Online Appendix

The Price of Experience

Hyeok Jeong*
KDI School of Public Policy and Management

Yong Kim†
Yonsei University

Iourii Manovskii‡
University of Pennsylvania

*KDI School of Public Policy and Management, 87 Hoegiro Dongdaemun-gu, Seoul 130-868, Korea. E-mail: hyeokj@gmail.com.

†School of Economics, Yonsei University, Daewoo Hall 509, 262 Seongsan-no, Seodaemun-gu, Seoul 120-749, Korea. E-mail: yongkim75@gmail.com.

‡Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA, 19104-6297 USA. E-mail: manovski@econ.upenn.edu.

A1 The Role of Age and Years of Prior Work: A Simple Two-by-Two Example

Consider an economy with 4 types of workers each earning $w^s_{j,t}$ in period t, where $s \in \{h, c\}$, h for high-school, c for college, and $j \in \{y, o\}$, y for young and o for old. Earnings of each group can be decomposed into a linear combination of payments for the two inputs – labor, L, and experience, E:

$$\begin{aligned} w^c_{j,t} &= c_t \left[a^c_j R_{L,t} + b^c_j R_{E,t}\right], \\ w^h_{j,t} &= a^h_j R_{L,t} + b^h_j R_{E,t}, \end{aligned}$$

where a^s_j denotes the time invariant quantity of input L supplied by schooling group s of age j, b^s_j denotes the corresponding supply of input E, c_t is a time varying aggregate shock to the productivity of college-educated workers, $R_{L,t}$ and $R_{E,t}$ denote the time varying economy wide prices of inputs L_t and E_t. Let $\Pi_t = R_{E,t}/R_{L,t}$ denote the relative price of the two inputs.

As in Katz and Murphy (1992), assume that all young workers exclusively supply one unit of L_t only ($a^c_y = a^h_y = 1, b^c_y = b^h_y = 0$). Old workers can potentially supply a combination of both inputs L_t and E_t, with the weights $\{a^c_o, a^h_o, b^h_o\}$ to be determined after normalizing $b^c_o = 1$. Decomposing log wages, we have

$$\begin{aligned} \ln w^c_{y,t} &= \ln c_t + \ln R_{L,t}, \\ \ln w^c_{o,t} &= \ln c_t + \ln R_{L,t} + \ln\left[a^c_o + \Pi_t\right], \\ \ln w^h_{y,t} &= \ln R_{L,t}, \\ \ln w^h_{o,t} &= \ln R_{L,t} + \ln\left[a^h_o + \Pi_t b^h_o\right]. \end{aligned}$$

The first and third equations imply that $\ln R_{L,t}$, $\ln c_t$ can be readily identified by the variation in $\ln w^h_{y,t}$ and $\ln w^c_{y,t} - \ln w^h_{y,t}$. The remaining parameters to be identified are a^c_o, a^h_o, b^h_o, and Π_t. It is the identification strategy for these parameters where our approach differs from that of Katz and Murphy.

A1.1 Katz and Murphy's Identification Strategy

Katz and Murphy assume that old workers completely stop supplying the input that young workers supply and simply set $a^c_o = a^h_o = 0$. This resolves the identification problem as the age premium for college and high school educated workers is respectively given by

$$\frac{w_{o,t}^c}{w_{y,t}^c} = \Pi_t \text{ and } \frac{w_{o,t}^h}{w_{y,t}^h} = \Pi_t b_o^h.$$

One consequence of this identifying assumption, evident from the equation above, is that the elasticity of the age premium with respect to Π_t is forced to be the same across college and high school educated workers (while these age premiums move differently in the data). As the ratio of age premiums between college and high school workers is equal to the ratio of college premiums between old an young workers, this identifying assumption also restricts the potential role of changes in Π_t in explaining the differential movement of college premiums across age groups that motivated the analysis in Card and Lemieux (2001).

A1.2 Our Identification Strategy

Our identification strategy is based on the idea that the relevant parameters can be identified if old workers of the same age differ in the number of years they have actually worked (and consequently accumulated different amounts of the experience input). To use an extreme example, note that we can identify a_o^c and Π_t if we observe some old college workers who never worked and thus accumulated no experience input. They earn $\tilde{w}_{o,t}^c = c_t R_{L,t} a_o^c$ in contrast to old college workers whose years of prior work endowed them with one unit of experience input and who earn $w_{o,t}^c = c_t \left[R_{L,t} a_o^c + R_{E,t}\right]$. Similarly, we can identify a_o^h and $\frac{b_o^h}{a_o^h}$ if we observe some old high school workers who accumulated no experience input and earn $\tilde{w}_{o,t}^h = R_{L,t} a_o^h$ and compare them with old high school workers whose years of prior work endowed them with b_o^h units of experience input so that they earn $w_{o,t}^h = R_{L,t} a_o^h + R_{E,t} b_o^h$. Using the PSID data we are able to effectively do this from the imperfect correlation between age and the number of years that individuals actually worked by that age. While the two-age example here is extreme, we show in Appendix A3 that a small amount of variation in the number of years worked is sufficient to fully identify the model in the main text non-parametrically.

Because we do not impose the assumption that old workers exclusively supply the experience input, the age premiums across schooling groups are given by

$$\frac{w_{o,t}^c}{w_{y,t}^c} = a_o^c + \Pi_t \text{ and } \frac{w_{o,t}^h}{w_{y,t}^h} = a_o^h + \Pi_t b_o^h,$$

which implies the elasticity of the age premium with respect to Π_t differs across schooling groups depending on $\frac{1}{a_o^c} \lesseqgtr \frac{b_o^h}{a_o^h}$. Our estimates of these parameters in the main text allow us to account for the different movements of the age premiums across schooling groups and college premiums across age groups.

A2 PSID Data

Sample. We use the Panel Study of Income Dynamics (PSID) data from the U.S. for the 1968-2007 period. The PSID consists of two main subsamples: the SEO (Survey of Economic Opportunity) sample and the SRC (Survey Research Center) sample. We use both samples and restrict ourselves to the core members with positive sampling weights (not the newly added family members through marriage) to maintain the consistent representativeness of the sample over time.[1] The sample is restricted to individuals between 18 and 65 years of age.

Years of Prior Work. The procedure we use to construct measures of actual years worked since age 18 is as follows. Questions regarding the number of years worked ("How many years have you worked for money since you were 18?" and "How many of these years did you work full time for most or all of the year?") were asked of every household's head and wife in 1974, 1975, 1976 and 1985.[2] These questions are also asked for every person in the year when that person first becomes a household head or wife.[3] Since there are some inconsistencies between the answers, we first adjust the 1974 report to be consistent with 1975 and 1976 values when possible. Next, we use 1974 as the base year; i.e., we assume that whatever is recorded in 1974 for the existing heads is true. For the entrants into the sample we assume that the number of years of prior work they report in their first year in the sample is true. If the report implies that an individual started working before the age of 18, we redefine it to be the number of years since age 18 for that individual. If the reported number of years worked in 1974 is smaller than that implied by the reports of hours between the individual entry into the sample (or 1968) and 1974, we replace the 1974 report with that implied by the accumulated reports of hours. We

[1] We use only the nonimmigrant sample. In 1990 the PSID added a new sample of 2000 Latino households, consisting of families originally from Mexico, Puerto Rico, and Cuba. Because this sample missed immigrants from other countries, Asians in particular, and because of a lack of funding, this Latino sample was dropped after 1995. Another sample of 441 immigrant families was added in 1997. Because of the inconsistencies in these samples, we restrict ourselves to the core SEO and SRC samples throughout the 1968-2007 period.

[2] By default, the head of household is the (male) husband if he is present or a female if she is single. In very few cases the head is a female, even when the male husband is present (but is, say, severely disabled).

[3] The PSID mistakenly did not ask some people in 1985 and fixed this mistake by asking them in 1987.

repeat this procedure for 1985 and for the reports of the new heads and wives. Finally, using the values in 1974, 1985, and the reports of the new heads and wives, we increment the years of work variables forward and backward as follows: increment the full-time measure by one if individual works at least 1500 hours in a given year.[4] If we observe an individual in the sample since age 18, we ignore his or her reports and instead directly use his or her reports of hours in each year using the cutoff above.[5]

Other Variables. Our hourly wage measure is equal to the total earnings last year divided by total hours worked last year. To get the real wage, we adjust the nominal wage using last year's CPI (equal to 100 in 1984).[6] We define the economically active population as the group of people who worked at least 700 hours last year.[7] Education is measured by years of final educational attainment.[8] Other control variables that we will use are gender (male dummy), race (black dummy), and region (Northeast, North Central and West dummies). The broad region variable is created using the state variable in the PSID.[9] South is the base category region.

[4]We experimented with using cutoff values other than 1500 hours of work or using directly the sum of accumulated hours of work to create other measures of prior work and found that our chosen measure shows the smoothest pattern of movements. The substantive results are not sensitive to this choice.

[5]The PSID switched from annual to bi-annual interviewing after 1997. Some data for the non-interview years is available but appears very noisy with large numbers of missing observations. This led us to use only the data from years when interviews took place. The only exception is hours worked in years between interviews which are needed to construct the measures of prior work. We imputed those hours as the maximum between the reported hours (if available) and the average hours in the two adjacent survey years.

[6]There is an alternative hourly wage measure in the PSID which reports the current hourly wage at the time of the interview. Unfortunately, this measure is only available for the household heads throughout the period. For wives it is available only in 1976 and after 1979 and it is not available at all for the other family members.

[7]As in the case of earnings, there is also an employment status variable at the time of the interview. We do not use this variable because (1) the reference period (current year) is different from that of the earnings measure (last year), and (2) this variable is available for the heads for all years but not for the wives before 1979 except in 1976 and is not available for the dependents.

[8]Education is reported in the PSID in 1968, 1975, and 1985 for existing heads of households, and every year for the people becoming household heads or wives. It is kept constant between the years in which it is updated. As a result, there would be a bias toward a lower educational level. For example, if education is 10 years in 1975 and 16 in 1985, it would be reported 10 between 1975 and 1985. If the individual, however, had 16 years of education already in 1980, then for five years he would be counted as less educated than he actually is. To minimize this bias, the education variable used in the estimation is fixed to be equal to its mode value among all the reports available. To make the education variable comparable across time we top code it at 16 years.

[9]We found that the broad region variable provided by the PSID appears to be error-ridden. For example, for some but not all Texas residents region is defined as West. Thus, we reconstructed the broad region variable directly from the state of residence.

A3 Variation in Years Worked Needed for Identification

As is well known, the variation in the number of years worked by a certain age is relatively small, especially for male workers. This might appear to pose a challenge for our identification strategy. We now show that the model is nonparametrically identified if we only have one year of difference in the number of years worked at each age, e.g., it is enough that some workers enter the labor market at age 18 while some others at age 19 so that at age 20, they have worked 2 and 1 years, respectively, and at age 21, they have worked 3 and 2 years, respectively, etc. Following this, we show that there is much more variation available in our data.

A3.1 Non-parametric Identification

We now establish non-parametric identification of the relative price of experience Π_t and of the $\lambda_L(j)$, $\lambda_E(j)$ and $g(e)$ schedules if within each age group j some individuals worked for j years and some others for $j-1$ years, i.e. $e \in \{j, j-1\}$ for $j \geq 1$.

Specify the log wage of a worker with age j and years worked e as

$$\ln w(j,e) = \ln R_{L,t} + \ln(\lambda_L(j) + \Pi_t \lambda_E(j) g(e)).$$

We have the restriction that $\lambda_L(0) = 1$ and $g(0) = 0$, so that

$$\begin{aligned} \ln w(0,0) &= \ln R_{L,t}, \\ \ln w(1,0) &= \ln R_{L,t} + \ln \lambda_L(1), \end{aligned}$$

which are used to identify $\ln R_{L,t}$ for all t and $\ln \lambda_L(1)$. Now consider

$$\begin{aligned} \ln w(j, e=j) &= \ln R_{L,t} + \ln(\lambda_L(j) + \Pi_t \lambda_E(j) g(e=j)) \\ \ln w(j, e=j-1) &= \ln R_{L,t} + \ln(\lambda_L(j) + \Pi_t \lambda_E(j) g(e=j-1)) \end{aligned}$$

for all $j \geq 1$ and for every t. Since $\ln R_{L,t}$ is known, these equations imply that we can determine the following after exponentiating

$$\begin{aligned} &\lambda_L(j) + \Pi_t \lambda_E(j) g(e=j), \\ &\lambda_L(j) + \Pi_t \lambda_E(j) g(e=j-1), \end{aligned} \tag{A1}$$

for all $j \geq 1$ and for every t. Since $\ln \lambda_L(1)$ is also known, we can further determine

$$\Pi_t \lambda_E(1) g(1)$$

for every t, from which we can determine $\frac{\Pi_{t+1}}{\Pi_t}$ for every t.

Using (A1) and the time differences of these values, we can find

$$(\Pi_{t+1} - \Pi_t)\, \lambda_E(j) g\,(e = j)\,,$$
$$(\Pi_{t+1} - \Pi_t)\, \lambda_E(j+1) g\,(e = j)\,,$$
$$(\Pi_{t+1} - \Pi_t)\, \lambda_E(j+1) g\,(e = j+1)\,.$$

Using ratios of these, we can determine $\frac{\lambda_E(j+1)}{\lambda_E(j)}$, $\frac{g(e=j+1)}{g(e=j)}$ beginning from $\frac{\lambda_E(2)}{\lambda_E(1)}$, $\frac{g(2)}{g(1)}$.

Next, to determine $\lambda_L(j)$ we can use

$$\frac{\lambda_L(j) + \Pi_t \lambda_E(j) g\,(e = j)}{\Pi_t \lambda_E(1) g\,(1)} = \frac{\lambda_L(j)}{\Pi_t \lambda_E(1) g\,(1)} + \frac{\lambda_E(j) g\,(e = j)}{\lambda_E(1) g\,(1)},$$

where the second term on the right hand side is known given the calculated $\frac{\lambda_E(j+1)}{\lambda_E(j)}$, and $\frac{g(e=j+1)}{g(e=j)}$, and $\Pi_t \lambda_E(1) g\,(1)$ is also known, allowing us to identify $\lambda_L(j)$.

Finally, although from $\Pi_{t=0}\lambda_E(1) g\,(1)$ we cannot separately identify $\Pi_{t=0}$ and $\lambda_E(1)$ and $g\,(1)$, we could normalize two of these, $\lambda_E(1)$ and $g\,(1)$, without loss of generality to identify $\Pi_{t=0}$.[10] As discussed in Appendix A7, the level of $\Pi_{t=0}$ does not affect our substantive results and it only scales the level of the estimated share parameter $\ln \delta$ in the aggregate technology (not the estimate of the complementarity parameter). In the specification used in the main text this normalization is not needed because $\lambda_E(0) = 1$, $g\,(0) = 0$ and the restricted functional forms of $\lambda_E(j)$ and $g\,(e)$.

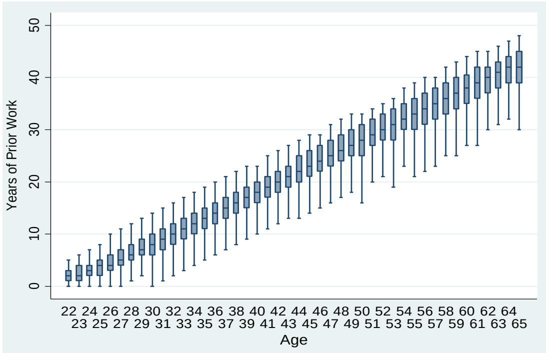
Note that given the nonparametric identification achieved, the parametric identification in the main text is guaranteed as a special case.

A3.2 Variation in Years Worked Available in the Data

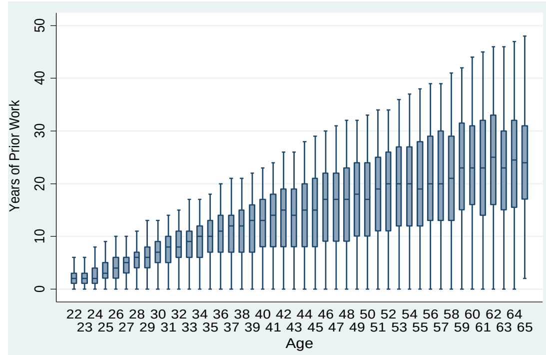
Figure A-1 uses boxplots to summarize the amount of variation in actual experience by age available in our data. Each plot shows the percentile statistics of the distribution such as median, 25th percentile, 75th percentile in a box and the upper and lower adjacent values in marking boundary values.[11] The figure illustrates that the range of variation of the number of years worked for every age group far exceeds the amount of variation needed for identification.

[10] This is the same normalization we used when setting $b_o^c = 1$ in Appendix A1.

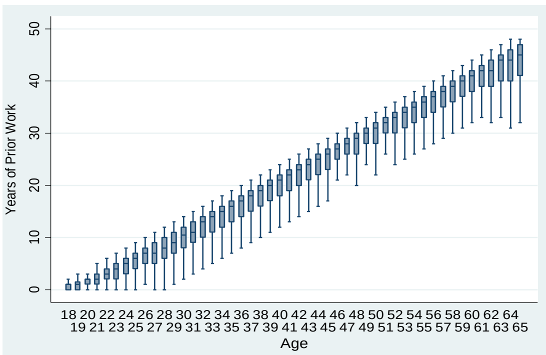
[11] The "upper and lower adjacent values" are the extreme values of ± 1.5 times of the inter-quartile range, which are suggested by Tukey (1977) to capture the "effective range" of the distribution.



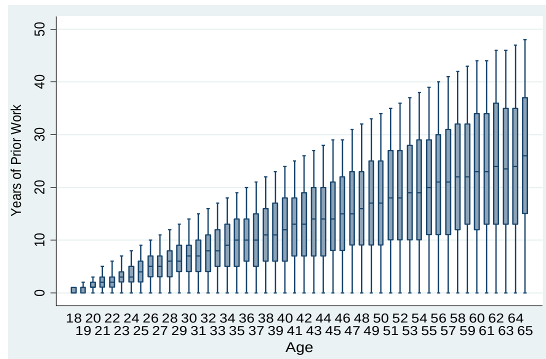
(a) College Male



(b) College Female



(c) High School Male



(d) High School Female

Figure A-1: Within-Age Variation in Years of Prior Work by Gender and Education.

Even among male workers, the effective range of within-age variation in the number of years worked is wider than 10 years for most age groups.

An alternative way to describe the amount of variation available for identification in the literature would be to to report the correlations between age and years of prior work. In our data these are 0.95 for college males, 0.95 for high-school males, 0.73 for college females, and 0.66 for high-school females. The interpretation of such correlations is, however, not straightforward in the context of establishing identification. This is because the overall correlation is dominated by the overall co-movement in age and years of prior work and does not immediately reveal the extent of years of prior work variation conditional on age, which determines the identification. The following simple example illustrates this.

Suppose that for each age the distribution of the number of years worked is constant around that age (the sample size N also ensures that this is the case). Let x_i denote age and y_i the years of prior work, then

$$y_i = x_i + e_i$$

where $e_i = e$ is from a given distribution independent of the level of x_i such that

$$\sum_i \left[e_i \left(x_i - \bar{x}\right)\right] = 0.$$

Without loss of generality, re-normalize the measure of years worked such that the average number of years worked and age are equal, $\bar{x} = \bar{y}$.

The aggregate correlation between age and years of prior work is given by

$$\begin{aligned}
r &= \frac{\sum_i \left[\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right]}{\sqrt{\sum_i \left(x_i - \bar{x}\right)^2 \sum_i \left(y_i - \bar{y}\right)^2}} \\
&= \frac{\sum_i \left[\left(x_i - \bar{x}\right)\left(x_i - \bar{x} + e_i\right)\right]}{\sqrt{\sum_i \left(x_i - \bar{x}\right)^2 \sum_i \left(x_i - \bar{x} + e_i\right)^2}} \\
&= \frac{\sum_i \left(x_i - \bar{x}\right)^2}{\sqrt{\sum_i \left(x_i - \bar{x}\right)^2 \left[\sum_i \left(x_i - \bar{x}\right)^2 + \sum_i \left(e_i\right)^2\right]}} \\
&= \sqrt{\frac{\sum_i \left(x_i - \bar{x}\right)^2}{\sum_i \left(x_i - \bar{x}\right)^2 + \sum_i \left(e_i\right)^2}}.
\end{aligned}$$

Thus, the correlation is falling in the ratio

$$\frac{\sum_i \left(e_i\right)^2}{\sum_i \left(x_i - \bar{x}\right)^2} = \frac{N\sigma_e^2}{\sum_i^N \left(x_i - \bar{x}\right)^2}.$$

Since the numerator is constant, this ratio is essentially falling in the range of ages in the sample, whereas the variation of age and years of prior work that is relevant for the identification is given by σ_e^2.

To provide a quantitative example, if we partition our sample into 10 equally spaced birth cohort bins, the average correlation within a bin is 0.9 for college males, 0.9 for high-school males, 0.58 for college females, and 0.57 for high-school females. To avoid arbitrariness of choosing such a partition, we think Figure A-1 is more informative in summarizing the ample variation available for identification in our data.

A4 Descriptive Analysis, Additional Results

A4.1 Descriptive Analysis, Benchmark Coefficient Estimates

Table A-1: Descriptive Analysis, Estimates of Time-Invariant Parameters.

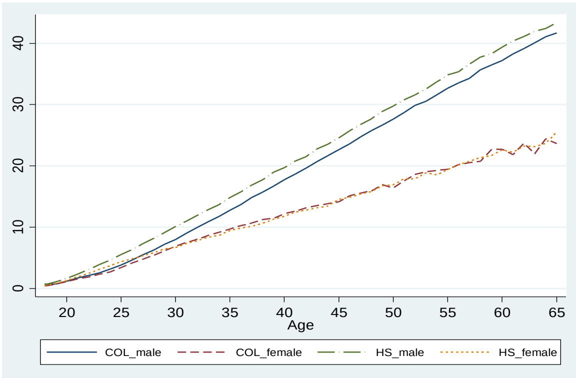
Parameter	Estimate	Standard error	t-statistic
$\lambda_{L,1}\,(HS, M)$	.0245	.00323	7.56
$\lambda_{L,2}\,(HS, M)$	-.000492	.0000803	-6.13
$\lambda_{L,0}\,(C, M)$	-.306	.0350	-8.74
$\lambda_{L,1}\,(C, M)$	.0671	.00363	18.47
$\lambda_{L,2}\,(C, M)$	-.00116	.0000846	-13.66
$\lambda_{E/L,1}\,(HS, M)$	-.0792	.00829	-9.55
$\lambda_{E/L,2}\,(HS, M)$	.00107	.000181	5.90
$\lambda_{E/L,0}\,(C, M)$	.332	.121	2.74
$\lambda_{E/L,1}\,(C, M)$	-0.148	.0132	-11.21
$\lambda_{E/L,2}\,(C, M)$	.00214	.000268	8.00
$\lambda_{L,1}\,(HS, F)$	.00109	.00203	.54
$\lambda_{L,2}\,(HS, F)$	.0000709	.0000463	1.53
$\lambda_{L,0}\,(C, F)$	-.0569	.0282	-2.02
$\lambda_{L,1}\,(C, F)$	.0345	.00265	12.98
$\lambda_{L,2}\,(C, F)$	-.000571	.0000612	-9.33
$\lambda_{E/L,1}\,(HS, F)$	-.0434	.00661	-6.57
$\lambda_{E/L,2}\,(HS, F)$	.0000404	.000137	.30
$\lambda_{E/L,0}\,(C, F)$	-.499	.127	-3.93
$\lambda_{E/L,1}\,(C, F)$	-.054	.0129	-4.23
$\lambda_{E/L,2}\,(C, F)$	.000142	.0000286	.50
θ_1	-.0234	.00476	-4.92
θ_2	.000979	.000184	5.32
θ_3	-.0000141	.00000238	-5.93
northeast	.19	.004	43.64
north central	.046	.004	11.46
west	.098	.0045	21.90
R^2	0.924		
$RMSE$	0.616		

Note - The entries represent the results of the reduced-form estimation of time-invariant parameters of the benchmark specification in Section II of the main text. For sample restrictions and variable construction procedures, see Appendix A2.

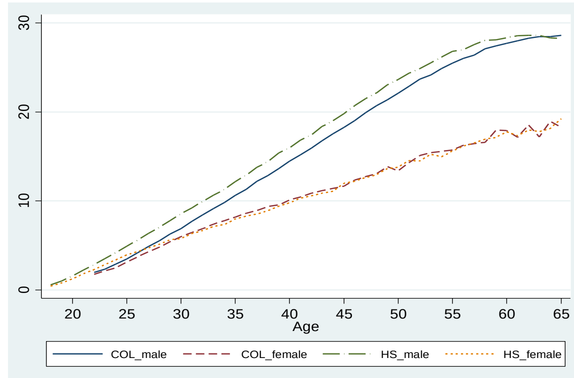
Table A-2: Descriptive Analysis, Estimates of Time-Varying Parameters.

Year	Price of Exp.	Schooling	Male	Black	Intercept
1968	.091(.015)	.072(.005)	.24(.031)	-.16(.048)	.46(.09)
1969	.078(.013)	.066(.005)	.25(.028)	-.16(.044)	.61(.08)
1970	.089(.013)	.063(.005)	.24(.027)	-.14(.042)	.63(.08)
1971	.087(.013)	.069(.005)	.22(.027)	-.10(.042)	.58(.08)
1972	.106(.014)	.074(.005)	.20(.026)	-.12(.041)	.49(.08)
1973	.104(.013)	.072(.004)	.25(.026)	-.086(.040)	.52(.07)
1974	.101(.012)	.064(.004)	.23(.023)	-.099(.036)	.65(.07)
1975	.109(.012)	.067(.004)	.21(.023)	-.087(.035)	.59(.07)
1976	.126(.013)	.069(.004)	.16(.023)	-.051(.035)	.54(.07)
1977	.131(.014)	.076(.004)	.19(.023)	-.037(.035)	.42(.07)
1978	.152(.015)	.075(.004)	.19(.023)	-.051(.035)	.40(.07)
1979	.141(.014)	.070(.004)	.20(.022)	-.087(.032)	.49(.07)
1980	.137(.013)	.066(.004)	.20(.021)	-.045(.031)	.53(.07)
1981	.170(.016)	.080(.004)	.17(.022)	-.092(.032)	.26(.07)
1982	.155(.015)	.071(.004)	.17(.022)	-.093(.032)	.38(.07)
1983	.206(.019)	.083(.004)	.10(.021)	-.075(.032)	.12(.07)
1984	.183(.016)	.081(.004)	.11(.020)	-.082(.030)	.20(.07)
1985	.217(.018)	.092(.004)	.15(.020)	-.078(.030)	-.03(.07)
1986	.207(.018)	.092(.004)	.12(.020)	-.12(.030)	.02(.07)
1987	.224(.015)	.099(.004)	.10(.020)	-.15(.029)	-.07(.07)
1988	.251(.021)	.108(.004)	.05(.020)	-.14(.029)	-.22(.07)
1989	.242(.019)	.111(.004)	.07(.018)	-.15(.026)	-.26(.07)
1990	.231(.019)	.114(.004)	.04(.018)	-.11(.026)	-.29(.07)
1991	.217(.018)	.123(.004)	.04(.018)	-.07(.027)	-.41(.07)
1992	.252(.020)	.122(.004)	.03(.018)	-.13(.026)	-.47(.07)
1993	.220(.018)	.111(.004)	.04(.018)	-.094(.026)	-.24(.07)
1994	.201(.017)	.119(.004)	.09(.018)	-.16(.026)	-.36(.07)
1995	.229(.019)	.110(.004)	.08(.018)	-.11(.026)	-.27(.07)
1996	.238(.020)	.106(.004)	.08(.018)	-.14(.026)	-.23(.08)
1997	.214(.018)	.115(.004)	.03(.017)	-.15(.025)	-.31(.07)
1999	.217(.018)	.113(.004)	.07(.017)	-.15(.025)	-.26(.07)
2001	.174(.016)	.117(.004)	.07(.017)	-.088(.024)	-.19(.07)
2003	.165(.015)	.106(.004)	.02(.016)	-.11(.024)	-.05(.07)
2005	.185(.016)	.120(.004)	.04(.016)	-.15(.024)	-.26(.07)
2007	.143(.014)	.118(.004)	.07(.016)	-.17(.023)	-.12(.07)

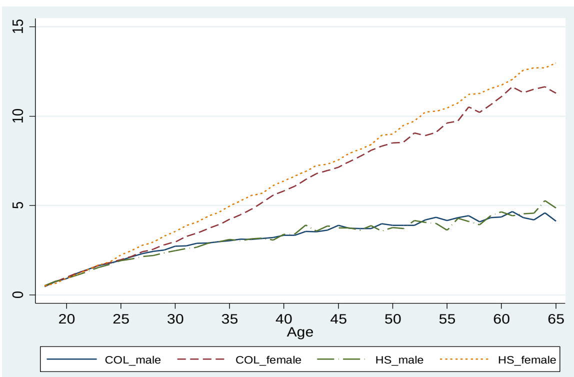
Note - The entries represent the results of the reduced-form estimation of time-varying parameters of the benchmark specification in Section II of the main text. Standard errors are in parenthesis. For sample restrictions and variable construction procedures, see Appendix A2.



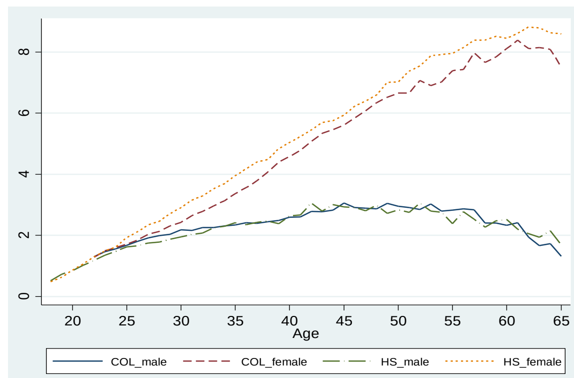
(a) Years of Prior Work, Mean



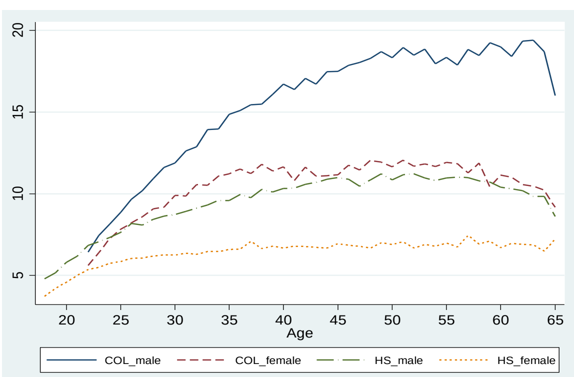
(b) Experience Input, $g(e)$, Mean



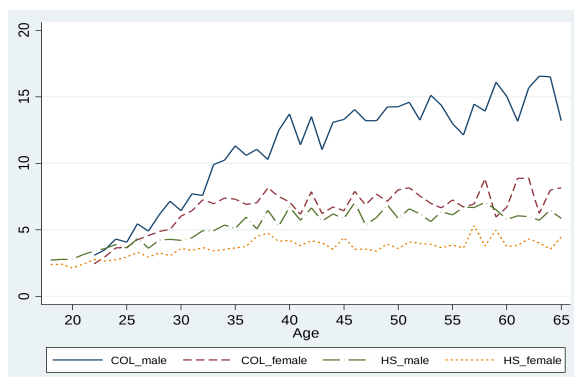
(c) Years of Prior Work, SD



(d) Experience Input, $g(e)$, SD

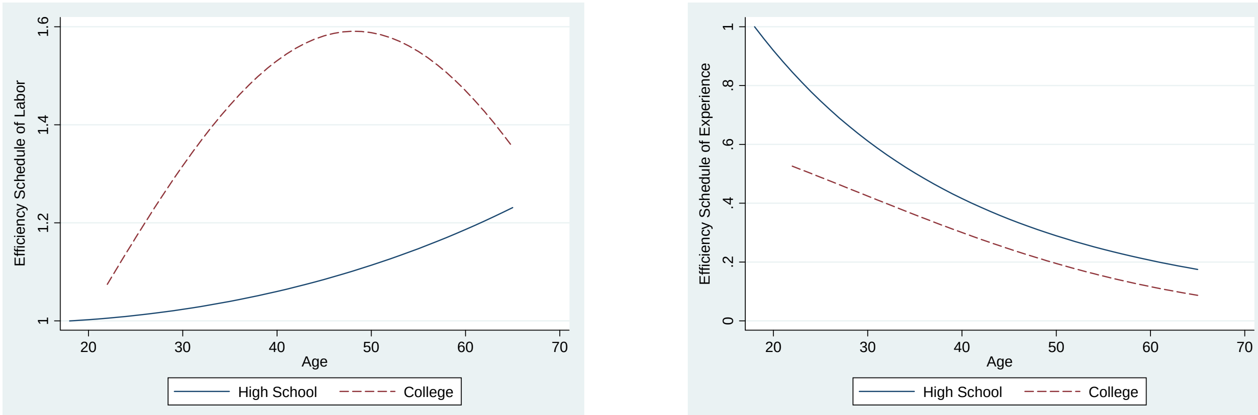


(e) Wage, Mean



(f) Wage, SD

Figure A-2: Age-Conditional Means and Standard Deviations of Years Worked, Experience Input, and Wages by Gender and Education.



(a) Age Efficiency Schedule of Labor.

(b) Age Efficiency Schedule of Experience.

Figure A-3: Estimated Age-Efficiency Schedules for Female Workers by Education.

A5 Time-Invariant Age Efficiency Schedules and Cohort Effects

While our identification strategy follows Katz and Murphy (1992) and subsequent literature in assuming that age efficiency schedules λ_L and λ_E are independent of time, this is potentially an important restriction ruling out certain cohort effects. In this Appendix we empirically assess this assumption through two experiments. First, we check for the presence of cohort effects not accounted for by the model with time-invariant age efficiency schedules. Second, we estimate the model separately on different cohorts and check whether the estimates differ significantly.

A5.1 Cohort Effects

To check for the presence of cohort effects that are not accounted for by our specification with constant age efficiency schedules, we we obtain wage residuals from our model and ask whether we can detect the presence of residual cohort effects in them. In particular, we regress the residuals on the full set of cohort dummies. The estimates of coefficients on those dummies are all statistically insignificant from zero. Moreover, they do not exhibit any particular trends as is evident from Figure A-4 in which the coefficients of the cohort dummies are plotted for all cohorts.

How does our model account for cohort effects? The time effect on wages is captured via the changes in the year dummies of the wage regression (equation (16) in the main text). The pure age effect is captured by the age efficiency schedule of labor $\lambda_L(j)$. The cohort effect is captured

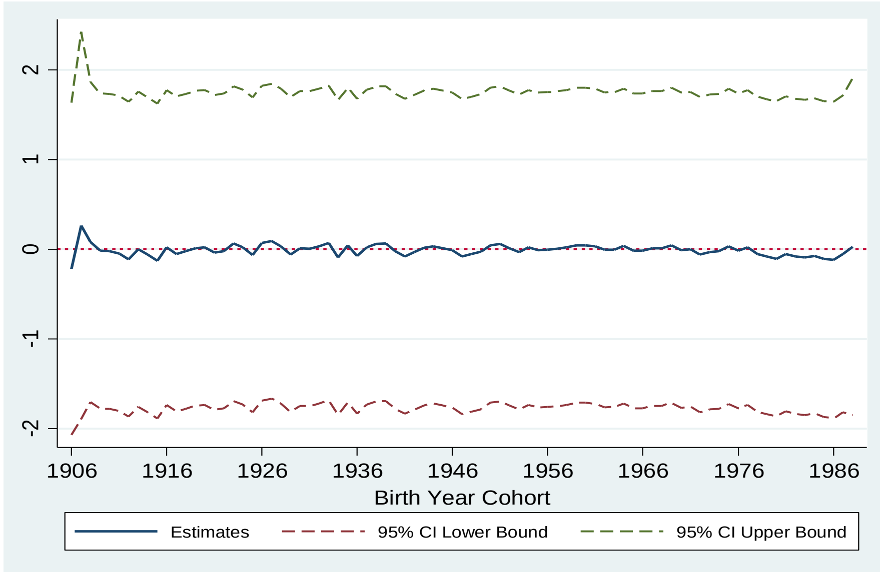


Figure A-4: Residual Cohort Dummy Estimates.

by the interactive term between the relative price of experience $\Pi_{E,t}$ (time effect) and the age efficiency schedule of labor $\lambda_{E/L}(j)$ (age effect).

Moreover, we also allow the time-varying coefficients for the characteristics of schooling, gender and race. Through the changing age composition of these demographic subgroups, allowing the time-varying coefficients on these characteristics also captures the cohort effect indirectly. These are clearly one particular way of capturing cohort effects and the question is whether we are missing other significant cohort effects than the ones captured. The results of the experiment in this section suggest that we are not.

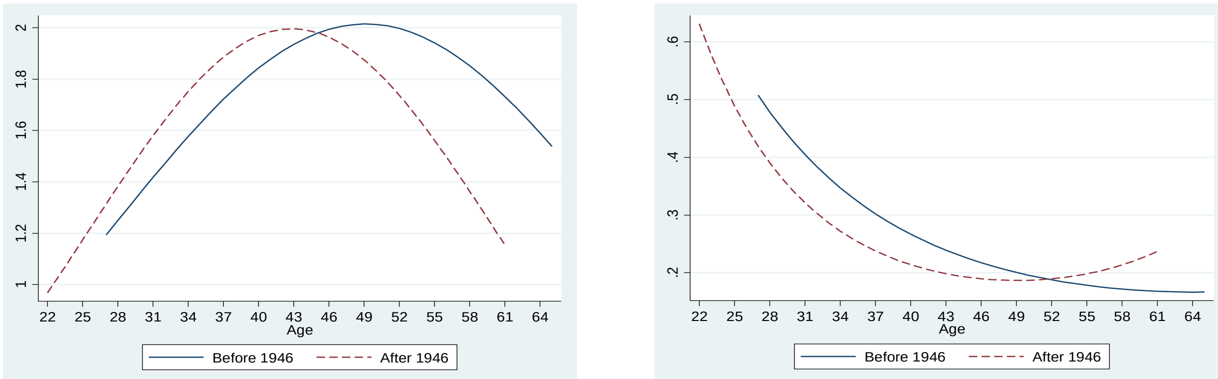
A5.2 Estimating the Model on Different Cohorts

We now allow the age efficiency schedules to differ across cohorts and check whether the estimates differ significantly. In particular, we allow the age efficiency schedules for males to be cohort specific as follows

$$\lambda_L^s(j,Z) = \exp(\lambda_{L,1}^s j + \lambda_{L,2}^s j^2 + I_Z\left(Z_{L,1}^s j + Z_{L,2}^s j^2\right)),$$
$$\lambda_E^s(j,Z) = \exp(\lambda_{E,1}^s j + \lambda_{E,2}^s j^2 + I_Z\left(Z_{E,1}^s j + Z_{E,2}^s j^2\right)),$$

where I_Z is an indicator for cohort Z.

Defining cohorts too finely runs into the problem of cohorts not having observations along the support of the age profile for the youngest and oldest cohorts, over and above sample size



(a) λ_L

(b) λ_E

Figure A-5: Estimated $\lambda_L(j), \lambda_E(j)$ for Cohorts of College-Educated Male Workers Born before or after 1946.

issues. Thus, we consider two cohort groups, those born before and those born after 1946. We choose the partition of cohorts at 1946 to generate the most overlap in terms of age between the two cohorts for comparison. This turns out to also imply similar sample sizes for the two groups.[12]

For the high school group, the four coefficients $\{Z^s_{L,1}, Z^s_{L,2}, Z^s_{E,1}, Z^s_{E,2}\}$ are all insignificant so there are no statistically relevant differences in λ^s_L and λ^s_E across cohorts. For the college group each of these four coefficients turned out to be statistically significant, and we investigated further the implications.

Plotting $\lambda^s_L(j, Z)$ and $\lambda^s_E(j, Z)$ over the support of age j for the pre-1946, post-1946 cohorts, in Figure A-5, we observe that the age efficiency schedules are quite similar in shape and position between the two cohort groups.

Moreover, allowing for cohort-specific $\lambda^s_L(j, Z)$, $\lambda^s_E(j, Z)$ does not affect the ability of the model to match the dynamics of the relative price of experience, the age premiums and college premiums which are the focus of our analysis. The results are virtually indistinguishable from the benchmark ones, leading us to prefer the more parsimonious benchmark specification.

We can statistically assess the similarity of predictions of the models with common or cohort-

[12]We also considered an alternative specification where we allowed the cohort to determine the level of efficiency units such that

$$\lambda^s_L(j, Z) = \exp(\lambda^s_{L,1} j + \lambda^s_{L,2} j^2 + I_Z Z^s_L).$$

In this simpler specification, we could consider differences in coefficient Z^s_L between finely defined cohorts since we do not run into the issue of not having observations along the support mentioned above. The coefficients for Z^s_L for cohorts differentiated by calender birth year (relative to the base cohort) all turn out to be insignificant.

specific age efficiency schedules by performing the nonparametric Kolmogorov-Smirnov test for the distributional equality between the two predicted wage distributions. Specifically, let $F_n^0(w)$ and $F_n^1(w)$ be the empirical distribution functions of the fitted log wage with sample size n from the specification of common age efficiency schedules and from that of cohort-differentiated age efficiency schedules, respectively. The Kolmogorov-Smirnov statistic for testing the equality between the two distributions is

$$K_n = \sup_w \left| F_n^0(w) - F_n^1(w) \right|,$$

where $\sup_w |\cdot|$ indicates the supremum and the empirical distribution function is defined as

$$F_n^j(w) = \frac{1}{n} \sum_{i=1}^{n} I\left(W^j \leq w\right).$$

The statistic $\sqrt{n}K_n$ converges to Kolmogorov distribution under the null of equality, which does not depend on the form of the true distribution of the log wage.

The Kolmogorov-Smirnov test statistic for the whole sample is 0.0026 with p-value of 0.639, hence we cannot reject the null hypothesis of equality between the two distributions for the whole sample. We also check the equality of the log wage distributions for the two cohorts. The test statistic for the cohort born before 1946 sample is 0.0048 with p-value of 0.439. The test statistic for the cohort born after 1946 is 0.0049 with p-value of 0.189. Thus, the equality of cohort-specific conditional distributions is not rejected either.

A6 Assessing Alternative Specifications

In this section, we compare the ability of our specification of the wage equation to fit the data relative to various alternative specifications. Two types of restrictions are of particular interest. First, our benchmark specification incorporates three potential sources of curvature in the life-cycle profiles ($g(e)$, λ_L, and λ_E). We assess whether all three components are essential for fitting the data or only a subset of them would be statistically sufficient. Second, our benchmark specification filters out the time-varying college premium, gender premium, etc. when assessing the relative price of experience, and incorporates these time varying premiums when calculating the aggregate effective supplies of experience and labor. The traditional cell-based correction for composition does not accommodate these features as we discuss in Footnote 2 in the main text. Consequently, it is of some interest to assess the consequences of this restriction for fitting the wage distribution. To do so we consider alternative specifications where the coefficients on these characteristics are forced to be time-invariant.

We measure the distance between our benchmark specification and each alternative one by the difference in estimated log likelihoods between them. Vuong (1989) shows that under regularity conditions, the likelihood ratio test statistic converges to a central chi-square distribution.[13] Specifically, suppose there are two competing models to explain the variable Y conditional on Z that are represented by the conditional distribution functions $\mathbf{F}_\theta \equiv \left\{F_{Y|Z}\left(\cdot|\cdot;\theta\right);\theta\in\Theta\subset R^p\right\}$ and $\mathbf{G}_\gamma \equiv \left\{G_{Y|Z}\left(\cdot|\cdot;\gamma\right);\gamma\in\Gamma\subset R^q\right\}$, respectively, and their density functions are denoted by $f\left(y|z;\theta\right)$ and $g\left(y|z;\gamma\right)$. In our case, $y=\ln w$ and $z=(1,s,x,j,e,\chi)$. Let $\widehat{\theta}_n$ and $\widehat{\gamma}_n$ be the corresponding maximum likelihood estimators for the sample $(y_\iota,z_\iota)_{\iota=1}^n$ of size n, i.e., $\widehat{\theta}_n = \arg\max_{\theta\in\Theta}\sum_{\iota=1}^n \log f\left(y_\iota|z_\iota;\theta\right)$ and $\widehat{\gamma}_n=\arg\max_{\gamma\in\Gamma}\sum_{\iota=1}^n\log g\left(y_\iota|z_\iota;\gamma\right)$. Then, under the regularity conditions, $2LR_n\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)\xrightarrow{D}\chi^2_{p-q}$, where

$$2LR_n\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)=2\sum_{\iota=1}^n\log\left[\frac{f\left(y_\iota|z_\iota;\widehat{\theta}_n\right)}{g\left(y_\iota|z_\iota;\widehat{\gamma}_n\right)}\right], \tag{A2}$$

and $p-q$ is the difference in the total number of parameters between the two models.

The alternative specifications that we consider in this section are nested by the benchmark specification. Hence the alternative hypothesis to the null hypothesis of the equivalence of the

[13] See assumptions A1-A5 and information matrix equivalence condition in equation (3.8) for regularity conditions and Theorem 3.3 and Corollary 3.4 for the characterization of the asymptotic distribution of the likelihood ratio test statistic in Vuong (1989).

Table A-3: Likelihood Ratio Test Statistics.

Specification #	Description	$\mathcal{LR}$	$\chi^2_{p-q}\,(0.01)$
Spec 1	$g(e) = e$; $\lambda_L = \lambda_E = 1$; $\alpha_t = \alpha$	6943.8	181.8
Spec 2	$g(e) = e$; $\lambda_L = \lambda_E = 1$; benchmark α_t	5661.5	41.6
Spec 3	benchmark $g(e)$; $\lambda_L = \lambda_E = 1$; $\alpha_t = \alpha$	3634.5	178.4
Spec 4	$g(e) = e$; benchmark λ_L and λ_E; $\alpha_t = \alpha$	1110.8	159.0
Spec 5	benchmark $g(e)$; $\lambda_L = \lambda_E = 1$; benchmark α_t	2371.0	37.6
Spec 6	benchmark $g(e)$; benchmark λ_L and λ_E; $\alpha_t = \alpha$	1033.4	155.5
Spec 7	benchmark $g(e)$; $\lambda_L = 1$; benchmark λ_E; benchmark α_t	863.9	23.2
Spec 8	benchmark $g(e)$; benchmark λ_L; $\lambda_E = 1$; benchmark α_t	765.0	23.2
Spec 9	benchmark $g(e)$; symmetric $\lambda_L = \lambda_E$; benchmark α_t	758.9	23.2
Spec 10	$g(e) = e$; benchmark λ_L and λ_E; benchmark α_t	32.4	11.3

compared models is that the benchmark model is *strictly superior* to the other candidate model in fitting the wage distribution. Thus, the statistic in (A2) allows us to perform a statistical significance test for the superiority of our benchmark specification over the alternatives.

Table A-3 provides the likelihood ratio test statistic (denoted by $\mathcal{LR}$) comparing our benchmark specification with various alternatives, along with the corresponding critical values of the chi-square distributions for the 1% significance level (denoted by $\chi^2_{p-q}\,(0.01)$).

The likelihood ratio test statistics are far larger than the 1% significance critical values for all of the alternative specifications. In fact, the test statistics also exceed the 0.1% significance critical values. Our benchmark specification fits the wage distribution strictly better than the other candidate specifications at any conventional significance level. Thus, the full incorporation of age efficiency schedules for both experience and labor, the curvature of experience input, and allowing for the time-varying coefficients for control variables provides critical improvements in fitting the wage distribution.

Furthermore, the likelihood ratio test statistic can be considered as the distance of each alternative specification from the benchmark. By comparing the magnitudes of the likelihood ratio statistics across specifications, we can infer the relative importance of each ingredient of the model specification. For example, the likelihood ratio falls from 6943.8 (Spec 1) to 5661.5 (Spec 2) by allowing for time-varying coefficients on the control variables, but falls to 3634.5 (Spec

3) after relaxing the linearity of $g(e)$ function. Thus, allowing for curvature in the experience accumulation technology seems more important than allowing for the time-varying coefficients for the control variables. The most substantial improvements come from introducing the age efficiency schedules. The likelihood ratio falls from 6943.8 (Spec 1) to 1110.8 (Spec 4), after incorporating the age efficiency schedules into the model. With curvature in the experience accumulation technology and time-varying coefficients on the control variables, the likelihood ratio increases to 2371.0 (Spec 5) from 1110.8 (Spec 4) when the efficiency of labor and experience is not allowed to depend on age.

Further evidence of the importance of the full consideration of the age efficiency schedules comes from the comparison of likelihood ratios among specifications 7, 8, 9, and 10. After allowing for the age efficiency schedules and time-varying coefficients for the control variables, the likelihood ratio falls to 32.4 (Spec 10), even when restricting the experience accumulation technology to be linear. However, even with full curvature of experience accumulation technology and time-varying coefficients on the control variables, an incomplete inclusion of the age efficiency schedules makes the model fit much worse: the likelihood ratio becomes 863.9 (Spec 7 for $\lambda_L = 1$), 765.0 (Spec 8 for $\lambda_E = 1$), and 758.9 (Spec 9 for $\lambda_L = \lambda_E$).

Spec 6 shows the likelihood ratio is high (1033.4) when we do not allow for time varying premiums to college, gender etc. $\alpha_t = \alpha$. Comparing Spec 1 with Spec 2, Spec 3 with Spec 5 and Spec 4 with Spec 10, we further confirm that setting these premiums as constant under other specifications for efficiency schedules substantially raises the likelihood ratio.

This evidence implies that all three potential sources of curvature of life-cycle profiles as well as the time variation in the coefficients on the control variables are essential for fitting the wage data. We emphasize once again, however, that while this evidence guides us in specifying the model of individual earnings, it is independent of the relationship between the aggregate relative supply of experience and its relative price.

A7 Identification of the Aggregate Production Function Parameters in the Structural Model

The log wage equation (20) in the main text includes all parameters of the model. In particular, given the measurement of aggregate inputs E_t and L_t, the variation of the relative price of experience Π_{E_t} in relation to the variation of the experience-labor ratio $\frac{E_t}{L_t}$ is the source of identification of the technology parameters μ and δ. The time-series correlation between the relative price Π_{E_t} and the relative factor endowment $\frac{E_t}{L_t}$ identifies μ (which is scale free). The average magnitude of Π_{E_t} relative to the magnitude of the $\frac{E_t}{L_t}$ identifies the scale parameter δ.

Note that the magnitudes of Π_{E_t} and $\frac{E_t}{L_t}$ depend on the normalization of some parameter of the age efficiency schedules, i.e., $\lambda_E(0, HS, x) = \lambda_L(0, HS, x) = 1$ for $x \in \{M, F\}$. Thus, the identification of δ is subject to this normalization. More precisely, it is the normalization of the *relative* efficiency of experience of the youngest workers that affects the identification of δ. That is, re-normalizing $\lambda_E(0, s, x) = \lambda_L(0, s, x) = l$ for any arbitrary constant l such that $\lambda_{E/L}(0, s, x) = 1$ leaves the estimate of δ unchanged. However, if we normalize the age efficiency schedules asymmetrically between experience and labor so that $\lambda_L(0, s, x) = a$ and $\lambda_E(0, s, x) = b$, hence $\lambda_{E/L}(0, s, x) = c = b/a \neq 1$, the coefficient function in front of experience in the log wage becomes $\delta \left(c\frac{E_t}{L_t}\right)^{\mu-1} c\lambda_{E/L}(j, s, x) = \widetilde{\delta} \left(\frac{E_t}{L_t}\right)^{\mu-1} \lambda_{E/L}(j, s, x)$, where $\widetilde{\delta} = \delta c^{\mu}$. Thus, the estimated value of δ may change. The normalization of the age efficiency schedule of *labor* affects the scale of the aggregate productivity term. Specifically, with $\lambda_L(0, s, x) = a$, the aggregate productivity term turns to $\ln aA_t$. Note, however, that estimates of μ as well as the age efficiency schedules, our key parameters, are *not* affected by this normalization.

A8 Structural Estimation, Additional Results

A8.1 Structural Estimation, Benchmark Coefficient Estimates

Table A-4: Structural Estimation, Estimates of Time-Invariant Parameters.

Parameter	Estimate	Standard error	t-statistic
$\lambda_{L,1}\,(HS, M)$	.0239	0.00335	7.13
$\lambda_{L,2}\,(HS, M)$	-.000502	0.0000846	-5.93
$\lambda_{L,0}\,(C, M)$	-.308	0.0353	-8.73
$\lambda_{L,1}\,(C, M)$	.0679	0.00366	18.52
$\lambda_{L,2}\,(C, M)$	-.00118	0.0000857	-13.82
$\lambda_{E/L,1}\,(HS, M)$	-.0773	0.00832	-9.29
$\lambda_{E/L,2}\,(HS, M)$	.00107	0.000183	5.88
$\lambda_{E/L,0}\,(C, M)$	.321	0.122	2.62
$\lambda_{E/L,1}\,(C, M)$	-0.148	0.0132	-11.19
$\lambda_{E/L,2}\,(C, M)$	0.00219	0.000266	8.25
$\lambda_{L,1}\,(HS, F)$	.000879	.00205	0.43
$\lambda_{L,2}\,(HS, F)$	.0000755	.0000467	1.62
$\lambda_{L,0}\,(C, F)$	-.0544	.0284	-1.92
$\lambda_{L,1}\,(C, F)$	.0344	.00269	12.81
$\lambda_{L,2}\,(C, F)$	-.000574	.0000622	-9.23
$\lambda_{E/L,1}\,(HS, F)$	-.0426	.00661	-6.45
$\lambda_{E/L,2}\,(HS, F)$	.0000302	.000137	.22
$\lambda_{E/L,0}\,(C, F)$	-.510	.127	-3.99
$\lambda_{E/L,1}\,(C, F)$	-.054	.0129	-4.17
$\lambda_{E/L,2}\,(C, F)$	.000158	.0000287	.55
θ_1	-0.0244	0.00462	-5.30
θ_2	.00101	0.000179	5.66
θ_3	-.0000145	0.00000231	-6.29
northeast	0.19	0.004	43.66
north central	0.046	0.004	11.49
west	0.098	0.0045	21.94
R^2	0.924		
$RMSE$	0.616		

Note - The entries represent the results of the structural estimation of time-invariant parameters of the benchmark specification. For sample restrictions and variable construction procedures, see Appendix A2. See Section III in the main text for details of the estimation procedure.

Table A-5: Structural Estimation, Estimates of Time-Varying Parameters.

Year	Schooling	Male	Black	Intercept
1968	.071(.005)	.25(.033)	-.16(.048)	.49(.066)
1969	.067(.004)	.24(.031)	-.16(.044)	.58(.061)
1970	.062(.004)	.24(.030)	-.14(.042)	.65(.060)
1971	.069(.004)	.22(.029)	-.10(.042)	.52(.059)
1972	.072(.004)	.21(.029)	-.12(.041)	.54(.059)
1973	.071(.004)	.25(.028)	-.087(.040)	.54(.057)
1974	.065(.004)	.23(.026)	-.100(.036)	.63(.054)
1975	.068(.004)	.21(.026)	-.086(.035)	.56(.054)
1976	.069(.004)	.16(.026)	-.051(.035)	.53(.055)
1977	.077(.004)	.19(.026)	-.037(.035)	.41(.055)
1978	.073(.004)	.19(.026)	-.052(.035)	.44(.055)
1979	.071(.004)	.20(.025)	-.086(.032)	.47(.053)
1980	.068(.004)	.19(.024)	-.044(.031)	.47(.054)
1981	.079(.004)	.17(.024)	-.092(.032)	.28(.054)
1982	.073(.004)	.16(.024)	-.093(.032)	.34(.055)
1983	.079(.004)	.12(.025)	-.075(.032)	.21(.057)
1984	.082(.004)	.11(.024)	-.082(.030)	.19(.055)
1985	.090(.004)	.16(.024)	-.078(.030)	.01(.055)
1986	.091(.004)	.13(.024)	-.12(.030)	.03(.057)
1987	.097(.004)	.11(.024)	-.15(.029)	-.03(.057)
1988	.105(.004)	.06(.024)	-.14(.029)	-.14(.059)
1989	.110(.004)	.07(.023)	-.15(.026)	-.23(.057)
1990	.114(.004)	.05(.023)	-.11(.026)	-.29(.058)
1991	.126(.004)	.04(.023)	-.071(.027)	-.47(.059)
1992	.120(.004)	.04(.023)	-.13(.026)	-.41(.059)
1993	.113(.004)	.04(.023)	-.094(.026)	-.30(.059)
1994	.125(.004)	.08(.023)	-.16(.026)	-.49(.059)
1995	.109(.004)	.08(.023)	-.12(.026)	-.24(.059)
1996	.102(.004)	.09(.023)	-.14(.026)	-.15(.059)
1997	.116(.004)	.03(.022)	-.15(.025)	-.34(.052)
1999	.110(.004)	.08(.022)	-.15(.025)	-.19(.052)
2001	.119(.004)	.07(.022)	-.087(.024)	-.24(.053)
2003	.109(.004)	.02(.022)	-.11(.024)	-.10(.052)
2005	.118(.004)	.05(.022)	-.15(.024)	-.20(.052)
2007	.120(.004)	.06(.022)	-.17(.023)	-.18(.052)

Note - The entries represent the results of the structural estimation of time-varying parameters of the benchmark specification. Standard errors are in parenthesis. For sample restrictions and variable construction procedures, see Appendix A2. See Section III in the main text for details of the estimation procedure.

A9 The Effect of Cohort Size on Earnings

Given the aggregate technology (3) in the main text, note that from the Euler theorem

$$\begin{aligned} G_{EE} &= -\frac{L_t}{E_t} G_{EL}, \\ G_{LL} &= -\frac{E_t}{L_t} G_{EL}. \end{aligned}$$

The aggregate stocks of labor and experience in period t can be constructed as the sum of effective supplies across cohorts indexed by age j

$$\begin{aligned} L_t &= \sum_j \lambda_L(j) N_{j,t}, \\ E_t &= \sum_j \lambda_E(j) g_t(j) N_{j,t}, \end{aligned}$$

where $N_{j,t}$ denotes the cohort size and $\lambda_L(j)$, $\lambda_E(j)$, $g(j)$ denote the efficiency schedules for a representative worker in cohort j. We suppress notation for sex and schooling and omitted productive characteristics z_{jt} and hours h_{jt} for clarity. The complementarity between two cohorts j and k is given by the condition $\frac{d^2 Y_t}{dN_{j,t} dN_{k,t}} > 0$. This cross derivative is given by

$$\begin{aligned} \frac{d^2 Y_t}{dN_{j,t} dN_{k,t}} &= A_t \left[\begin{array}{c} G_{EE}\lambda_E(j) g_t(j) \lambda_E(k) g_t(k) + G_{EL}\lambda_E(j) g_t(j) \lambda_L(k) \\ +G_{LE}\lambda_L(j)\lambda_E(k) g_t(k) + G_{LL}\lambda_L(j)\lambda_L(k) \end{array} \right] \\ &= A_t G_{EL} \frac{L_t}{E_t} \left[\begin{array}{c} -\lambda_E(j) g_t(j) \lambda_E(k) g_t(k) + \lambda_E(j) g_t(j) \lambda_L(k) \frac{E_t}{L_t} \\ +\lambda_L(j) \lambda_E(k) g_t(k) \frac{E_t}{L_t} - \left(\frac{E_t}{L_t}\right)^2 \lambda_L(j)\lambda_L(k) \end{array} \right] \\ &= A_t G_{EL} \frac{L_t}{E_t} \lambda_L(j)\lambda_L(k) \left[\begin{array}{c} -\frac{\lambda_E(j) g_t(j) \lambda_E(k) g_t(k)}{\lambda_L(j)\lambda_L(k)} + \frac{\lambda_E(j) g_t(j)}{\lambda_L(j)} \frac{E_t}{L_t} \\ +\frac{\lambda_E(k) g_t(k)}{\lambda_L(k)} \frac{E_t}{L_t} - \left(\frac{E_t}{L_t}\right)^2 \end{array} \right] \\ &= A_t G_{EL} \frac{L_t}{E_t} \lambda_L(j)\lambda_L(k) \left[\frac{E_t}{L_t} - \frac{\lambda_E(j) g_t(j)}{\lambda_L(j)} \right] \left[\frac{\lambda_E(k) g_t(k)}{\lambda_L(k)} - \frac{E_t}{L_t} \right], \end{aligned}$$

using the implications of the Euler theorem above.

Since aggregate experience-labor complementarity implies $G_{EL} > 0$, cohorts are complements when the cohort specific experience-labor ratios $\frac{\lambda_E(j) g_t(j)}{\lambda_L(j)}$ and $\frac{\lambda_E(k) g_t(k)}{\lambda_L(k)}$, are respectively lower and higher than the aggregate experience-labor ratio $\frac{E_t}{L_t}$. This is because cohorts complement each other through the effect on the aggregate experience-labor ratio. When both cohort specific experience-labor ratios are either lower or higher than the aggregate ratio, they are substitutes since $\frac{d^2 Y_t}{dN_{j,t} dN_{k,t}} < 0$.

Complementarity or substitutability is stronger the larger is $\frac{\lambda_E(j)g_t(j)}{\lambda_L(j)} - \frac{E_t}{L_t}$, i.e. the more distant the cohort specific experience-labor ratios are from the aggregate ratio. For a cohort where the experience-labor ratio coincides with the aggregate ratio, i.e. $\frac{\lambda_E(j)g_t(j)}{\lambda_L(j)} = \frac{E_t}{L_t}$, the marginal product is not affected by changes in the population of other cohorts (at the margin). For the same reason, the effect of own cohort size $\frac{d^2Y_t}{d^2N_{j,t}}$ on reducing the marginal product is rising in the absolute distance of the cohort specific experience-labor ratio from the aggregate ratio, i.e. $\frac{\lambda_E(j)g_t(j)}{\lambda_L(j)} - \frac{E_t}{L_t}$.

References

Card, D., and T. Lemieux (2001): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics*, 116(2), 705–746.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963-87: Supply and Demand Factors," *Quarterly Journal of Economics*, 107(1), 35–78.

Tukey, J. W. (1977): *Exploratory Data Analysis.* Addison-Wesley, Reading, MA.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), 307–333.