleaning and maintenance | |
| 15 | Personal care and service | |
| 16 | Sales | Routine |
| 17 | Office and administrative support | |
| 18 | Farming, fishing, and forestry | Not classified |
| 19 | Construction and extraction | Routine |
| 20 | Installation, maintenance, and repairs | |
| 21 | Production | |
| 22 | Transportation | |
| 23 | Material moving | |
| 24 | Military | Not classified |

Figure G-1 reports log earnings inequality, along with between-group and within-group (residual) inequality, based on DER wage measures in the SIPP/W-2. The general trends are qualitatively similar to those for the PSID reported in Figure 1; although, the variance of total log earnings inequality and residual inequality is notably higher than their counterparts for log wages in the PSID.

Figure G-2 shows $\mathrm{E}\left[w_t | w_b \in Q_b^j\right]$ for different $t$ years where $Q_b^j$ reflects quartile $j$ in 'base' year $b$, while Figure G-3 shows residual autocovariances $\mathrm{Cov}(w_t, w_b)$ over years $t \geq b + 6$ for fixed base year $b$. Both figures are based on samples of non-college and college men with 21–25 years of experience in each base year, $b$. Together, these indicate declines in the return to skills over the late-1980s and 1990s, consistent with our PSID-based results.

Tables G-2 and G-3 report 2SLS estimates of skill return growth rates using SER- and DER-based lagged log earnings residuals ($w_{t-7}$), respectively, as instruments. (See Figure 22 in the paper.) Corresponding standard errors and sample sizes are also reported.

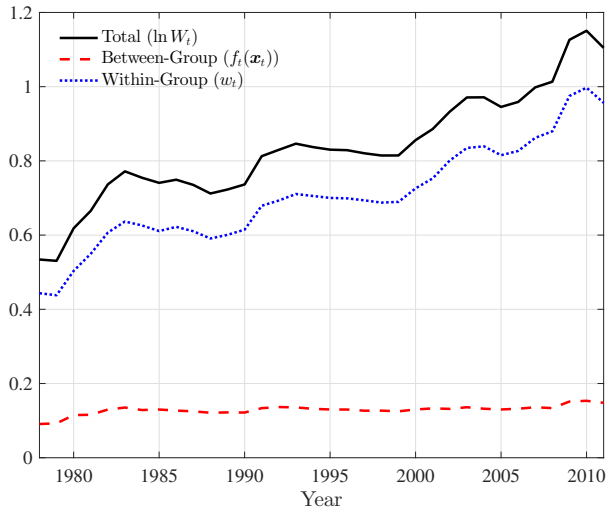
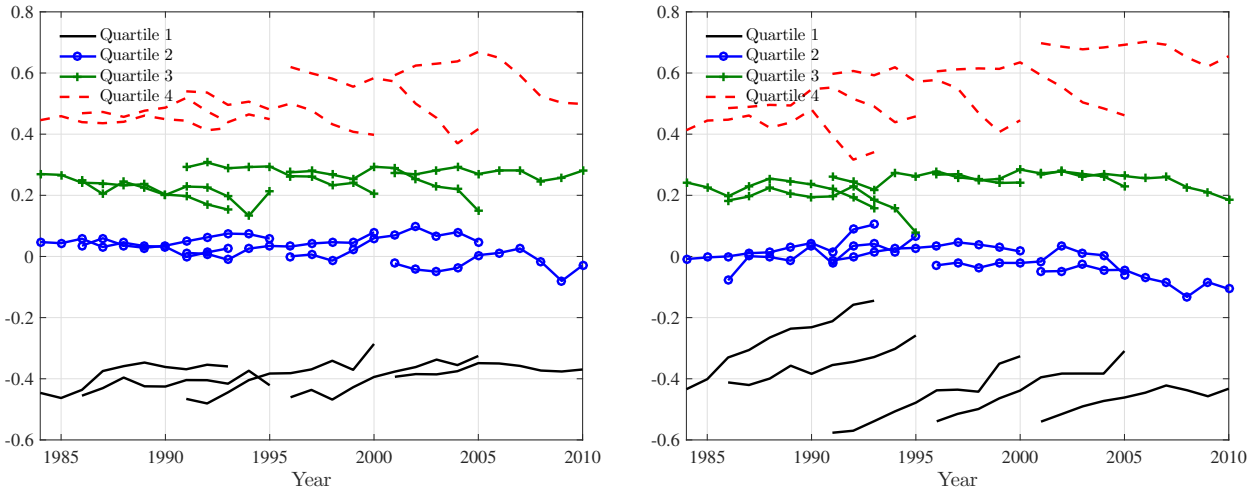
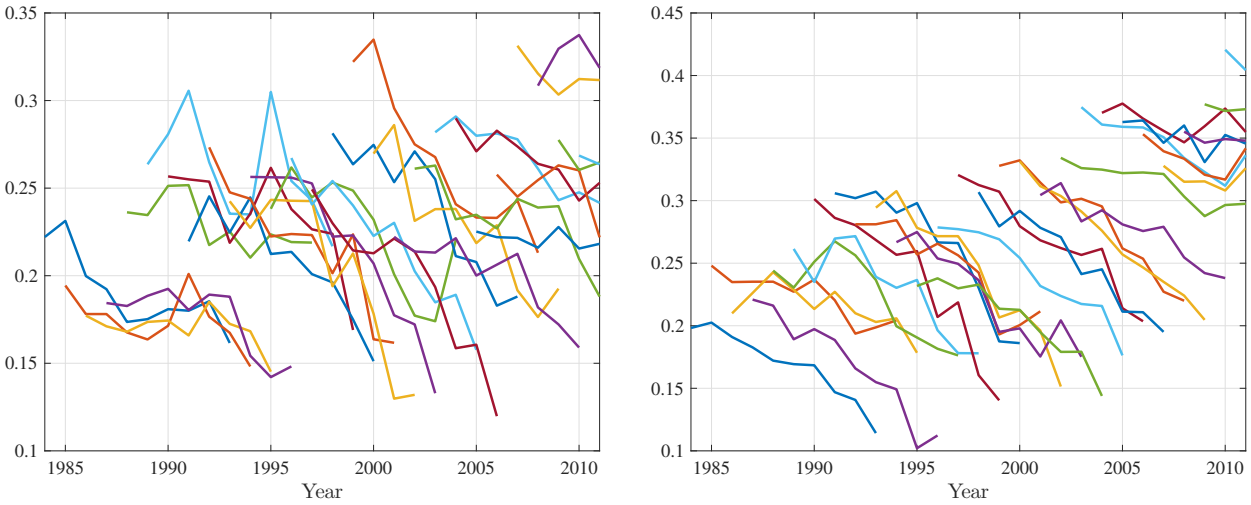



Figure G-1: Between- and within-group variances of log earnings, ages 16–64 with 5–40 years of experience (SIPP/W-2)




(a) Non-College (b) College

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




(a) Non-College (b) College

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

Table G-2: 2SLS estimates of $\Delta\mu_t/\mu_{t-1}$ (instrument: $w_{t-7}$ SER earnings), 32–40 years of experience in year $t$ (SIPP/W-2)

| Year | Non-College | | | College | | |
|---|---|---|---|---|---|---|
| | Estimate | Standard Error | Observations | Estimate | Standard Error | Observations |
| 1979 | 0.024 | 0.059 | 3,600 | 0.122 | 0.075 | 1,900 |
| 1980 | -0.037 | 0.043 | 3,700 | 0.047 | 0.063 | 2,100 |
| 1981 | 0.053 | 0.051 | 3,800 | -0.135 | 0.077 | 2,200 |
| 1982 | 0.077 | 0.045 | 3,800 | 0.015 | 0.063 | 2,400 |
| 1983 | -0.044 | 0.043 | 3,700 | -0.002 | 0.059 | 2,500 |
| 1984 | -0.024 | 0.041 | 3,700 | 0.010 | 0.069 | 2,600 |
| 1985 | -0.068 | 0.039 | 3,600 | -0.024 | 0.058 | 2,700 |
| 1986 | -0.039 | 0.040 | 3,600 | 0.010 | 0.050 | 2,900 |
| 1987 | -0.139 | 0.037 | 3,700 | -0.041 | 0.032 | 3,000 |
| 1988 | -0.043 | 0.038 | 3,800 | -0.101 | 0.030 | 3,100 |
| 1989 | 0.020 | 0.033 | 3,800 | -0.043 | 0.043 | 3,100 |
| 1990 | -0.078 | 0.033 | 3,800 | -0.009 | 0.044 | 3,200 |

Notes: Reports coefficient estimates from 2SLS regression of $\Delta w_t$ on $w_{t-1}$ using $w_{t-7}$ as an instrument. The number of observations is rounded to the nearest 100 due to confidentiality.

Table G-3: 2SLS estimates of $\Delta\mu_t/\mu_{t-1}$ (instrument: $w_{t-7}$ DER earnings), 32–40 years of experience in year $t$ (SIPP/W-2)

| Year | Non-College | | | College | | |
|---|---|---|---|---|---|---|
| | Estimate | Standard Error | Observations | Estimate | Standard Error | Observations |
| 1985 | -0.060 | 0.029 | 3,800 | -0.026 | 0.032 | 2,900 |
| 1986 | -0.072 | 0.030 | 3,800 | -0.101 | 0.026 | 3,100 |
| 1987 | -0.074 | 0.030 | 3,900 | -0.049 | 0.027 | 3,200 |
| 1988 | -0.055 | 0.037 | 4,000 | -0.084 | 0.028 | 3,400 |
| 1989 | 0.016 | 0.030 | 3,900 | -0.025 | 0.028 | 3,400 |
| 1990 | -0.065 | 0.029 | 4,000 | 0.012 | 0.030 | 3,500 |
| 1991 | 0.010 | 0.032 | 4,000 | -0.009 | 0.028 | 3,600 |
| 1992 | -0.061 | 0.026 | 4,100 | -0.096 | 0.025 | 3,800 |
| 1993 | -0.073 | 0.023 | 4,100 | -0.040 | 0.024 | 3,900 |
| 1994 | -0.025 | 0.025 | 4,200 | -0.055 | 0.027 | 4,200 |
| 1995 | 0.012 | 0.036 | 4,300 | -0.045 | 0.023 | 4,400 |
| 1996 | -0.018 | 0.028 | 4,300 | -0.094 | 0.022 | 4,700 |
| 1997 | -0.058 | 0.024 | 4,400 | -0.020 | 0.021 | 5,000 |
| 1998 | -0.088 | 0.019 | 4,400 | -0.073 | 0.020 | 5,500 |
| 1999 | -0.057 | 0.023 | 4,400 | -0.084 | 0.020 | 6,000 |
| 2000 | -0.107 | 0.028 | 4,400 | -0.048 | 0.022 | 6,600 |
| 2001 | -0.102 | 0.029 | 4,400 | -0.039 | 0.018 | 7,100 |
| 2002 | -0.045 | 0.029 | 4,300 | -0.034 | 0.021 | 7,700 |
| 2003 | -0.085 | 0.028 | 4,400 | -0.045 | 0.018 | 8,300 |
| 2004 | -0.010 | 0.029 | 4,800 | -0.019 | 0.016 | 8,900 |
| 2005 | -0.052 | 0.024 | 5,000 | -0.064 | 0.015 | 9,400 |
| 2006 | -0.002 | 0.024 | 5,100 | -0.011 | 0.014 | 9,800 |
| 2007 | -0.031 | 0.023 | 5,400 | -0.023 | 0.015 | 10,000 |
| 2008 | -0.011 | 0.022 | 5,600 | -0.014 | 0.015 | 10,500 |
| 2009 | 0.006 | 0.024 | 5,700 | -0.042 | 0.015 | 10,500 |
| 2010 | -0.001 | 0.024 | 5,700 | 0.004 | 0.014 | 10,000 |
| 2011 | -0.048 | 0.023 | 5,700 | 0.003 | 0.013 | 9,900 |

Notes: Reports coefficient estimates from 2SLS regression of $\Delta w_t$ on $w_{t-1}$ using $w_{t-7}$ as an instrument. The number of observations is rounded to the nearest 100 due to confidentiality.