Demographic Change and the Return to Experience*

Hyeok Jeong
Vanderbilt University

Yong Kim
University of Southern California

Iourii Manovskii
University of Pennsylvania

Abstract

We propose and estimate a model in which changes in the demographic composition of the labor force may affect the returns to labor market experience. We consider workers as providing two distinct productive services – physical effort, or "labor," and services of the skill accumulated with labor market experience, or "experience." The key element in the model is the aggregate production function that allows for complementarity between the appropriately measured aggregate stocks of labor and experience. The parameters of the aggregate technology are identified by estimating individual earnings equations that consistently aggregate. Both time-series and cross-sectional data confirm strong experience-labor complementarity. We find that the observed demographic changes that drive the aggregate experience to labor ratio account nearly perfectly for the substantial changes in the experience premium over time.

JEL Classification: E24, E25, J24, J31

Keywords: Experience Premium, Return to Experience, Experience-labor Complementarity, Wage Dispersion, Consistent Aggregation

*First version: December 2005. This version: March 2008. We would like to thank Aureo De Paula, Jeremy Greenwood, Elena Krasnokutskaya, Bob Lucas, Ludo Visschers and seminar participants at New York University, the University of Pennsylvania, the Federal Reserve Bank of Philadelphia, the Federal Reserve Board of Governors, the 2007 NBER Summer Institute, and the Search & Matching Club meeting at the University of Pennsylvania for their comments. We are very grateful to Eva Meyersson Milgrom and Jeremy Fox for help with Swedish SAF data and to Fane Naja Groes for her help with the Danish data. This research has been supported by the National Science Foundation Grant No. SES-0617876.

1 Introduction

In this paper we evaluate the extent to which changing demographics alone can account for the large changes in the return to labor market experience over time. We reach a surprising conclusion that most of the observed changes in return to experience are accounted for by changes in the relative supply of experience. We find this conclusion surprising because, while it is implied by the basic forces of supply and demand, it has been overlooked by an extensive existing literature. This may be partly due to the fact that it is not the absolute supply of experience that determines its reward but its supply relative to other productive inputs. We provide a theory that enables the measurement of this relationship.

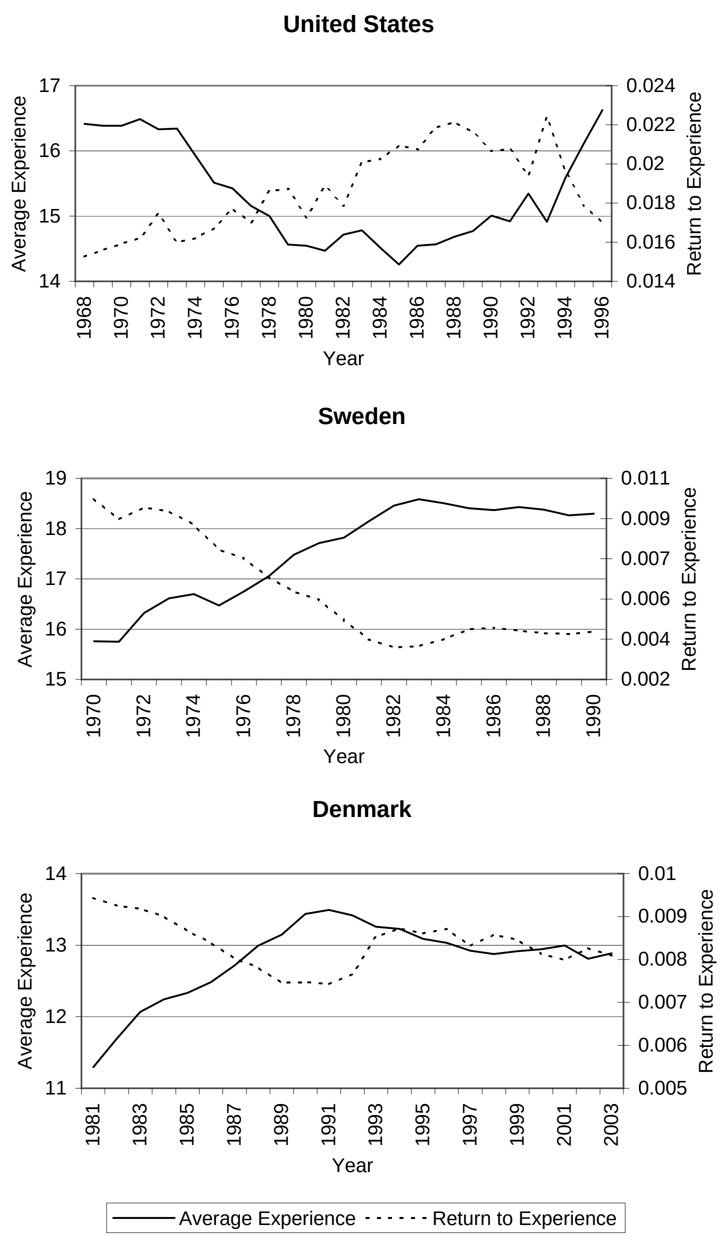
The motivation for our analysis comes from a strong negative correlation between the return to experience and the experience per worker in the data. Consider a year t cross-sectional regression of log wages of individual i on years of labor market experience

$$log(w_{it}) = \beta_t Experience_{it} + \gamma_t C_{it} + \epsilon_{it}, \tag{1}$$

where C_{it} includes a constant, education and sex controls. We estimate this equation on the U.S. PSID data for 1968-1996, Swedish SAF data for 1975-1995 that cover approximately 60% of private-sector employment in Sweden, and Danish data that cover 100% of the population over the 1980-2003 period. The data are described in the Appendices A1.1, A1.3, and A1.4. In each year, and in each country the sample is restricted to 20- to 65-year-old individuals with available observations on wages. In Figure 1, we plot the estimated rate of return to experience β_t against time for each country. On the same figure for each country, we plot the series of years of labor market experience per worker. The relationship between these two series is striking. The correlation coefficient between the rate of return to experience and the experience per worker is -0.75 in the U.S., -0.97 in Sweden, and -0.8 in Denmark. This correlation becomes substantially stronger once we consistently measure the efficiency units of experience using the model we develop below.[1]

[1]We are very grateful to Eva Meyersson Milgrom and Jeremy Fox for computing these statistics from the Swedish data and to Fane Naja Groes for computing them from the Danish data. Since we do not have personal

Figure 1: Experience per Worker vs. Return to Experience



Note - Source data and sample restrictions are described in Appendices A1.1, A1.3, and A1.4. Average experience is measured as experience per worker. Return to experience in each year represents the estimated coefficient $\hat{\beta}_t$ from regression 1.

For the U.S., the changes in the rate of return to experience over time are well documented. The evidence summarized in Topel (1997) and Katz and Autor (1999) implies that the cross-section age-earnings profiles steepened within education groups throughout the 1970s and 1980s and flattened in the 1990s. We find that these changes in the return to experience are nearly perfectly aligned with changes in the aggregate experience per worker. Throughout the 1970s there was a large entry of inexperienced workers born during the baby boom into the labor force and a sizable increase in female labor force participation. This reduced the experience per worker and coincided with the large increase in the return to experience. As these workers accumulated labor market experience over time, experience per worker increased substantially, and the return to experience has declined.

We propose a theory that enables the measurement and quantitatively accounts for this relationship. We think of workers as providing two distinct productive services - physical effort, which we refer to as "labor," and services of the skill accumulated with labor market experience, which we refer to as "experience." Depending on their labor market histories, workers accumulate different amounts of experience. Competitive firms can bundle workers to maintain the desired labor to experience ratio as in Heckman and Scheinkman (1987). This implies that prices of the two services provided by workers are competitively determined in the market. We consider an aggregate technology that allows for complementarity between labor and experience. The aggregate production function approach (with the competitive pricing of the bundled inputs) has been recently used in the literature to study earnings dynamics, e.g., Heckman, Lochner, and Taber (1998) and Guvenen and Kuruscu (2007a,b).

Allowing for complementarity between labor and experience provides a possible link between demographic composition and the return to experience. If labor and experience are complementary, any demographic change that affects their ratio will also affect the return to experience and hence relative wages. This effect is absent if labor and experience are perfect substitutes in aggregate production. The model enables us to estimate the magnitude of the

access to the microdata from Denmark and Sweden, we conduct our analysis in the rest of the paper based on U.S. data only.

experience-labor complementarity. Given this estimate we can quantitatively evaluate the effect of demographic change on the evolution of the rate of return to experience.

In our benchmark estimation, we use the U.S. PSID data to measure the aggregate stocks of labor and experience in each year between 1968 and 1996. We identify the parameters of aggregate technology using disaggregated earnings data, by estimating individual earnings equations that consistently aggregate. Assuming competitive factor markets, we are able to identify the technology parameters from the time-series variation in the ratio of the stock of labor to the stock of experience. We find a strong experience-labor complementarity with the elasticity of substitution between experience and labor at 0.3. Having estimated the technology parameters, we quantitatively evaluate to what extent the observed changes in the experience to labor ratio can account for the changing return to experience. We conclude that these changes account for the evolution of the return to experience nearly perfectly.

To confirm the existence and the magnitude of complementarity between labor and experience, we also estimate the model using the U.S. decennial census data, where the identification comes from the *cross-sectional* variation of the return to experience and the ratio of experience to labor across U.S. states. Despite the different source of identification, we obtain a similar estimate of the elasticity of substitution between experience and labor.

Our paper contributes to the large body of literature devoted to measuring and understanding the substantial change in the return to skill in the U.S. labor market. Skill is typically defined according to its two observable dimensions: education and labor market experience. Katz and Autor (1999) note that this definition of skill follows naturally from the classic models that link higher earnings for more educated workers and upward sloping age-earnings profiles to acquisition of human capital through education and on-the-job training (Becker (1962, 1993), Ben-Porath (1967), Mincer (1974)). In the influential Ben-Porath (1967) formulation, skills acquired through education and on-the-job training are perfectly substitutable. Thus, the literature has focused on accounting for the returns to education implicitly assuming that since work experience provides the same fundamental skill, the same explanation would also apply to the return to experience.[2] Our findings in this paper suggest that skills acquired through schooling and labor market experience may represent distinct forms of human capital priced separately. This follows because the rising returns to education due to, say, skill-biased technical change or capital skill complementarity did not affect the relationship between the relative supply of experience and its return.

Another branch of the literature has explored the relationship between the size of a cohort and relative earnings of its members. Motivated by the baby boom generation experience, Freeman (1979), Welch (1979) and Berger (1985) provided early empirical evidence that larger cohorts suffer depressed earnings upon entry into the labor market. More recent evidence is summarized in Wasmer (2001) and Triest, Sapozhnikov, and Sass (2006). Kim and Topel (1995) found that a sharp decline in the share of young workers in South Korea was associated with an increase in their relative earnings. Despite this suggestive evidence, Topel (1997) summarizes this literature by saying: "The effects of cohort size on earnings tend to be a sideline in the inequality literature." Our theory and quantitative results show that demographic change arising from changes in cohort size may be key to understanding the dynamics of the returns to experience and the associated wage inequality across cohorts.[3]

The paper offers several additional contributions. First, the existing literature proxies actual work experience by the age-based potential experience (equal to age minus years of education minus six). In contrast, we directly measure actual work experience, which enables us to separately identify the effects of age and work experience on wages. We find that the hump shape in the return to experience over the life-cycle is not driven by the decreasing returns in accumulating experience. Instead, it is driven by a less efficient utilization of accumulated experience with age. Second, on a methodological level, we show that it is possible to identify parameters of the aggregate production function from individual data by maintaining a consis-

[2]Heckman, Lochner, and Taber (1998) relax this assumption in their analysis.

[3]There is also a literature that investigates the effects of cohort size on human capital aquisition through schooling (e.g., Flinn (1993a) and refernces therein) and on-the-job training (Flinn (1993b)). At the more general level, demographic change was recently found to be relevant for understanding other economic phenomena. Jaimovich and Siu (2007) explore the effects of the demographic change on business cycle volatility. Fisher and Gervais (2007) study the effects of demographic change on the reduction in volatility of residential investment since the mid-1980s.

tency between individual earnings equations and the aggregate production function. Moreover, we show that estimation of the technology parameters using the microdata can identify the size and direction of the residual technical change.

The remainder of the paper is organized as follows. In Section 2 we describe the model that rationalizes the relationship between the return to experience and the aggregate experience-labor ratio. In Section 3 we describe our estimation procedure. In Section 4 we present estimation results and conduct sensitivity analysis on model specification. Section 5 concludes.

2 Model

We construct an aggregate production function, the parameters of which will be estimated, and characterize an individual earnings equation to evaluate whether the changes in the relative abundance of the aggregate factor inputs, and hence their marginal products, can account for the observed dynamics of the return to experience in the individual earnings equation. We show that the individual earnings consistently aggregate to the aggregate earnings as is implied by our aggregate production function, when aggregate stocks of labor and experience are consistently measured in terms of effective units. Household decisions determine the aggregate stocks of labor and experience, but we take the household decisions as given in estimating the aggregate technology parameters because the marginal products depend only on the aggregate stocks of those inputs. Such an approach is similar to that of Krusell, Ohanian, Ríos-Rull, and Violante (2000). Thus, we can simplify the analysis considerably by abstracting from modeling household decisions.

2.1 Individual Earnings

Time is discrete, measured in years, and indexed by t. We assume that each worker is endowed with one unit of labor each year throughout his or her life. Each year in which an individual works more than a critical level of hours of work $\underline{h}$, the worker accumulates one unit of experience (or learning-by-doing) skill. Thus, the total stock of experience e, of a j-year-old worker is given by

$$e = \sum_{k=0}^{j} I(h_k > \underline{h}) \,, \tag{2}$$

where I is equal to one if $(h_k > \underline{h})$ and zero otherwise. Thus, we allow for possible differences between *actual* work experience and *potential* work experience, the years since completion of schooling (the typical age-based measure of experience).

One might expect the transformation of years of work experience e into the efficiency units of experience supplied by the worker to the market to depend on age. For example, a worker who accumulated, say, 5 years of experience by the age of 30 may be supplying a different amount of efficiency units of experience than a worker who accumulated 5 years of experience by the age of 60. This relationship may also depend on the schooling level s of the worker. We partition the workforce into a low-education group with final years of schooling less than or equal to 12 ($s = 0$) and a high-education group with final years of schooling beyond 12 ($s = 1$). Similar to work experience, we allow the effective units of labor to depend on age and schooling level. This is meant to summarize the evolution of physical aptitude and, possibly, the evolution of utilization of physical abilities over the life-cycle.

Let $\lambda_E(j, s)$ and $\lambda_L(j, s)$ be such effective-unit transforming factors for experience and labor respectively, where we do not limit their range except the normalization $\lambda_E(0, s) = \lambda_L(0, s) = 1$. We will refer to these transforming factors as "life-cycle efficiency" schedules. Thus, a j-year-old worker in schooling group s who acquired e years of work experience provides $\lambda_E(j, s)e$ units of effective experience and $\lambda_L(j, s)$ units of effective labor to the labor market. Modeling these life-cycle efficiency schedules enables us to separately identify the return to age and the return to experience (conditional on schooling level).[4] As we will discuss in Section 4.2 modeling life-cycle efficiency schedules allows us to better measure the aggregate stocks of labor and experience but it does not drive our finding of a strong complementarity between labor and experience in the aggregate production function. We also allow for differences in individual

[4]We have assumed that the technology that maps years of work experience e into units of experience skill is linear. In Section 4.2.1 we will relax this assumption and allow general polynomial specifications to capture the possible curvature of this mapping, e.g., diminishing returns in experience accumulation. We find that the linear mapping is indeed supported as the best specification by the data.

productivity per hours of work z that are determined by a set of exogenous characteristics.

Let R_{L_t} and R_{E_t} denote the market prices of labor and experience, respectively. The total earnings y_{it} of worker i at date t with age j_{it}, work experience e_{it} and schooling level s_{it}, who works h_{it} hours and whose individual productivity equals z_{it}, is given by

$$\begin{aligned} y_{it} &= \left[R_{L_t}\lambda_L(j_{it}, s_{it}) + R_{E_t}\lambda_E(j_{it}, s_{it})e_{it}\right] h_{it}z_{it} \\ &= R_{L_t}\lambda_L(j_{it}, s_{it}) \left[1 + \frac{R_{E_t}}{R_{L_t}}\frac{\lambda_E(j_{it}, s_{it})}{\lambda_L(j_{it}, s_{it})}e_{it}\right] h_{it}z_{it}. \end{aligned} \tag{3}$$

We label the ratio of the price of experience to the price of labor as the "experience premium" and denote it by

$$\Pi_{E_t} \equiv \frac{R_{E_t}}{R_{L_t}}.$$

Note that we distinguish the "experience premium" from the "rate of return to experience," which we will define below in Equation (22) as an appropriately evaluated derivative of the log wage with respect to experience. This distinction is useful because the experience premium depends only on aggregate state variables in the market, not on individual worker characteristics, while the return to experience may depend on individual characteristics.

It will also be convenient to define the relative life-cycle efficiency schedule of experience as the ratio of the life-cycle efficiency schedule of experience to that of labor

$$\lambda_{E/L}(j_{it}, s_{it}) \equiv \frac{\lambda_E(j_{it}, s_{it})}{\lambda_L(j_{it}, s_{it})}.$$

2.2 Aggregate Technology

Consider an aggregate production function that maps the aggregate stock of labor L_t and the aggregate stock of experience E_t into aggregate earnings Y_t such that

$$Y_t = A_t G\left(L_t, E_t\right), \tag{4}$$

where G is a constant-returns-to-scale function and A_t represents the aggregate productivity of labor and experience.

The Euler theorem implies

$$Y_t = A_t\left(G_{L_t}L_t + G_{E_t}E_t\right),$$

where $G_{L_t} = \frac{\partial G}{\partial L_t}$ and $G_{E_t} = \frac{\partial G}{\partial E_t}$ will be referred to as marginal products of labor and experience, respectively. The assumption of perfect competition determines the prices of labor and experience such that

$$R_{L_t} = A_t G_{L_t}, \tag{5}$$

$$R_{E_t} = A_t G_{E_t}. \tag{6}$$

2.3 Consistent Aggregation

Summing the individual earnings equation in (3) over individuals i at a given date t, we have

$$\begin{aligned}
\sum_i y_{it} &= R_{L_t} \sum_i \lambda_L(j_{it}, s_{it}) z_{it} h_{it} + R_{E_t} \sum_i \lambda_E(j_{it}, s_{it}) z_{it} h_{it} e_{it} \\
&= A_t G_{L_t} L_t + A_t G_{E_t} E_t \\
&= Y_t
\end{aligned}$$

where the aggregate inputs L_t, E_t are measured as

$$L_t = \sum_i \lambda_L(j_{it}, s_{it}) z_{it} h_{it}, \tag{7}$$

$$E_t = \sum_i \lambda_E(j_{it}, s_{it}) e_{it} z_{it} h_{it}, \tag{8}$$

and prices R_{L_t}, R_{E_t} are determined by Equations (5) and (6), respectively.

Thus, the individual earnings equations in (3) consistently aggregate to the aggregate earnings as is implied by the aggregate production function in (4). Note that this consistent aggregation holds for any homogeneous of degree one function G as long as the aggregate inputs L_t and E_t are consistently measured as in equations (7) and (8).

2.4 Labor-Experience Complementarity

To conduct our quantitative analysis we must choose a specific functional form for G. We restrict our attention to the commonly used class of constant elasticity of substitution (CES) production technologies. This type of production function is tractable, parsimonious, and provides a simple specification that allows us to evaluate the role of the change in the demographic composition of the workforce in driving the change in the experience premium. Specifically,

$$Y_t = A_t \left(L_t^{\mu} + \delta E_t^{\mu}\right)^{\frac{1}{\mu}}, \tag{9}$$

where the elasticity of substitution between L_t and E_t is measured by $\frac{1}{1-\mu}$ (where $\mu \leq 1$), and the parameter $\delta > 0$ adjusts the relative scale between L_t and E_t.

The degree of substitutability between labor and experience is governed by the value μ. If $\mu = 1$, labor and experience are perfect substitutes. In this case, the demographic change affecting the ratio of labor to experience does not affect the experience premium. However, if labor and experience are not perfect substitutes, changes in the demographic composition of the workforce will affect the experience premium. This opens up the possibility of understanding the relative wage dynamics through demographic changes and bears a directly testable implication. The lower the value of μ, the stronger the complementarity between these two inputs (or the less substitutable they are).[5]

3 Estimation

3.1 Log Wage Equation

Using the aggregate production function in (9) and individual earnings equation in (3), the log wage equation (for hourly earnings $w_{it} = \frac{y_{it}}{h_{it}}$) is

$$\ln w_{it} = \ln A_t + \ln G_{L_t} + \ln \lambda_L(j_{it}, s_{it}) + \ln \left[1 + \Pi_{E_t} \lambda_{E/L}(j_{it}, s_{it}) e_{it}\right] + \ln z_{it}, \tag{10}$$

where

$$G_{L_t} = \left(1 + \delta \left(\frac{E_t}{L_t}\right)^{\mu}\right)^{\frac{1}{\mu}-1}, \tag{11}$$

$$\Pi_{E_t} = \delta \left(\frac{E_t}{L_t}\right)^{\mu-1}. \tag{12}$$

Having no a priori knowledge on the functional form for the labor and experience life-cycle efficiency schedules $\lambda_L(j, s)$ and $\lambda_E(j, s)$, we approximate each of them by an exponential

[5] A special case of $\mu = 0$ implies unit elasticity of substitution, i.e., the Cobb-Douglas specification. Values of $\mu < 0$ indicate lower substitutability than in the Cobb-Douglas case.

function of a second-degree polynomial for each schooling group s[6]

$$\lambda_L(j,s) = \exp(\lambda_{L,1}(s)\,j + \lambda_{L,2}(s)\,j^2), \tag{13}$$

$$\lambda_E(j,s) = \exp(\lambda_{E,1}(s)\,j + \lambda_{E,2}(s)\,j^2). \tag{14}$$

This implies that the relative life-cycle efficiency schedule of experience is given by

$$\lambda_{E/L}(j,s) = \exp(\lambda_{E/L,1}(s)\,j + \lambda_{E/L,2}(s)\,j^2),$$

where

$$\begin{aligned} \lambda_{E/L,1}(s) &= \lambda_{E,1}(s) - \lambda_{L,1}(s), \\ \lambda_{E/L,2}(s) &= \lambda_{E,2}(s) - \lambda_{L,2}(s). \end{aligned}$$

The individual productivity variable $\ln z_{it}$ is decomposed such that

$$\ln z_{it} = \alpha_t \chi_{it}, \tag{15}$$

where χ_{it} is a vector of the observable characteristics, including years of schooling, sex, race and geographic region.

Substituting these expressions into equation (10), we obtain the log wage equation to be estimated

$$\begin{aligned} \ln w_{it} &= \ln A_t + \ln G_{L_t} + (\lambda_{L,1}(s_{it})\,j_{it} + \lambda_{L,2}(s_{it})\,j_{it}^2) \\ &\quad + \ln\left[1 + \delta\left(\frac{E_t}{L_t}\right)^{\mu-1} \exp(\lambda_{E/L,1}(s_{it})\,j_{it} + \lambda_{E/L,2}(s_{it})\,j_{it}^2)\,e_{it}\right] + \alpha_t\chi_{it} + \epsilon_{it}, \end{aligned} \tag{16}$$

where ϵ_{it} represents the classical measurement error.[7] In contrast to a typical Mincerian specification, our earnings equation involves an explicit aggregate state variable, the experience-labor ratio $\frac{E_t}{L_t}$, which is the driving force of the experience premium dynamics. If E_t and L_t are perfect substitutes, i.e., $\mu = 1$, the log wage equation turns into a Mincerian earnings equation with a time-invariant experience premium.

[6] We experimented using higher order polynomials but found the coefficients on higher order terms to be insignificant.

[7] There exists an alternative interpretation of ϵ_{it} as the *i.i.d.* normal idiosyncratic productivity shock, in which case the appropriate specification is $\ln z_{it} = \alpha_t\chi_{it} + \epsilon_{it}$. Under such interpretation idiosyncratic productivity realizations will enter the calculation of the aggregate stocks of L and E. While theoretically the consistency of estimated technology parameters cannot be established with such formulation, in practice we estimated the model with this specification and found that the estimates are virtually unaffected by this choice of specification.

3.2 Identification

The log wage equation in (16) includes all parameters of the model. In particular, given the measurement of aggregate inputs E_t and L_t, the variation of experience premium 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 *correlation* between the relative price Π_{E_t} and the relative factor endowment $\frac{E_t}{L_t}$ over time t identifies μ (which is scale free).The *average magnitude* of the 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 the life-cycle efficiency schedules, i.e., $\lambda_E(0,s) = \lambda_L(0,s) = 1$. 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, renormalizing $\lambda_E(0,s) = \lambda_L(0,s) = l$ for any arbitrary constant l such that $\lambda_{E/L}(0,s) = 1$ leaves the estimate of δ unchanged. However, if we normalize the life-cycle efficiency schedules asymmetrically between experience and labor such that $\lambda_L(0,s) = a$ and $\lambda_E(0,s) = b$, hence $\lambda_{E/L}(0,s) = c = b/a \neq 1$, the coefficient function in front of experience in the log wage becomes $\delta \left(c\frac{E_t}{L_t}\right)^{\mu-1} c\lambda_{E/L}(j,s) = \widetilde{\delta}\left(\frac{E_t}{L_t}\right)^{\mu-1} \lambda_{E/L}(j,s)$, where $\widetilde{\delta} = \delta c^{\mu}$. Thus, the estimated value of δ may change. The normalization of the life-cycle efficiency schedule of *labor* affects the scale of the aggregate productivity term. Specifically, with $\lambda_L(0,s) = a$, the aggregate productivity term turns to $\ln aA_t$. Note, however, that estimates of μ as well as the life-cycle efficiency schedules, our key parameters, are *not* affected by this normalization.

Consistent aggregation requires that the aggregate inputs L_t and E_t be measured as in equations (7) and (8); hence, they depend on the parameters of life-cycle efficiency schedules $\lambda_L(j,s)$ and $\lambda_E(j,s)$ for $s \in \{0,1\}$, and the coefficient vector α_t of the observable individual productivity attributes χ_{it}. That is, for a consistent estimation of the entire model, the aggregate inputs L_t and E_t need to be expressed in terms of these life-cycle efficiency and productivity parameters and to be estimated at the same time. This creates another complicated nonlinearity inside the arguments of the already nonlinear log wage equation. Thus, we employ the following two-step estimation procedure, which allows us to have *statistical consistency* in estimating the technology parameters, and also *consistent aggregation* of the data in precisely the same way our model implies.

In the *first stage*, we obtain the estimates of eight parameters of the life-cycle efficiency schedules $\widehat{\lambda}_L(j,s)$ and $\widehat{\lambda}_{E/L}(j,s)$ and the estimates of the time-varying coefficients ($\widehat{\alpha}_t$, $\widehat{D}_t$ and experience premium $\widehat{\Pi}_{E_t}$) applying a nonlinear least-squares method to the following log wage equation

$$\begin{aligned}\ln w_{it} \quad &= \quad D_t + \left(\lambda_{L,1}\left(s_{it}\right) j_{it} + \lambda_{L,2}\left(s_{it}\right) j_{it}^2\right) \qquad (17)\\ &\quad + \ln\left[1 + \Pi_{E_t}\exp(\lambda_{E/L,1}\left(s_{it}\right) j_{it} + \lambda_{E/L,2}\left(s_{it}\right) j_{it}^2)\, e_{it}\right] + \alpha_t\chi_{it} + \epsilon_{it},\end{aligned}$$

where $\ln A_t + \ln G_{L_t}$ in log wage equation (16) is replaced by D_t, and $\delta\left(\frac{E_t}{L_t}\right)^{\mu-1}$ by Π_{E_t}. Thus, in the first stage, we treat the marginal product of labor and the experience premium as time-varying constants. Furthermore, the estimates of the $\widehat{\lambda}_L(j,s)$, $\widehat{\lambda}_{E/L}(j,s)$ parameters imply the estimate of $\widehat{\lambda}_E(j,s)$ parameters such that

$$\widehat{\lambda}_E(j,s) = \widehat{\lambda}_{E/L}(j,s)\widehat{\lambda}_L(j,s).$$

These estimates allow us to construct the estimated aggregate inputs $\widehat{L}_t$ and $\widehat{E}_t$ at each date t as in equations (7) and (8) such that

$$\begin{aligned}\widehat{L}_t \quad &= \quad \sum_i \widehat{\lambda}_L(j_{it}, s_{it})\widehat{z}_{it}h_{it}, \qquad (18)\\ \widehat{E}_t \quad &= \quad \sum_i \widehat{\lambda}_E(j_{it}, s_{it})\widehat{z}_{it}h_{it}e_{it}, \qquad (19)\end{aligned}$$

where

$$\widehat{z}_{it} = \exp\left(\widehat{\alpha}_t\chi_{it}\right). \qquad (20)$$

In the *second stage,* we obtain the estimates of the technology parameters μ and δ by again applying the nonlinear least-squares to the log wage equation in (16), but using the first-stage estimates of the life-cycle efficiency schedules parameters, $\widehat{\alpha}_t$ and $\widehat{D}_t$, and the implied experience-labor ratio $\frac{\widehat{E}_t}{\widehat{L}_t}$ constructed in equations (18) and (19)

$$\begin{aligned} \ln w_{it} &= \widehat{D}_t + (\widehat{\lambda}_{L,1}(s_{it})\, j_{it} + \widehat{\lambda}_{L,2}(s_{it})\, j_{it}^2) \\ &\quad + \ln\left[1 + \delta \left(\frac{\widehat{E}_t}{\widehat{L}_t}\right)^{\mu-1} \exp(\widehat{\lambda}_{E/L,1}(s_{it})\, j_{it} + \widehat{\lambda}_{E/L,2}(s_{it})\, j_{it}^2) e_{it}\right] + \widehat{\alpha}_t \chi_{it} + \epsilon_{it}. \end{aligned} \tag{21}$$

Separation of the second-stage estimation from the first-stage estimation helps us better isolate the relation between the experience-labor ratio and the experience premium as well as its effect on technology parameters μ and δ.

4 Results

We first use the Panel Study of Income Dynamics (PSID) data over the 1968-1996 period. In each year, we measure the experience premium and the experience-labor ratio. The co-movement of these variables over time identifies the parameters of the production function. We then use the estimated parameter values to evaluate how accurately our model predicts the dynamics of the experience premium. We find that our two-parameter production function quantitatively matches the time series of the experience premium very well.

The identifying source of the technology parameters is the correlated variation between the experience premium and the experience-labor ratio. The source of the variation in these two variables can come from either across time or across space. In the second experiment, we use cross-sectional data from the U.S. census and estimate the production parameters from the variation of experience premium and experience-labor ratio across U.S. states. Remarkably, the cross-sectional estimate of the elasticity of substitution between labor and experience is similar to the time-series estimate from the PSID. Thus, in both experiments, we find strong evidence of complementarity between labor and experience. This implies that demographic composition can play an important role in determining the experience premium and hence relative wages.

4.1 PSID Time Series

4.1.1 Estimation

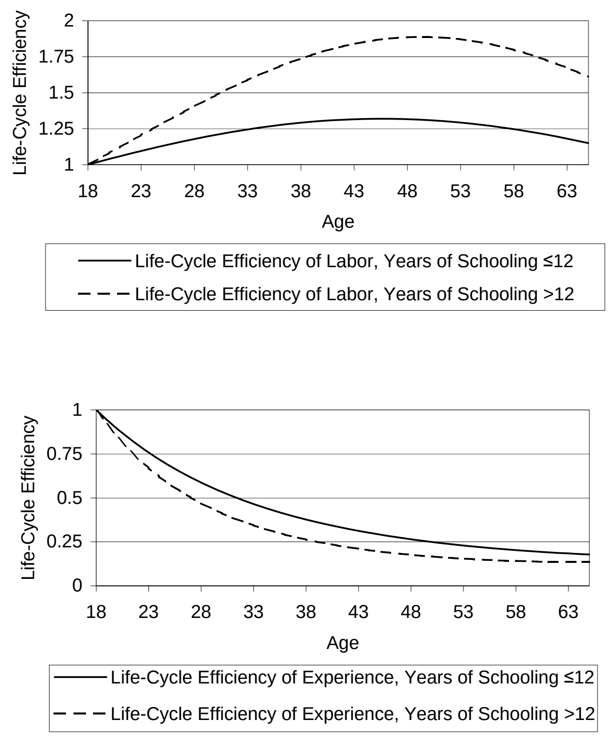
In our benchmark estimation, we use the PSID data. See Appendix A1.1 for the details of our sample selection and the construction of the variables. An important contribution of our analysis is to construct measures of *actual* work experience, utilizing the panel structure of the PSID (the procedure is described in Appendix A1.1). It is typical in the literature to proxy actual experience with *potential* experience, defined as age minus years of schooling minus six. By construction, once years of schooling are controlled for, that experience measure can capture only the effects of age on the return to experience. Using actual experience, however, we are able to separate the effects of age and experience on earnings.

In the first stage, we obtain estimates of the parameters of life-cycle efficiency schedules for labor and experience $\left(\widehat{\lambda}_{L,k}(s), \widehat{\lambda}_{E,k}(s)\right)_{k=1,2;s=0,1}$ and the time-varying parameters $\left(\widehat{\alpha}_t, \widehat{D}_t, \widehat{\Pi}_{E_t}\right)_{t=1968}^{1996}$, applying a nonlinear least-squares method to the log-wage equation (17). The estimates of these coefficients and their standard errors are reported in Appendix Tables A-2 and A-3. All the parameters are estimated precisely. This is particularly important in the case of the experience premium, $\left(\widehat{\Pi}_{E_t}\right)_{t=1968}^{1996}$, because we will evaluate the performance of the model by its ability to replicate this estimated series.

Figure 2 plots the estimated life-cycle efficiency schedules of labor and experience for each schooling group. The values of the life-cycle efficiency schedules for labor exceed one for all ages for both schooling groups; hence, the effective units of labor increase over the life cycle. The increase is much larger for the high-education group (workers with more than 12 years of schooling) than the low-education group (workers with up to 12 years of schooling). However, they are hump-shaped for both schooling groups, peaking at age 46 for the low-education group and at 49 for the high-education group.

In contrast, the estimated life-cycle efficiency schedule for experience is monotonically decreasing and below unity over the entire age range as shown in Figure 2. Thus, there is a substantial benefit from accumulating experience early in life, and this benefit is larger for the

Figure 2: Life-Cycle Efficiency Schedules for Labor and Experience by Schooling Group



high-education workers, since their stock of experience depreciates at a faster rate than that of the low-education workers.

These life-cycle efficiency schedules together with the estimated hourly productivity term $\widehat{z}_{it} = \exp(\widehat{\alpha}_t \chi_{it})$ are reflected in calculating the aggregate stocks of labor and experience as in equations (18) and (19). Thus, changes in the composition of age groups as well as the composition of the productivity characteristics χ_{it} in the workforce affect these aggregate stocks directly via life-cycle efficiency schedules and $\widehat{z}_{it}$, and indirectly via the correlation between age and the characteristics χ_{it}.

Given the series of the experience premium $\widehat{\Pi}_{E_t}$ and the implied experience-labor ratio $\frac{\widehat{E}_t}{\widehat{L}_t}$ from the first-stage estimation above, we obtain the estimates of technology parameters μ and δ from the second-stage estimation, which are reported in Table 1. The correlation coefficient between the experience premium and the experience-labor ratio is remarkably high at -0.96. This correlation identifies the curvature parameter $\hat{\mu} = -2.35$. The scale parameter is estimated at $\hat{\delta} = 6.93$, adjusting the relative magnitudes between the average experience premium of 18% and the average experience-labor ratio of 3. Both parameters are fairly precisely estimated with low standard errors of 0.060 and 0.447, respectively, for μ and δ. In addition, given the two-stage estimation, we computed bootstrap standard errors for μ and δ, which are equal to 0.295 and 3.26, respectively. According to the two most commonly used goodness-of-fit measures, adjusted R^2 at 0.926 and the root mean-squared error ($RMSE$) at 0.596, the fit seems very good.[8]

Table 1: PSID estimates of technology parameters.

Parameter	Estimate	Second Stage standard error	Bootstrap standard error
μ	-2.35	0.060	0.295
δ	6.93	0.447	2.260
R^2	0.926		
$RMSE$	0.596		

Note - The entries represent the results of the second-stage estimation of technology parameters of the benchmark specification from the PSID. Complete first-stage estimation results are provided in Appendix Tables A-2 and A-3. For sample restrictions and variable construction procedures, see Appendix A1.1. See Section 3 for details of the estimation procedure.

4.1.2 Experience-Labor Complementarity

The estimate of μ implies an elasticity of substitution between labor and experience of 0.3. Thus, we find substantial complementarity between labor and experience. To our knowledge, this is the first estimate for the elasticity of substitution between labor and experience for the U.S. The only alternative estimate of the experience-labor complementarity is reported in Jeong and Kim (2006). They find it to be equal to 0.4 in Thai data, very similar to our estimate for the U.S.

For comparison, the typical estimates of elasticity of substitution between college and high school workers are between 1.4 and 2.5 (e.g., Heckman, Lochner, and Taber (1998), Card and

[8]The goodness-of-fit measures of a nonlinear model are to be interpreted with caution. The R^2 of the nonlinear model does not tell us the precise share of the actual variation of the data that is predicted because of the curvature. The closer to zero the $RMSE$ is, the better the fit, although it has no upper bound.

Lemieux (2001), Gallipoli, Meghir, and Violante (2007)). The point estimate of this elasticity is 1.41 in Katz and Murphy (1992), although they suggest that a range of estimates with the elasticity as low as 0.5 are also consistent with the data. Krusell, Ohanian, Ríos-Rull, and Violante (2000) estimate the elasticity of substitution between the high school workers and a composite input of college workers and capital equipment at 1.7. Their estimate of the complementarity between the two inputs inside the composite input, i.e., college workers and capital equipment, labeled "capital-skill complementarity," is 0.67. The elasticity of substitution between capital and labor in the neoclassical Cobb-Douglas production function is equal to 1. Thus, the magnitude of the experience-labor complementarity is stronger but in line with the other types of complementarity measured in the literature.

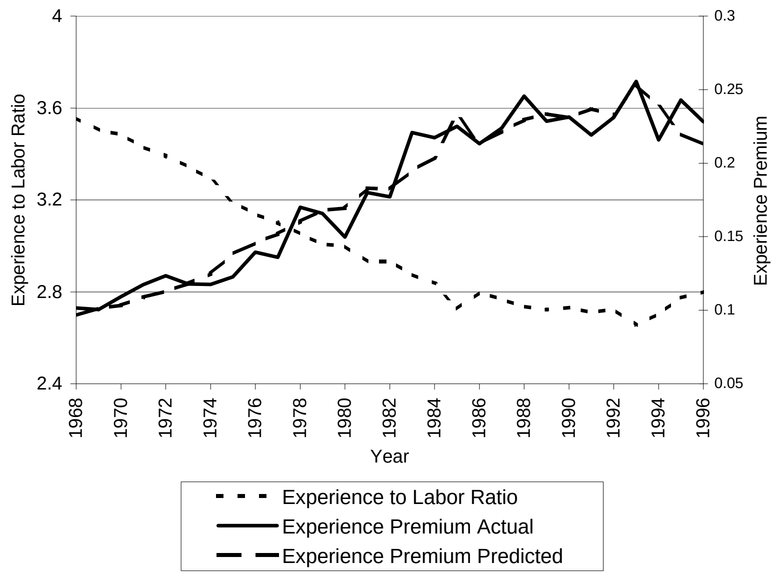
4.1.3 Experience Premium

Figure 3 displays the estimated experience premium $\widehat{\Pi}_{E_t}$ over the 1968-1996 period. There is a substantial movement of the premium, which increases with an average growth rate of 2.93% per year over the sample period.[9] Figure 3 also shows the estimated experience-labor ratio over the same period, showing a clear negative co-movement with the experience premium. Thus, by measuring the relative supply of experience by the ratio of aggregate experience relative to aggregate labor (each weighted by its life-cycle efficiency factors and hourly productivity), we uncover a remarkably tight relationship between the relative price of experience (i.e., the experience premium) and the relative supply of experience.

Figure 3 also displays the experience premium predicted by our model at the above estimates of the aggregate production function and the aggregate experience-labor ratio. Despite the parsimonious specification (two parameters μ and δ, and a single state variable $\frac{E}{L}$), the model tracks the actual time-path of the experience premium very closely. The correlation coefficient between the actual experience premium $\widehat{\Pi}_{E_t}$ (implied by the first-stage estimates) and the predicted experience premium (implied by the second-stage estimates) is 0.97.

[9]Recall that the level of $\widehat{\Pi}_{E_t}$ depends on the normalization of $\lambda_{E/L}$. Its growth rate is independent of this normalization.

Figure 3: Actual and Predicted Experience Premium and Experience to Labor Ratio



4.1.4 Rate of Return to Experience

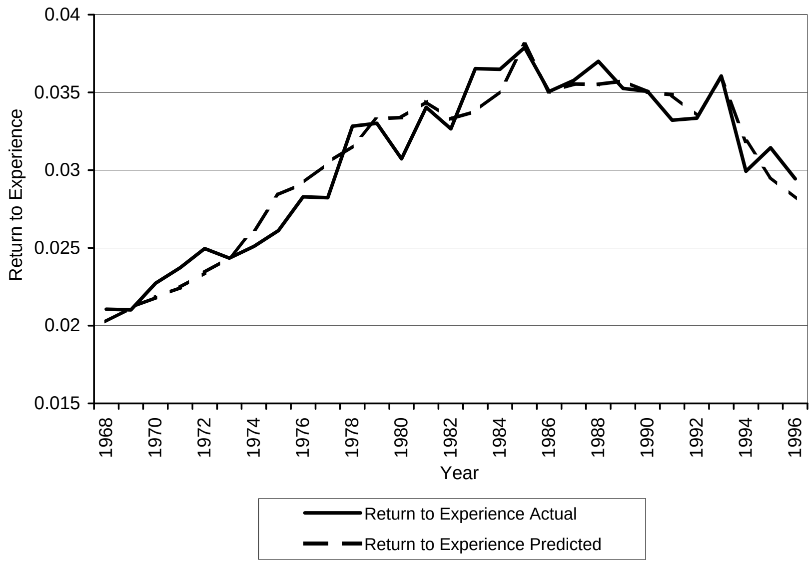
The experience premium in our model measures a relative price of experience. It is common to all workers, independent of their individual characteristics. Over the life-cycle, however, the *rate of return to experience* of individual workers may vary depending on their age and schooling level (because of the life-cycle efficiency schedules) as well as their experience level (because of the nonlinearity of the log wage equation). Define the rate of return to experience as the marginal wage increment to the addition of one more year of experience such that

$$\frac{d \ln w_{it}}{de_{it}} = \frac{\Pi_{E_t} \lambda_{E/L}(j_{it}, s_{it})}{1 + \Pi_{E_t} \lambda_{E/L}(j_{it}, s_{it}) e_{it}}. \tag{22}$$

Thus, the individual rate of return to experience is rising in the aggregate experience premium Π_{E_t}, and falling in the individual level of experience e_{it}. It is falling in worker's age j_{it} when the relative ratio of life-cycle efficiency schedules $\lambda_{E/L}(j_{it}, s_{it})$ is falling in age, and it is rising in age when this ratio is rising in age. Note that both the level and the growth rate of the rate of return to experience are independent of the normalization of $\lambda_{E/L}$.

To summarize the evolution of the returns to experience in a single time-series, we must decide on where to evaluate the function in (22). Consider a "representative worker" whose

Figure 4: Actual and Predicted Return to Experience for the Representative Worker



rate of return to experience is given by $\frac{\Pi_{E_t}\Lambda_t}{1+\Pi_{E_t}\Xi_t}$ with Λ_t and Ξ_t measured as follows:

$$\Lambda_t \equiv \frac{\sum_i \lambda_E(j_{it}, s_{it}) z_{it} h_{it}}{\sum_i \lambda_L(j_{it}, s_{it}) z_{it} h_{it}}, \tag{23}$$

$$\Xi_t \equiv \frac{\sum_i \lambda_E(j_{it}, s_{it}) e_{it} z_{it} h_{it}}{\sum_i \lambda_L(j_{it}, s_{it}) z_{it} h_{it}}. \tag{24}$$

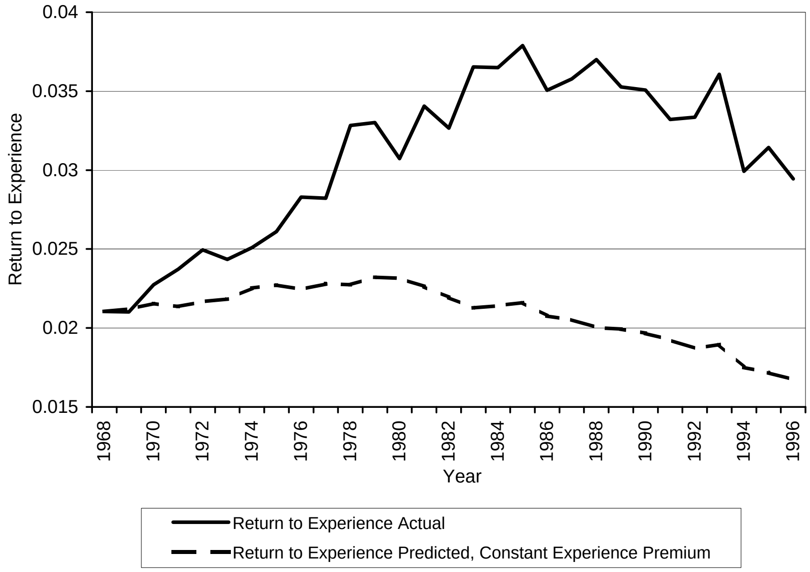
Hence, this worker supplies the aggregate effective labor and aggregate effective experience.[10]

Figure 4 plots the rate of return to experience for the representative worker using estimates of $\widehat{\Pi}_{E_t}$ and $\left(\widehat{\lambda}_{L,k}(s), \widehat{\lambda}_{E,k}(s)\right)_{k=1,2;s=0,1}$ from the first-stage estimation. The rate of return is sizable and changes substantially over time from 2.1% in 1968 to 3.8% in 1988, and then to 2.8% in 1996. Figure 4 also compares this path with the predicted path of the return to experience implied by the technology estimates $\hat{\mu}$, $\hat{\delta}$ from the second-stage estimation, which indicates that the fit is very good.

Equation 22 implies that there are two potential sources of change in the return to experience: first, changes in the experience premium $\widehat{\Pi}_{E_t}$, and second, changes in the age and schooling composition of workers affecting the measured $\lambda_{E/L}$. The first effect is driven by the

[10]By construction, the effective units of experience to labor for this representative worker coincide with the aggregate experience to labor ratio $\Xi_t = \frac{E_t}{L_t}$. We weight the life-cycle efficiency schedules of experience and labor by $z_{it}h_{it}$ to obtain the effective units of experience and labor at the aggregate level.

Figure 5: Actual Return to Experience for the Representative Worker and Counterfactual with Constant Experience Premium

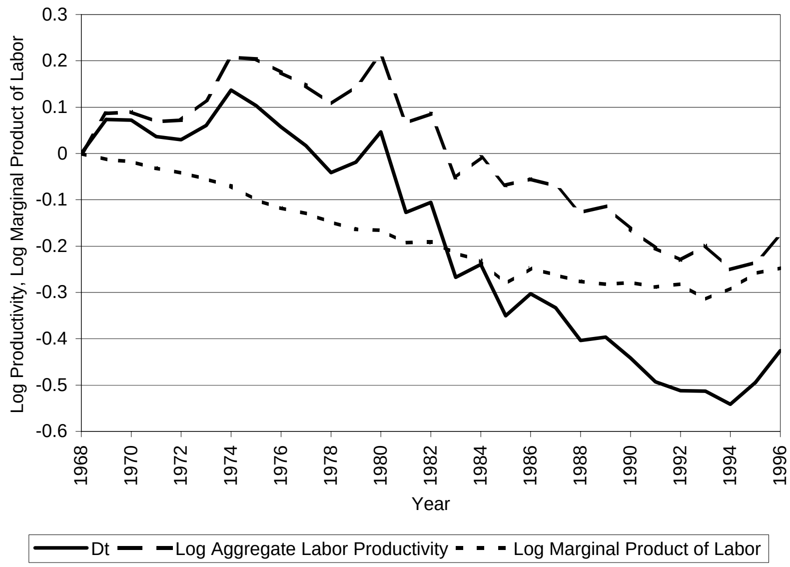


complementarity between experience and labor, and the second is a composition effect that can arise without such complementarity. Our model provides a way to isolate the contribution of the composition effect. To do so, we hold the experience premium $\widehat{\Pi}_{E_{68}}$ fixed at its level in 1968 and generate a counterfactual series of the rate of return to experience.[11] This is compared with the actual series (implied by the first-stage estimates) in Figure 5. The figure shows that the composition effect alone does not account for the observed changes in the aggregate rate of return to experience. Instead, it is the complementarity effect that drives the changes in the aggregate rate of return to experience.

We obtain similar results when considering the average rate of return to experience across workers instead of the return to experience of a "representative worker." The model predicts the average rate of return to experience well, and its movement over time is again driven by the complementarity effect, as shown in Appendix Figures A-1 and A-2.

[11]The choice of when to fix the experience premium is inconsequential.

Figure 6: $\widehat{D}_t$, Aggregate Labor Productivity and Marginal Product of Labor



4.1.5 Aggregate Productivity

Given the estimates of $\widehat{\mu}$ and $\widehat{\delta}$, we can uncover the marginal product of labor G_{L_t} using equation (11) such that

$$\widehat{G}_{L_t} = \left(1 + \widehat{\delta}\left(\frac{\widehat{E}_t}{\widehat{L}_t}\right)^{\widehat{\mu}}\right)^{\frac{1}{\widehat{\mu}}-1}. \tag{25}$$

Combining $\widehat{G}_{L_t}$ and the estimates of the time-varying constant terms $\widehat{D}_t$, the log of the aggregate productivity term A_t can be identified by

$$\ln \widehat{A}_t = \widehat{D}_t - \ln \widehat{G}_{L_t}. \tag{26}$$

Thus, we can decompose the changes in D_t into a component due to changes in the experience-labor ratio $\frac{E_t}{L_t}$ and a component due to changes in aggregate productivity level A_t.

As shown in Figure 6, $\widehat{D}_t$ displays no trend until 1974 and then decreases until 1994 before rising again. The figure also presents the decomposition of this term into log of marginal product of labor $\ln \widehat{G}_{L_t}$ and log of aggregate productivity $\ln \widehat{A}_t$, according to (26). To facilitate comparisons of the movements among these three variables, we normalize values in 1968 to zero (by subtracting a constant). The log of marginal product of labor $\ln \widehat{G}_{L_t}$ has decreased over the sample period accounting for most of the slowdown in the growth of D_t and its eventual decline. This translates into a substantial 26% fall in the level of the marginal product of labor between 1968 and 1996. It is encouraging that the model not only accounts for the dynamics of the return to experience but also endogenously generates a substantial decline in the interceprt of the wage equation. It is not a priori clear that these two features of the data might be related, but the model provides a tight link between them. When the experience to labor ratio declines, the marginal product of labor declines as well. This is exactly what the intercept of the wage equation captures.

The share of aggregate earnings accruing to labor input implied by (9) is

$$\frac{\frac{dY_t}{dL_t}L_t}{Y_t} = \frac{1}{1+\delta\left(\frac{E_t}{L_t}\right)^{\mu}} = \frac{1}{1+\Pi_t\frac{E_t}{L_t}}.$$

Our estimate of $\mu < 0$ implies that the labor input's share of aggregate earnings is rising in $\frac{E_t}{L_t}$. Given our estimates for $\widehat{\Pi}_{E_t}$ and $\frac{\widehat{E}_t}{\widehat{L}_t}$ from the first-stage estimation, we find that the share of labor falls steadily from 72% in 1968 to 61% in 1996 (with an average of 66%). This implies that the relative importance of experience skill over labor has increased during the sample period.

4.2 Alternative Specifications

In this subsection we first evaluate whether model specifications with nonlinear experience production technology are supported by the data. Next, we progressively make the model specification more and more coarse by first restricting the life-cycle efficiency schedules to be the same across schooling groups, then dropping the life-cycle efficiency schedule of labor from the model (while keeping the life-cycle efficiency schedule of experience common across schooling groups), and finally dropping the life-cycle efficiency schedule of experience as well. Across these specifications the model's ability to fit the data in the first-stage estimation becomes progressively weaker. The estimates of the complementarity between labor and experience, however, remain robust. We perform the experiments in a particular order, but the findings are robust to this ordering.

4.2.1 Curvature in Production of the Experience Skill

In the benchmark model, the experience skill is a linear function of years of experience. We now explore the effect of allowing for curvature in this relationship. Consider approximating it by a fourth order polynomial such that the effective units of experience supplied by individual i of age j_{it} and years of experience e_{it} is given by $\lambda_E(j_{it}, s_{it})\,(e_{it} + \eta_1 e_{it}^2 + \eta_2 e_{it}^3 + \eta_3 e_{it}^4)$ and the aggregate experience E_t in Equation (8) is replaced by

$$E_t = \sum_i \lambda_E(j_{it}, s_{it})\left(e_{it} + \eta_1 e_{it}^2 + \eta_2 e_{it}^3 + \eta_3 e_{it}^4\right) z_{it} h_{it}. \tag{27}$$

The identification and estimation of the model remain essentially unchanged. The only difference is that three additional parameters are estimated in the first stage. However, this may yield different values for the experience-labor ratios and the first-stage estimates of the experience premium, which in turn will affect the second-stage estimation in obtaining the technology parameters.

The resulting estimates for the technology parameters and coefficients for higher order experience terms are reported in Table 2. The estimates for η_1, η_2, and η_3 are all insignificant at the 5% level.[12] The estimates of the aggregate technology parameters remain very similar to those of our benchmark linear specification. This suggests the robustness of the linear specification for experience skill.

Given the insignificance of higher order experience terms, the presence of complementarity and the shape and slope of the life-cycle efficiency schedule is robust to the introduction of higher order experience terms. As is indicated in Appendix Figure A-3, the fit of the model remains virtually unchanged. The robustness of the life-cycle efficiency schedules confirms that the effective units of experience monotonically decrease over the life-cycle. We also find that the linear technology of experience production is supported by the data compared to specifications with curvature. This implies that the hump shape in the return to experience over the life-cycle is not driven by the decreasing returns in accumulating experience. Instead, it is driven by less efficient utilization of accumulated experience with age.

[12]Higher order experience terms remain insignificant in the quadratic and cubic specification of experience production technology.

Table 2: PSID estimates, with curvature in the experience production technology.

parameter	estimate	standard error	t-statistic
μ	-2.32	0.060	-38.73
δ	5.83	0.35	16.52
η_1	-0.0051	.0066	-0.78
η_2	0.0004	.00027	1.35
η_3	-5.33e-06	4.05e-06	-1.32
R^2	0.926		
$RMSE$	0.596		

Note - Entries for μ, δ and the goodness-of-fit represent the results of the second-stage estimation on PSID data of technology parameters of the specification that allows for curvature in the experience production technology. See Section 4.2.1 for the specification. Parameters η_1, η_2, and η_3 are estimated in the first stage. See Section 3 for details of the estimation procedure. For sample restrictions and variable construction procedures, see Appendix A1.1.

4.2.2 Common Life-Cycle Efficiency Schedules across Schooling Groups

In our benchmark specification, we allow the life-cycle efficiency schedules for labor and experience to differ between the two schooling groups. This flexible specification improves the fit of the model, and we did find significant differences in the life-cycle efficiency schedules between the two schooling groups. However, we find that existence of the strong complementarity and the explanatory power of our model for experience premium are preserved even under a less flexible specification where the life-cycle efficiency schedules are forced to be the same for both schooling groups, i.e., $\lambda_L(j,0) = \lambda_L(j,1)$ and $\lambda_E(j,0) = \lambda_E(j,1)$. Estimates for this common-life-cycle-efficiency-schedules specification are reported in Table 3. The estimated $\hat{\mu} = -4.49$, which implies stronger complementarity. Obviously, restricting the life-cycle efficiency schedules may change the magnitudes of the estimated experience premium and the experience-labor ratio. However, the estimated values of the experience premium (17% on average) and the experience-labor ratio (3.6 on average) are similar to those of the benchmark case. Given this, the stronger complementarity results in a substantially higher estimate for the scaling parameter at $\hat{\delta} = 195$ than before. However, this does not imply that the properties of the aggregate production function change substantially. One way to show this is to compare the labor shares

Table 3: PSID estimates, common life-cycle efficiency schedules across schooling groups.

parameter	estimate	standard error	t-statistic
μ	-4.49	0.121	-37.08
δ	195.49	30.14	6.49
$\lambda_{L,1}$	0.026	0.0012	22.23
$\lambda_{L,2}$	-0.00044	0.00003	-15.62
$\lambda_{E/L,1}$	-0.86222	0.0332	-25.97
$\lambda_{E/L,2}$	0.00095	0.000074	12.83
R^2	0.925		
$RMSE$	0.598		

Note - Entries for μ, δ and the goodness-of-fit represent the results of the second-stage estimation on PSID data of technology parameters of the specification that restricts labor and experience life-cycle efficiency schedules to be the same across schooling groups. See Section 4.2.2 for the specification. Parameters $\lambda_{L,1}$, $\lambda_{L,2}$, $\lambda_{E-L,1}$, and $\lambda_{E-L,2}$ are estimated in the first stage. See Section 3 for details of the estimation procedure. For sample restrictions and variable construction procedures, see Appendix A1.1.

of aggregate earnings between the two sets of estimates. The share of raw labor of aggregate earnings is now varying from 72% in 1968 to 57% in 1996, which mirrors the changes implied by the benchmark estimates. As shown in Appendix Figure A-4, the ability of the model to predict the experience premium from the experience-labor ratio remains very good under this restriction.

4.2.3 Constant Life-Cycle Efficiency of Labor

We now further restrict the specification by not only imposing the same life-cycle efficiency schedules across schooling groups but also assuming that only the efficiency of the experience skill varies over the life-cycle, such that each worker supplies one unit of effective labor throughout the life cycle. Thus, we estimate the following restricted log wage equation:

$$\ln w_{it} = \ln A_t + \ln G_{L_t} + \ln \left[1 + \delta \left(\frac{E_t}{L_t} \right)^{\mu-1} \exp(\lambda_{E,1} j_{it} + \lambda_{E,2} j_{it}^2)\, e_{it} \right] + \alpha_t \chi_{it} + \epsilon_{it}.$$

We experiment on this specification to evaluate to what extent the estimate of complementarity and explanatory power of the model depends on the concavity of the life-cycle profile of wages. In the first-stage estimation the model attempts to fit the life-cycle wage profile by the

Table 4: PSID estimates, specification with constant life-cycle efficiency of labor.

parameter	estimate	standard error	t-statistic
μ	-1.64	0.084	-19.46
δ	21.11	3.53	5.98
$\lambda_{E,1}$	-0.0244	0.0018	-13.78
$\lambda_{E,2}$	-0.00012	0.000034	-3.82
R^2	0.9245		
$RMSE$	0.6		

Note - Entries for μ, δ and the goodness-of-fit represent the results of the second-stage estimation on PSID data of technology parameters of the specification that imposes constant life-cycle efficiency of labor and restricts life-cycle efficiency schedules for experience to be the same across schooling groups. See Section 4.2.3 for the specification. Parameters $\lambda_{E,1}$ and $\lambda_{E,2}$ are estimated in the first stage. See Section 3 for details of the estimation procedure. For sample restrictions and variable construction procedures, see Appendix A1.1.

appropriate choice of the experience production technology parameters as well as the shapes of the life-cycle efficiency schedules for labor and experience. By dropping the life-cycle efficiency schedule of labor, we put the burden of matching life-cycle profiles on the life-cycle efficiency schedule of experience as well as the estimated experience premium.

Estimates for this specification are reported in Table 5. In this case, the estimated $\hat{\mu} = -1.64$. Taken together with the estimate of the scaling parameter $\hat{\delta} = 21.11$, these estimates again imply properties of the production function similar to those in our benchmark specification. Interestingly, the life-cycle efficiency schedule for experience still implies a monotonic decline in the efficiency units of experience over the life-cycle. Appendix Figure A-5 confirms that the co-movement between the experience premium and the experience-labor ratio remains strong and the model continues to predict the experience premium dynamics well. This is remarkable because estimates of both the experience premium and the experience-labor ratio are affected by the restriction imposed on the model.

4.2.4 Constant Life-Cycle Efficiency of Labor and Experience

Finally, we restrict the specification further by assuming that the efficiency of neither labor nor experience changes over the life-cycle. Thus, we estimate the following restricted log wage equation:

$$\ln w_{it} = \ln A_t + \ln G_{L_t} + \ln \left[1 + \delta \left(\frac{E_t}{L_t} \right)^{\mu - 1} e_{it} \right] + \alpha_t \chi_{it} + \epsilon_{it}.$$

Table 5: PSID estimates, specification with constant life-cycle efficiency of labor and experience.

parameter	estimate	standard error	t-statistic
μ	-1.77	0.111	-15.88
δ	59.65	18.00	3.31
R^2	0.922		
$RMSE$	0.61		

Note - Entries for μ, δ and the goodness-of-fit represent the results of the second-stage estimation on PSID data of technology parameters of the specification that does not include life-cycle efficiency schedules of labor and experience. See Section 4.2.4 for the specification. See Section 3 for details of the estimation procedure. For sample restrictions and variable construction procedures, see Appendix A1.1.

In this specification the model's ability to match the life-cycle earnings profiled is severely restricted. This implies a decline of the goodness-of-fit in the first-stage estimation (first-stage RMSE increases to 0.61 from 0.596 in the benchmark). It also affects the first-stage estimates of the experience premium and the experience-labor ratio. Remarkably, the second-stage estimate of complementarity between labor and experience remains robust. Estimates for this specification are reported in Table 5. The estimated $\hat{\mu} = -1.77$ and the scaling parameter $\hat{\delta} = 59.65$ once again imply properties of the production function fairly similar to those in our benchmark specification. Appendix Figure A-6 confirms the clear co-movement between the experience premium and the experience-labor ratio and the ability of the restricted model to predict the experience premium dynamics quite well.

4.3 Census Cross-section

We now re-estimate the model using *cross-sectional* variation of the experience premium and experience-labor ratio across the U.S. states. Thus, time-specific variables in equation (16) for the previous estimation now take on an interpretation of state-specific variables (resulting in changes in the subscript from time index t to state index v). Our two-step estimation procedure is applied to the following equation:

$$\begin{aligned}\ln w_{iv} &= \ln A_v + \ln G_{L_v} + \left(\lambda_{L,1}\left(s_{iv}\right) j_{iv} + \lambda_{L,2}\left(s_{iv}\right) j_{iv}^2\right) \\ &\quad + \ln\left[1 + \delta\left(\frac{E_v}{L_v}\right)^{\mu-1} \exp(\lambda_{E/L,1}\left(s_{iv}\right) j_{iv} + \lambda_{E/L,2}\left(s_{iv}\right) j_{iv}^2)\, e_{iv}\right] + \alpha_v \chi_{iv} + \epsilon_{iv}.\end{aligned} \tag{28}$$

We use the U.S. decennial census data rather than the PSID to have sufficient within-state samples, which helps us to better measure the cross-state variation of the experience premium and the experience-labor ratio. We use data from the 1980 census, which falls within the mid-range of our PSID sample period, and from where we can measure hourly wages and all the control variables in a way compatible with the PSID. The only exception is work experience. Unfortunately, census data do not contain data on actual work experience. To maximize the compatibility of the results between the time-series and cross-sectional estimation, we use the fact that we do observe the actual experience in PSID for the same census year 1980 and hence can predict actual experience from the observable variables both in the PSID and in the 1980 census. Thus, we impute actual experience for the 1980 census from the observable variables according to the prediction formula obtained from the PSID and use it in our cross-sectional estimation.[13]

The first-stage estimates for $\widehat{\Pi}_{E_v}$ and $\frac{\widehat{E}_v}{\widehat{L}_v}$ imply an average raw labor share of aggregate earnings of 60% across states, which compares well to the labor share implied by the time-series estimates. The estimates for the technology parameters using the census data are reported in Table 6. The cross-sectional estimate of the curvature parameter $\hat{\mu} = -1.58$ is similar to our time-series estimate from the PSID. We show in Appendix A2 that these estimates can indeed be compared to each other despite one being for the aggregate production function and the other for the state-level production function. Thus, the presence of complementarity between labor and experience is confirmed both in the time-series and cross-sectional data for the U.S. Moreover, the time-series and cross-sectional elasticities of substitution between labor and experience are similar, despite the differences in the identifying source of variation.

[13]We find a correlation coefficient between actual experience and predicted actual experience in the PSID of 0.91. The details of this estimation can be found in Appendix A1.2.

Table 6: Census cross-section estimates of technology parameters.

parameter	estimate	standard error	t-statistic
μ	-1.58	0.031	-51.64
δ	5.75	0.234	24.54
R^2	0.936		
$RMSE$	0.55		

Note - The entries represent the results of the second-stage estimation of technology parameters of the benchmark specification from the 1980 decennial census data. For sample restrictions and variable construction procedures, see Appendix A1.2. See Section 4.3 for the specification and Section 3 for details of the estimation procedure.

5 Conclusion

We investigate the importance of demographic change in driving the evolution of the return to experience. While the idea of an inverse relation between the relative supply of experience and the economic return it commands has a strong intuitive appeal, there has been no explicit quantitative evaluation of this relationship. This paper contributes by providing a theory that can rationalize this relationship and measure its importance.

The insight of the model is to note that it is not the stock of the experience that matters but its stock relative to the other productive inputs and to propose what these other inputs are. It appears natural to consider workers as supplying two productive services, i.e., "labor" and "experience skill" accumulated with work experience. These inputs constitute the arguments in the aggregate production function and are competitively priced in the market. If they are complementary, a change in the aggregate ratio of experience to labor would affect the marginal products of the two inputs. Demographic change, such as a sizable entry of inexperienced workers into the market, will increase the marginal product of experience. We identify parameters of the aggregate technology using disaggregate earnings data, by estimating individual earnings equations that consistently aggregate. Our estimates imply that labor and experience are indeed complementary, with an elasticity of substitution of 0.3. Moreover, the model with a very parsimonious specification of the aggregate production function and a single state variable, the experience-labor ratio, quantitatively accounts very well for the large changes in the experience premium over time. For example, we find a correlation coefficient between actual and predicted experience premia of 0.97 over the sample period 1968-1996. This is our main contribution.

By separately identifying the effects of age from the effects of experience on wages, we find that the hump shape in the return to experience over the life-cycle is not driven by the decreasing returns in accumulating experience. Instead, it is driven by a less efficient utilization of accumulated experience with age. This insight is not only important for model building but has a clear relevance for the design of labor market policies. Moreover, we find that the effective units of labor are also hump shaped over the life-cycle. Taken together, these findings imply that the concavity of the age-wage profile or the experience-wage profile (where experience is measured by age-based potential experience) in typical Mincerian regressions is due to age effects rather than decreasing returns in experience skill production.

On a methodological level, we show that it is possible to identify parameters of the aggregate production function from individual data by maintaining a consistency between individual earnings equations and the aggregate production function. Moreover, we show that estimation of the technology parameters using the microdata can identify the size and direction of the residual technical change.

Finally, our findings contribute to the literature that attempts to isolate age, time, and cohort effects in the earnings equation (e.g., Heathcote, Storesletten, and Violante (2005)). By construction, only two of these effects can be simultaneously identified, when the time effect is identified with the calendar date. We model the time and age effects, where the changes in demographic composition, summarized by the experience-labor ratio, determines a time effect via the experience premium and the age effect is captured by the life-cycle efficiency profiles. As entering cohort vary in size, distribution of age as well as the experience-labor ratio in the labor market and hence relative wages will change. Thus, we indirectly capture cohort size effects through the structure of the model.

References

Abowd, J. M., and F. Kramarz (1999): “The Analysis of Labor Markets using Matched Employer-Employee Data,” in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3B, chap. 26, pp. 2629–2710. Amsterdam: North-Holland.

Becker, G. (1962): “Investment in Human Capital: A Theoretical Analysis,” *Journal of Political Economy*, 70, 9–49.

——— (1993): *Human Capital*. University of Chicago Press, Chicago, Third edn.

Ben-Porath, Y. (1967): “The Production of Human Capital and the Life Cycle of Earnings,” *Journal of Political Economy*, 75(4), 352–365.

Berger, M. C. (1985): “The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence,” *Journal of Political Economy*, 93(3), 561–573.

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.

Fisher, J. D. M., and M. Gervais (2007): “First-Time Home Buyers and Residential Investment Volatility,” WP 2007-15, Federal Reserve Bank of Chicago.

Flinn, C. J. (1993a): “Cohort Size and Schooling Choice,” *Journal of Population Economics*, 6, 31–55.

——— (1993b): “The Implications of Cohort Size for Human Capital Investment,” in *Labour Markets in an Aging Europe*, ed. by P. Johnson, and K. F. Zimmermann, chap. 5. Cambridge: Cambridge University Press.

Fox, J. T. (2008): “Firm Size Wage Gaps, Job Responsibility and Hierarchical Matching,” mimeo, University of Chicago.

Freeman, R. B. (1979): "The Effect of Demographic Factors on Age-Earnings Profiles," *Journal of Human Resources*, 14(3), 289–318.

Gallipoli, G., C. Meghir, and G. L. Violante (2007): "Equilibrium Effects of Education Policies: a Quantitative Evaluation," mimeo, UBC, UCL, NYU.

Guvenen, F., and B. Kuruscu (2007a): "A Quantitative Analysis of the Evolution of the U.S. Wage Distribution: 1970-2000," NBER Working Paper 13095.

——— (2007b): "Understanding the Evolution of the U.S. Wage Distribution: A Theoretical Analysis," NBER Working Paper 13096.

Heathcote, J., K. Storesletten, and G. L. Violante (2005): "Two Views of Inequality Over the Life-Cycle," *Journal of the European Economic Association*, 3, 765–775.

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

Heckman, J. J., and J. A. Scheinkman (1987): "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *Review of Economic Studies*, 54(2), 243–55.

Jaimovich, N., and H. E. Siu (2007): "The Young, the Old, and the Restless: Demographics and Business Cycle Volatility," Research Department Staff Report 387, Federal Reserve Bank of Minneapolis.

Jeong, H., and Y. Kim (2006): "Complementarity and Transition to Modern Economic Growth," Institute for Economic Policy Research Working Paper Series 06.44.

Katz, L. F., and D. H. Autor (1999): "Changes in Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, ed. by O. C. Ashtenfelter, and D. Card, vol. 3A, pp. 1463–1555. Amsterdam: North-Holland.

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

KIM, D.-I., AND R. H. TOPEL (1995): "Labor Markets and Economic Growth: Lessons from Korea's Industrialization, 1970-1990," in *Differences and Changes in Wage Structures*, ed. by R. B. Freeman, and L. F. Katz, pp. 227–64. Chicago: University of Chicago Press for NBER.

KRUSELL, P., L. E. OHANIAN, J.-V. RÍOS-RULL, AND G. L. VIOLANTE (2000): "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 68, 1029–1053.

KWON, I., AND E. M. MEYERSSON-MILGROM (2004): "Boundaries of Internal Labor Markets: The Relative Importance of Firms and Occupations," mimeo, Stanford University Graduate School of Business.

MINCER, J. (1974): *Schooling, Experience, and Earnings*. Columbia University Press, New York, NY.

RUGGLES, S., M. SOBEK, T. ALEXANDER, C. A. FITCH, R. GOEKEN, P. KELLY HALL, M. KING, AND C. RONNANDER (2004): *Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]*. Minnesota Population Center [producer and distributor], Minneapolis, MN.

TOPEL, R. H. (1997): "Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality," *Journal of Economic Perspectives*, 11(2), 55–74.

TRIEST, R. K., M. SAPOZHNIKOV, AND S. A. SASS (2006): "Population Aging and the Structure of Wages," CRR WP 2006-5, Center for Retirement Research at Boston College.

WASMER, E. (2001): "Between-group Competition in the Labor Market and the Rising Returns to Skill: US and France 1964-2000," IZA Discussion Paper No. 292.

WELCH, F. (1979): "Effects of Cohort Size on Earnings: The Baby Boom Babies Financial Bust," *Journal of Political Economy*, 87(5), S65–S97.

APPENDICES

A1 Data Appendix

A1.1 PSID Data

Sample. We use the Panel Study of Income Dynamics (PSID) data from the U.S. for the period 1968-1997. 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.[14] The sample is restricted to individuals between 18 and 65 years of age.

Actual Labor Market Experience. The procedure we use to construct measures of actual work experience since age 18 is as follows. Questions regarding overall labor market experience ("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.[15] These questions are also asked for every person in the year when that person first becomes a household head or wife.[16] Since there are some inconsistencies between the answers, we first adjust 1974 overall experience to be consistent with 1975 and 1976 values where 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 experience they report in their first year in the sample is true. If reported experience 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 experience 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 repeat this procedure for 1985 and the reports of the new heads and wives. Finally, using the values of experience in 1974, 1985, and the reports of the new heads and wives, we increment experience variables forward and backward as follows: increment the full-time experience measure by one if individual works at least 1500 hours in a given year.[17] If

[14] 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 1968-1996 period.

[15] 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).

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

[17] We experimented with using cutoff values other than 1500 hours of work or using directly the sum of

we observe an individual in the sample since age 18, we ignore his or her reports of experience and instead directly use his or her reports of hours in each year using the cutoff above. If we do not observe an individual in the sample since age 18, we attempt to construct tenure based on the earliest report of his or her overall experience.

Other Variables. Our hourly wage measure is equal to the total earnings last year divided by total hours worked last year. To get real wage, we adjust the nominal wage using last year's CPI (equal to 100 in 1984).[18] We define the economically active population as the group of people who worked at least 700 hours last year.[19] Education is measured by years of final educational attainment.[20] Other control variables that we will use are gender (male dummy), race (black dummy), and region (Northeast, North Central (i.e., Midwest) and West dummies). The broad region variable is created using the state variable in the PSID.[21] South is the (poorest) base category region.

A1.2 U.S. Census Data

Sample. We use U.S. census data from the 1980 U.S. Census 5% Public Use Micro-samples (see Ruggles, Sobek, Alexander, Fitch, Goeken, Kelly Hall, King, and Ronnander (2004)). The sample represents a 1-in-20 national random sample of the population, with approximately 11,337,000 person records. Sample restrictions and variable definitions are imposed to ensure maximum comparability with the PSID. The sample is restricted to individuals between 18 and 65 years of age with nonimputed values of all the variables used in the analysis. The economically active population is defined as those who worked at least 700 hours last year.

---

accumulated hours of work to create other measures of experience and found that our chosen measure shows the smoothest pattern of movements. The results are not sensitive to this choice.

[18] 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.

[19] As in the case of earnings, there is also an employment status variable at the time of the interview. We do not use tis variable because (1) the reference period (current year) is different from that of 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.

[20] 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 wifes. 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.

[21] 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.

Actual Labor Market Experience. Unfortunately, census data do not contain information on actual worker experience. We exploit the relationship between measures of actual experience in the PSID and worker characteristics to impute actual experience to workers in the census data. In particular, using the 1980 PSID data, we regress actual experience on a polynomial function of age for each subgroup and household demographic variables, such that

$$e_i = \sum_{k=1}^{K} \left(\alpha_{k0} + \alpha_{k1} j_i + \alpha_{k2} j_i^2 + \alpha_{k3t} j_i^3 + \alpha_{k4t} j_i^4\right) d_{k,i} + \beta\chi_i + \zeta_i \tag{A1}$$

to obtain the predicted experience $\widehat{e}_i$ such that

$$\widehat{e}_i = \sum_{k=1}^{K} \left(\widehat{\alpha}_{k0t} + \widehat{\alpha}_{k1} j_i + \widehat{\alpha}_{k2} j_i^2 + \widehat{\alpha}_{k3} j_i^3 + \widehat{\alpha}_{k4} j_i^4\right) d_{k,i} + \widehat{\beta}\chi_i \tag{A2}$$

where j_i denotes age, $d_{k,i}$ the dummy variable for subgroup k, χ_i the household demographic variables, and ζ_i the *i.i.d.* idiosyncratic shock to work experience. The characteristics partitioning the subgroups are college education, household headship, four regions (Northeast, North Central, South and West), race, marital status and gender. We include the number of children in the household and its squared value for χ_{it}. The estimated coefficients are reported in Appendix Table A-1. In the PSID, the correlation between actual and predicted experience is equal to 0.91. Applying (A2), we construct predicted levels of actual experience $\widehat{e}_i$ using the individual observable characteristics from the 1980 census.

Table A-1: Coefficient estimates for predicted experience.

Variable	dummy	age	age-squared	age-cubic	age-quartic
age		.5083873	.0483154	-.0015294	7.19e-06
college	.6203082	-.6386097	.059012	-.0019724	.0000214
head	-.4141898	.2635762	-.0382591	.0015935	-.0000127
northeast	.2163412	-.1804537	.0238301	-.001135	.000016
north central	.2091296	-.1472149	.0195273	-.0009212	.0000128
west	-.1581706	.1143491	-.0155347	.0004851	-3.77e-06
black	-.4370619	.0386038	-.0033227	.0001884	-2.33e-06
unmarried male	-.0216989	-.0237478	-.0288911	.0018001	-.0000271
unmarried female	.4193616	-.3648575	.0162269	-.0009843	.0000137
married female	-.3881132	.2097724	-.0583758	.0013988	-1.64e-06
children	-.5025105				
children squared	.0431225				
constant	.5184911				

Other Variables. Education is measured as the number of years of school completed, topcoded at 16. Earnings are measured as total pre-tax wage and salary income for the preceding calendar year. The codes represent the midpoints of ten-dollar intervals. Earnings are topcoded at \$75,000. To obtain hourly earnings we divide the total earnings last year by the product of weeks worked last year and usual hours worked per week last year. Other control variables that we use are gender (male dummy), race (black dummy), and region categorized consistently with the PSID based on the state of residence.

A1.3 SAF Data

The data were collected and compiled by the Swedish Employers' Confederation (SAF) from its database on wage statistics assembled from establishment-level personnel records. The data contain information on roughly 60% of the Swedish private-sector workforce. Included in the data are blue- and white-collar workers in every industry (except insurance and banking) in the private sector from 1970 to 1990. The blue- and white-collar worker samples are separate and cannot be combined to produce a representative combined sample. We report the results based on the white-collar sample only. The results based on the blue-collar sample are similar and are available upon request. Labor market experience is measured as potential experience. These data are used in the yearly wage negotiations and are monitored not only by the SAF but also by the labor unions. Data quality is considered to be exceptionally high. See Kwon and Meyersson-Milgrom (2004) and Fox (2008) for a more detailed description of the data and Swedish labor market conditions.

A1.4 Danish Data

The data from Denmark are based on administrative files covering 100% of the Danish population for the years 1981 to 2003. It is administrative register data from the Integrated Database for Labor Market Research (IDA) and the Income Registers, which are both longitudinal databases with annual observations for individuals. Wages are extracted from the Income Registers and include all wage earners working in every industry (both public and private). Socioeconomic variables and experience come from the Integrated Database for Labor Market Research (IDA). Information on actual work experience after 1964 is available for everyone in the data set. For those who entered the labor market before 1964 their experience in that year is measured as potential experience (age minus education minus 6). After that it is incremented using observed actual experience. The quality and the validity of the IDA data are regarded as high (see Abowd and Kramarz (1999)).

A1.5 First-Stage Estimation Results from the PSID

Table A-2: PSID first-stage estimates, time-invariant parameters.

	Estimate	Standard Error
northeast	.18452	.00438
north central	.05052	.00412
west	.09726	.00469
$\lambda_{L,1}\,(0)$	.02008	.00128
$\lambda_{L,2}\,(0)$	-.00036	.00003
$\lambda_{L,1}\,(1)$	.04057	.00148
$\lambda_{L,2}\,(1)$	-.00065	.00004
$\lambda_{E-L,1}\,(0)$	-.07787	.00340
$\lambda_{E-L,2}\,(0)$	.00081	.00008
$\lambda_{E-L,1}\,(1)$	-.12513	.00478
$\lambda_{E-L,2}\,(1)$	.00154	.00013
R^2	0.9258	
$RMSE$	0.5955	

Note - The entries represent the results of the first-stage estimation of time-invariant parameters of the benchmark specification from the PSID. For sample restrictions and variable construction procedures, see Appendix A1.1. See Section 3 for details of the estimation procedure.

Table A-3: PSID first-stage estimates, time-varying parameters.

Year	Male	Black	Schooling	Exp. premium	Constant
1968	.32327 (.02787)	-.18287 (.04361)	.07320 (.00506)	.10161 (.01345)	.75483 (.05298)
1969	.32086 (.02488)	-.16414 (.03946)	.06775 (.00465)	.10044 (.01186)	.82818 (.04735)
1970	.31857 (.02418)	-.13311 (.03766)	.06679 (.00456)	.10931 (.01192)	.82679 (.04609)
1971	.28301 (.02385)	-.10426 (.03757)	.07323 (.00449)	.11741 (.01219)	.79148 (.04571)
1972	.27196 (.02308)	-.12598 (.03634)	.07586 (.00443)	.12352 (.01210)	.78455 (.04505)
1973	.31703 (.02257)	-.08361 (.03574)	.07323 (.00435)	.11788 (.01156)	.81554 (.04389)
1974	.30313 (.02036)	-.10620 (.03217)	.06471 (.00406)	.11764 (.01051)	.89162 (.04056)
1975	.28381 (.02038)	-.08534 (.03177)	.06764 (.00414)	.12265 (.01096)	.85795 (.04142)
1976	.23595 (.02046)	-.04879 (.03194)	.07093 (.00416)	.13952 (.01168)	.81198 (.04149)
1977	.26641 (.01998)	-.03969 (.03110)	.07592 (.00410)	.13599 (.01132)	.77157 (.04039)
1978	.24862 (.01979)	-.05903 (.03110)	.07664 (.00412)	.17006 (.01293)	.71333 (.04133)
1979	.26996 (.01854)	-.08691 (.02804)	.07386 (.00390)	.16582 (.01197)	.73624 (.03846)
1980	.26920 (.01839)	-.05236 (.02770)	.06585 (.00399)	.14980 (.01126)	.80149 (.03883)
1981	.24269 (.01856)	-.08848 (.02828)	.07885 (.00401)	.17996 (.01295)	.62780 (.03989)
1982	.22503 (.01862)	-.08382 (.02842)	.07433 (.00407)	.17712 (.01315)	.64944 (.04096)
1983	.16874 (.01854)	-.06402 (.02883)	.08543 (.00414)	.22084 (.01532)	.48736 (.04228)
1984	.16126 (.01738)	-.07211 (.02660)	.08580 (.00401)	.21729 (.01419)	.51520 (.04031)
1985	.22537 (.01713)	-.07033 (.02615)	.09376 (.00397)	.22509 (.01436)	.40447 (.03963)
1986	.20594 (.01696)	-.10794 (.02614)	.09276 (.00408)	.21326 (.01415)	.45178 (.04128)
1987	.18265 (.01687)	-.13030 (.02610)	.09901 (.00408)	.22388 (.01452)	.42192 (.04132)
1988	.13757 (.01682)	-.12790 (.02578)	.10712 (.00413)	.24552 (.01547)	.35109 (.04164)
1989	.16448 (.01577)	-.12214 (.02339)	.10726 (.00392)	.22844 (.01395)	.35821 (.03910)
1990	.13416 (.01587)	-.09420 (.02338)	.11278 (.00401)	.23144 (.01445)	.31307 (.04041)
1991	.12592 (.01584)	-.06615 (.02401)	.11998 (.00404)	.21916 (.01403)	.26205 (.04077)
1992	.12921 (.01574)	-.14385 (.02381)	.11758 (.00411)	.23110 (.01456)	.24268 (.04145)
1993	.10802 (.01552)	-.08157 (.02396)	.11625 (.00410)	.25555 (.01545)	.24155 (.04110)
1994	.17647 (.01646)	-.15186 (.02419)	.12557 (.00439)	.21578 (.01577)	.21333 (.04642)
1995	.16136 (.01618)	-.12000 (.02364)	.11307 (.00432)	.24281 (.01737)	.26023 (.04705)
1996	.17025 (.01626)	-.13694 (.02391)	.10611 (.00435)	.22821 (.01705)	.32920 (.04781)

Note - The entries represent the results of the first-stage estimation of time-varying parameters of the benchmark specification from the PSID. Standard errors are in parenthesis. For sample restrictions and variable construction procedures, see Appendix A1.1. See Section 3 for details of the estimation procedure.

A2 Aggregation of State-Level Production Functions

Consider a common state-level production function

$$Y_v = AG(L_v, \widetilde{E}_v) = A\left(L_v^{\mu} + \delta \widetilde{E}_v^{\mu}\right)^{\frac{1}{\mu}}. \tag{A3}$$

We allow the effective workforce size L_v to vary across states to capture the difference in state size. The effective state-level experience $\widetilde{E}_v$ also varies across states such that $\widetilde{E}_v = \epsilon_v E_v$, where E_v is the deterministic component and ϵ_v represents an ex-post *i.i.d* shock over states with mean of unity. This formulation allows for a variation of experience-labor ratio across states despite the common technology.

The aggregate production function $F(L, E)$, defined on aggregate labor L and aggregate experience E, can be characterized by the following planning problem to maximize the expected sum of state outputs subject to the factor feasibility conditions:

$$\mathbb{E}\left\{F(L,E)\right\} = \mathbb{E}\left\{\max_z \sum_{v=1}^{N} z_v Y_v\right\} \text{ subject to}$$

$$\sum_{v=1}^{N} z_v L_v = L, \tag{A4}$$

$$\mathbb{E}\left\{\sum_{v=1}^{N} z_v \widetilde{E}_v\right\} = E, \tag{A5}$$

where the expectation operator $\mathbb{E}$ is for the ex-post shock ϵ_v, and $z = (z_v)_{v=1}^{N}$ denotes the production plan. Given the common technology and the ex-post and *i.i.d.* nature of the shock ϵ_v, the optimal production plan adjusts only the difference in state size (measured by L_v) such that $z_v^* = \frac{1}{N}\frac{1}{p_v}$, where $p_v = \frac{L_v}{L}$, and the *ex-ante* experience-labor ratios $\frac{E_v}{L_v}$ are equalized across states to the aggregate experience-labor ratio $\frac{E}{L}$.[22]

[22] The feasibility conditions (A4) and (A5) for this production plan can be verified as follows:

$$\begin{aligned}
\sum_{v=1}^{N} z_v^* L_v &= \sum_{v=1}^{N} \frac{1}{N}\frac{1}{p_v} L_v = \frac{1}{N}\sum_{v=1}^{N} \frac{L}{L_v} L_v = L \\
\mathbb{E}\left\{\sum_{v=1}^{N} z_v^* \widetilde{E}_v\right\} &= \sum_{v=1}^{N} \frac{1}{N}\frac{1}{p_v}\mathbb{E}\left\{\widetilde{E}_v\right\} = \frac{1}{N}\sum_{v=1}^{N}\frac{L}{L_v}\mathbb{E}\left\{\widetilde{E}_v\right\} = \frac{1}{N}\sum_{v=1}^{N} L\mathbb{E}\left\{\frac{\widetilde{E}_v}{L_v}\right\} \\
&= \frac{1}{N}\sum_{v=1}^{N} L\mathbb{E}\left\{\frac{\epsilon_v E_v}{L_v}\right\} = \frac{1}{N}\sum_{v=1}^{N} L\left(\frac{E_v}{L_v}\right)\mathbb{E}\left\{\epsilon_v\right\} = \frac{1}{N}\sum_{v=1}^{N} L\left(\frac{E}{L}\right) = E.
\end{aligned}$$

Thus, the aggregate production function implied by the state production function (A3) is

$$\begin{aligned} F(L,E) &= \sum_{v=1}^{N} z_v^* Y_v = \frac{1}{N}\sum_{v=1}^{N} A\frac{L}{L_v}G(L_v,\epsilon_v E_v) = \frac{1}{N}\sum_{v=1}^{N} AG\left(\frac{L}{L_v}L_v, \epsilon_v\frac{E_v}{L_v}L\right) \\ &= \frac{1}{N}\sum_{v=1}^{N} AG\left(L,\epsilon_v\frac{E}{L}L\right) = \frac{1}{N}\sum_{v=1}^{N} AG\left(L,\epsilon_v E\right) = \frac{1}{N}\sum_{v=1}^{N} A\left(L^\mu + \delta_v E^\mu\right)^{\frac{1}{\mu}}, \end{aligned}$$

where

$$\delta_v = \delta\left(\epsilon_v\right)^\mu .$$

The marginal rate of technical substitution of the implied aggregate production function is

$$MRTS_{L,E} \equiv \frac{F_L}{F_E} = \left(\frac{E}{L}\right)^{1-\mu}\frac{f\left(\frac{E}{L}\right)}{g\left(\frac{E}{L}\right)}, \tag{A6}$$

where

$$\begin{aligned} f(\frac{E}{L}) &= \sum_{v=1}^{N}\left[1+\delta_v\left(\frac{E}{L}\right)^\mu\right]^{\frac{1}{\mu}-1}, \\ g(\frac{E}{L}) &= \sum_{v=1}^{N}\delta_v\left[1+\delta_v\left(\frac{E}{L}\right)^\mu\right]^{\frac{1}{\mu}-1}, \end{aligned}$$

and the elasticity of substitution between experience and labor σ is

$$\sigma \equiv \frac{d(E/L)}{dMRTS_{L,E}}\frac{MRTS_{L,E}}{(E/L)} = \frac{1}{\left[1-\mu+\eta\left(\frac{E}{L}\right)\right]}, \tag{A7}$$

where

$$\eta\left(\frac{E}{L}\right) = \left[\frac{f'\left(\frac{E}{L}\right)}{f\left(\frac{E}{L}\right)} - \frac{g'\left(\frac{E}{L}\right)}{g\left(\frac{E}{L}\right)}\right]\frac{E}{L}.$$

The elasticity of substitution between experience and labor implied by the census cross-sectional estimates of the parameters ($\delta = 5.75$ and $\mu = -1.58$) of the state production function in (A3) is 0.3876. We can calculate the elasticity of substitution between experience and labor of the aggregate production function according to the formula in equation (A7), using the aggregate and the state-level experience-labor ratios data. At the cross-sectional estimates above, we obtain an estimate of the *aggregate* elasticity of substitution between experience and labor of 0.3878, which is virtually the same as the one implied by the estimate of the curvature parameter of the state production function $\mu = -1.58$.

Figure A-1: Actual and Predicted Returns to Experience, Average across Workers

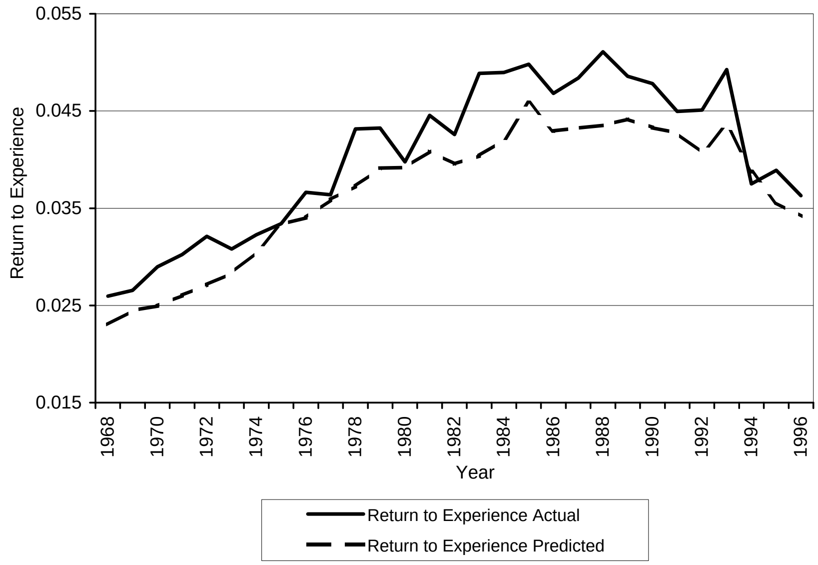


Figure A-2: Actual Return to Experience and Counterfactual with Constant Experience Premium, Average across Workers

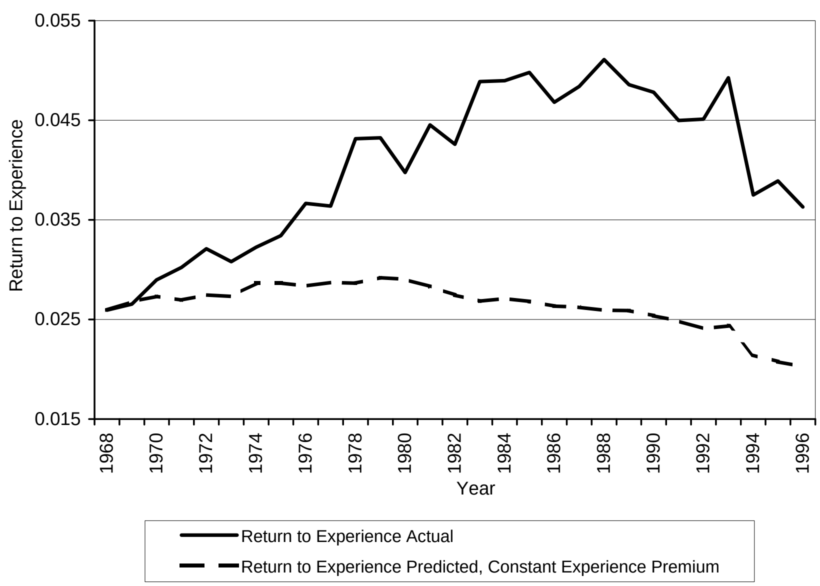


Figure A-3: Actual and Predicted Experience Premium and Experience to Labor Ratio, Quartic Specification for Experience Production Technology

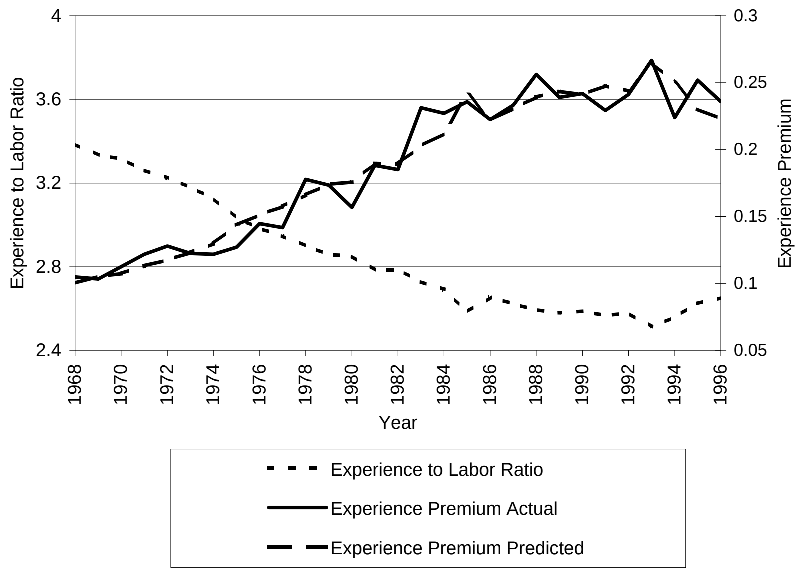


Figure A-4: Actual and Predicted Experience Premium and Experience to Labor Ratio, Common Life-Cycle Efficiency Schedules across Schooling Groups

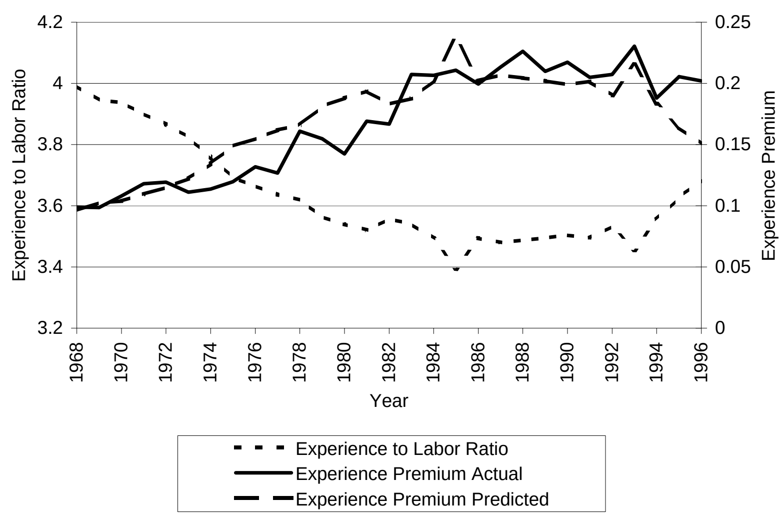


Figure A-5: Actual and Predicted Experience Premium and Experience to Labor Ratio, Common Life-Cycle Efficiency Schedules across Schooling Groups and Constant Life-Cycle Efficiency Schedule of Labor

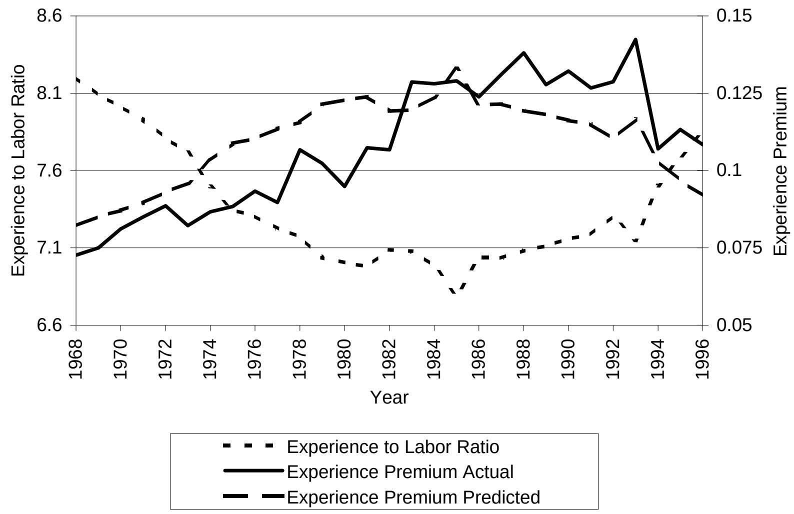


Figure A-6: Actual and Predicted Experience Premium and Experience to Labor Ratio, Constant Life-Cycle Efficiency of Labor and Experience