The Ben-Porath Model and Age-wage Profiles*

Lutz Hendricks†

January 3, 2013

Abstract

This paper asks how far one can go towards identifying the parameters of the Ben-Porath model without imposing strong assumptions about cohort endowments and skill prices. The main finding is that the Ben-Porath model generates nearly identical age-wage profiles for a wide range of two key parameters: the curvature of the human capital technology and the average growth rate of skill prices. Modeling school choice facilitates tighter identification of these model parameters. In contrast to a recent literature which argues that the Ben-Porath technology is nearly linear, my approach implies curvature parameters near 0.6 and human capital depreciation rates near 0.05. With these parameters, a number of empirical implications that have been derived from the Ben-Porath model are substantially modified.

JEL: E24, J24 (human capital), I21 (analysis of education).

Key words: Ben-Porath model. Human capital. Job-training.

*For helpful comments I am grateful to Oksana Leukhina. The usual disclaimer applies.

†University of North Carolina, Chapel Hill; lutz@lhendricks.org

1 Introduction

The Ben-Porath model is commonly used to study wage growth over the life-cycle. Its central equation describes a technology by which a worker can accumulate human capital h_t by investing time l_t:

$$h_{t+1} = (1 - \delta)h_t + \theta(h_t l_t)^\alpha \tag{1}$$

Here $\delta \in [0, 1)$ is the depreciation rate of human capital, $\theta > 0$ is a cohort specific productivity parameter, $\alpha \in (0, 1)$ is a curvature parameter, and t denotes age. Recent applications of the Ben-Porath model cover a wide range of questions, such as the sources of cross-country income differences (Manuelli and Seshadri, 2010), the evolution of wage inequality (Heckman, Lochner, and Taber, 1998) and of skill prices (Bowlus and Robinson, 2012), the sources of lifetime earnings inequality (Huggett, Ventura, and Yaron, 2006, 2011), and the effect of training on lifetime earnings (Kuruscu, 2006a).

Since the Ben-Porath model is often used to answer quantitative questions, there have been numerous attempts at estimating its parameters by fitting the model to observed age-wage profiles. An early literature yielded diverse parameter estimates.[1] A more recent literature follows Heckman, Lochner, and Taber (1998) and estimates the Ben-Porath model using panel data.[2] The resulting estimates are quite similar and have led researchers to conclude that the Ben-Porath technology is nearly linear with a human capital depreciation rate of 0. For example, Guvenen and Kuruscu (2010, p. 244) summarize the evidence as: "estimates of α – the curvature of the human capital accumulation function – typically vary between 0.80 and 0.95." Manuelli and Seshadri (2010, p. 15) argue that "more recent estimates ... are around 0.93." I label this view the "*emerging consensus*."

The near linear technology is central for a number of findings in the recent literature:

1. Manuelli and Seshadri (2010) argue that human capital accounts for a large share of cross-country income differences. In their theory, when α is close to 1, small exogenous differences in total factor productivity induce large differences in human capital.

2. Heckman, Lochner, and Taber (1998) develop a flat spot method for estimating skill prices, which is motivated by a Ben-Porath model with a near linear technology and

[1] See Browning, Hansen, and Heckman (1999) for a survey and Section 1.1 for additional detail.

[2] Taber (2002) and Kuruscu (2006a) follow Heckman, Lochner, and Taber (1998) in their estimation methods and in using NLSY data. Manuelli and Seshadri (2010) use PSID data.

no depreciation. Recently, Bowlus and Robinson (2012) have applied this method to U.S. data, arriving at the striking conclusion that the entire increase in the college wage premium since 1980 has been due to changes in human capital rather than wage rates.

3. Kuruscu (2006a) finds that training has negligible effects on lifetime earnings.

All of these findings depend on the Ben-Porath technology being close to linear (see Section 4 for details). Other quantitative applications of Ben-Porath models also employ near linear technologies. Examples include Heckman, Lochner, and Taber (1998) and Guvenen and Kuruscu (2010), who study the evolution of the U.S. wage distribution, and Taber (2002) who studies the effects of progressive taxes on human capital investment. In these cases it is not known whether the findings depend on α being close to 1.[3]

The question: All existing estimates of the Ben-Porath model use a limited amount of wage data, often covering only a single cohort or a single year, and must therefore impose strong assumptions on skill prices or cohort effects to achieve identification (see Section 1.1 for details). Notably, most recent studies closely follow Heckman, Lochner, and Taber (1998) in using NLSY data, which cover men born around 1960 who are at most 39 years old (and even younger in the earlier studies). With data that cover only partial wage profiles for a narrow range of birth cohorts, skill prices cannot be identified. Heckman, Lochner, and Taber (1998) and Kuruscu (2006a) therefore assume that skill prices are constant over time.[4] Taber (2002) experiments with alternative assumptions about expected skill prices and finds some of the model parameters sensitive to these assumptions.[5] Lacking wage data for older workers, all of these studies assume that human capital does not depreciate. This

[3] A partial exception is Guvenen and Kuruscu (2010) who show that their findings are qualitatively similar for α as low as 0.6. However, some of their results are reversed when α is below a threshold, the value of which is not known (cf. p. 244).

[4] There is, of course, a good reason why Heckman, Lochner, and Taber (1998) chose NLSY data. Estimating their model requires a proxy for individuals' cognitive abilities, which is not available is most datasets. Heckman, Lochner, and Taber (1998) also check whether their estimates are robust when the assumption of constant skill prices is relaxed. This is done by feeding the equilibrium skill prices implied by their general equilibrium model into the household problem. How this affects the estimated technology parameters is not reported.

[5] Manuelli and Seshadri (2010) do not explain how they estimate age-wage profiles from PSID data.

is motivated by the observation that wage profiles do not decline with age for older workers who are presumed to not invest in human capital.

The purpose of this paper is to estimate a Ben-Porath model without imposing strong assumptions on cohort effects, skill prices, and human capital depreciation. Specifically, the paper addresses the following questions:

1. How far can one go towards identifying the parameters of the Ben-Porath model when only weak restrictions are imposed on skill prices and cohort endowments?

2. Is the emerging consensus of a near linear human capital technology and no depreciation robust?

3. Are the implications of models with near linear human capital technologies robust?

Relaxing auxiliary assumptions about skill prices and cohort effects is desirable for a number of reasons:

1. The assumption of constant skill prices potentially conflicts with a large literature that documents movements in the relative wages of skilled and unskilled workers, and in particular a large increase in the college wage premium since about 1980 (see Goldin and Katz 2008 for a summary and further references).

2. The expansion of U.S. educational attainment during the 20th century suggests that the relative human capital endowments of different school groups may have changed over time (see Laitner 2000; Bowlus and Robinson 2012; Hendricks and Schoellman 2011).

3. The assumption that human capital does not depreciate is not based on an explicit model, but on visual inspection of empirical wage profiles (see Section 2.8 for details).

Approach. Following previous work, I estimate the model parameters by fitting the model to observed age profiles for mean log wages. The details of the model are described in Section 2. I depart from previous work in two ways. First, I allow cohort endowments (h_1), productivities (θ), and skill prices to vary over time in a flexible way. Second, I use a much larger set of wage observations. Specifically, in Section 2.3 I construct age-wage profiles for 12 cohorts of men born between 1935 and 1968 and observed in the Current

Population Survey (CPS) between 1964 and 2011. I distinguish four school groups: high school dropouts (HSD), high school graduates (HSG), college dropouts (CD), and college graduates (CG).

The role of two model parameters is of particular interest. The first is the curvature parameter α, which is central for a number of the model's quantitative applications (see Section 4). The second parameter of interest is the average growth rate of skill prices over the sample period, which I denote by Δw. Understanding the evolution of skill prices has been the focus of a large literature, much of it focused on skill-biased technical change (e.g., Berman, Bound, and Griliches 1994; Berman, Bound, and Machin 1998). Estimating Δw using the Ben-Porath model is the purpose of the flat spot method proposed by Heckman, Lochner, and Taber (1998). Since I am interested in the identification of both parameters, I calibrate the model on a grid of fixed values of α and Δw (Section 2.4).

Results. The results, reported in Section 2.5, are essentially the opposite of the emerging consensus. For values of α below 0.8, the model closely matches the observed age-wage profiles for all cohorts and school groups. The implied human capital depreciation rates are positive, at least for low values of α. However, the model fit deteriorates markedly as α rises above 0.8 and enters the range emphasized by the emerging consensus. One reason is that the wages of older workers comove almost perfectly in models with near linear technologies, which is at odds with the data.

A similar finding is obtained for the growth rate of skill prices. For negative values of Δw, the model closely matches observed age-wage profiles, but the model fit deteriorates as Δw turns positive.

Without imposing strong assumptions on the human capital endowments of different cohorts, identification of α and Δw is weak. The model generates nearly identical age-wage profiles over a wide range of $(\alpha, \Delta w)$ combinations. This point is illustrated in Figure 1, which shows age-wage profiles for college graduates for four birth cohorts. In addition to the empirical wage profiles, the Figure shows the model wage profiles for $\alpha \in \{0.3, 0.6\}$ and $\Delta w \in \{-2\%, 0\}$. The four model profiles are visually hard to distinguish. Similar results are obtained for the other school groups.

These results suggest that estimating the parameters of the Ben-Porath technology or the evolution of skill prices solely based on mean age-wage profiles is not promising. Section 3

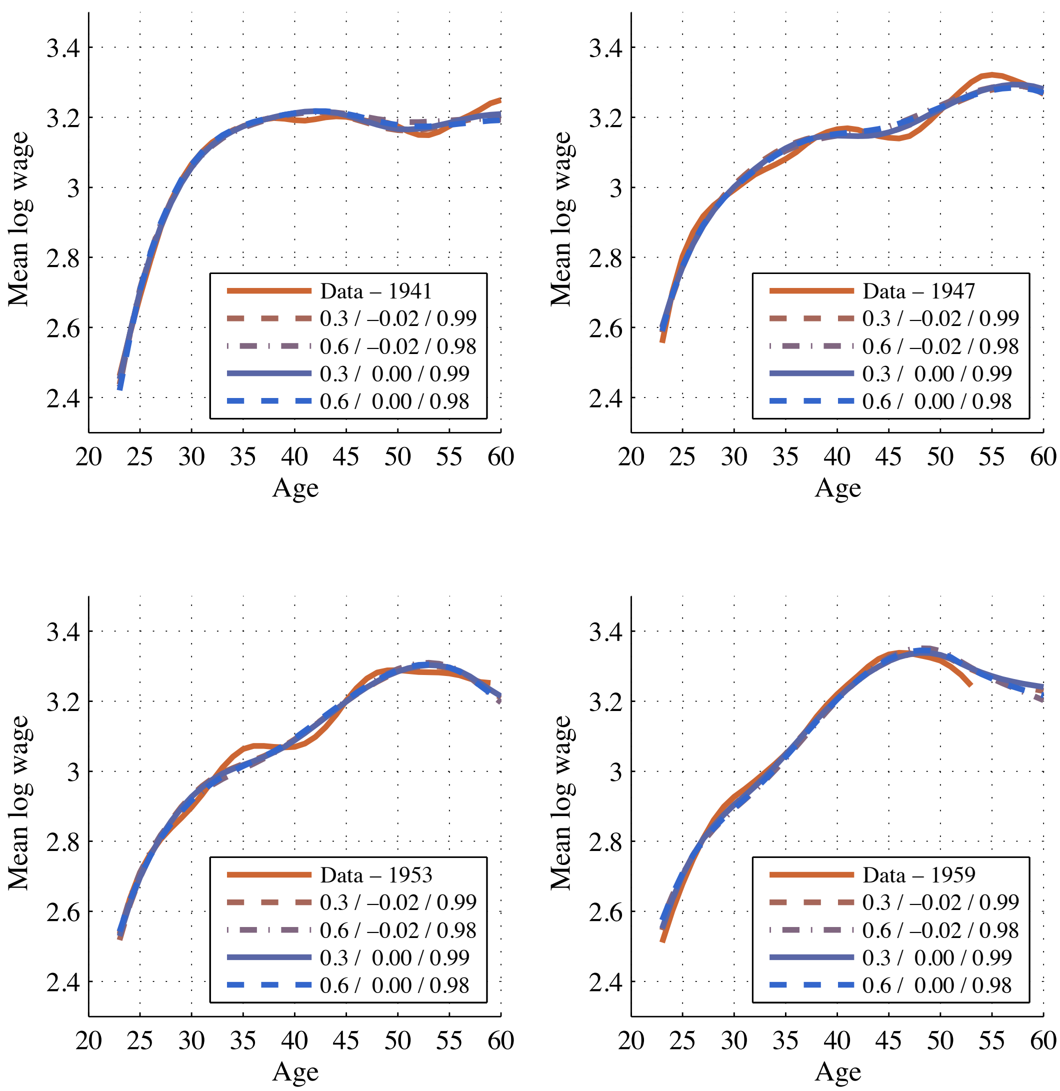


Figure 1: Age-wage profiles for college graduates

Notes: Each panel shows one cohort. Each legend entry shows the fixed values of the curvature parameter α and the average skill price growth rate Δw used in the calibration, as well as a pseudo-R^2 measure of fit.

examines how much progress can be made by imposing restrictions on cohort endowments (h_1 and θ). I develop a model of school choice which implies that a cohort's mean learning ability and human capital endowment are a function of educational attainment. These restrictions help to identify the key model parameters more precisely. The calibrated model implies curvature parameters near $\alpha = 0.6$. The calibrated skill price growth rates are near zero, but not precisely identified. Human capital depreciations rates take on values near 5%. Parameter values consistent with the emerging consensus ($\alpha_s \geq 0.8$ and $\delta_s < 0.01$) fit the data quite poorly.

Section 4 re-examines a number of empirical applications of the Ben-Porath model in light of these findings. The first application is the flat spot method. Its idea is to estimate the growth rate of skill prices from the growth rate of wages, following each cohort over an age range when human capital investment is likely small. Since the Ben-Porath model generates very similar age-wage profiles for a wide range of skill price growth rates, the flat spot method contains essentially no information about skill prices. Except for near linear human capital technologies, I show that both of its assumptions are violated: the depreciation rate of capital is not zero and human capital investment does not cease until very close to retirement.

The second application studies the effect of job training on lifetime earnings. Kuruscu (2006a) develops an ingenious argument why these effects must be small. Section 4.2 shows that this conclusion only holds for values of α near 1. For low values of α, training increases lifetime earnings by up to 65 log points, compared with about 1 log point as estimated by Kuruscu.

1.1 Related Literature

Browning, Hansen, and Heckman (1999) summarize existing estimates of Ben-Porath models. A set of papers from the 1970s imposed strong assumptions, partly due to data limitations. Heckman (1976) uses a single cross section dataset, the 1960 Census, which forces him to abstract from cohort effects. Brown (1976) uses a short panel dataset which covers 323 young men. Haley (1976) relies on 6 years of cross-sectional total income (rather than wage) data, which forces him to impose fixed linear time effects. The parameter estimates vary greatly between these studies.

The more recent literature has access to better data, but assumes that human capital does not depreciate. The work by Heckman, Lochner, and Taber (1998) and Kuruscu (2006a) was discussed earlier. Taber (2002) studies an extension of Heckman, Lochner, and Taber (1998) with progressive taxes. He finds that some of the Ben-Porath technology parameters are sensitive to assumptions about future skill prices. Manuelli and Seshadri (2010) use PSID data to measure the growth of wages between the ages of 25 and 55. Their treatment of time and cohort effects is not described. From this single data point, they estimate $\alpha = 0.936$ (this is the sum of the exponents on time and goods inputs).

Guvenen and Kuruscu (2010) study a variation of the Ben-Porath model where individuals supply raw labor and human capital to the market. Skill biased technical change raises the relative price of human capital. This model can replicate the evolution of the college wage premium and of wage inequality statistics constructed from CPS data. Huggett, Ventura, and Yaron (2011) use measures of earnings dispersion to calibrate a stochastic Ben-Porath model which yields an estimated curvature of $\alpha = 0.7$.

2 A Model with Unrestricted Cohort Endowments

This section studies how far one can go towards identifying the parameters of the Ben-Porath model imposing only weak restrictions cohort endowments and skill prices.

2.1 The Environment

I study a partial equilibrium Ben-Porath model with four school groups, indexed by s. In the data, these will be identified with high school dropouts (HSD), high school graduates (HSG), college dropouts (CD), and college graduates (CG).

Time is discrete and indexed by τ. The representative agent in cohort c and school group s is endowed with exogenous human capital $h_{s,c,1}$ and learning ability $\theta_{s,c}$. He works and trains for D_s periods. The objective is to maximize the discounted present value of lifetime earnings, which is given by

$$Y_{s,c} = \sum_{t=1}^{D_s} R^{-t} Y_{s,c,t} \tag{2}$$

where R is the time invariant gross interest rate and

$$Y_{s,c,t} = w_{s,\tau(c,t)} h_{s,c,t} (\ell_{s,c,t} - l_{s,c,t}) \tag{3}$$

denotes earnings and $y_{s,c,t} = Y_{s,c,t}/\ell_{s,c,t}$ denotes wages. In each period, the agent is endowed with $\ell_{s,c,t} > 0$ units of time, which are used for working or training, $0 \leq l_{s,c,t} \leq \ell_{s,c,t}$. $w_{s,\tau}$ is the skill price for labor of type s in period $\tau(c,t) = c - 1 + t$. Human capital evolves according to

$$h_{s,c,t+1} = (1 - \delta_s) h_{s,c,t} + \theta_{s,c} (h_{s,c,t} l_{s,c,t})^{\alpha_s} \tag{4}$$

with $0 \leq \delta_s < 1$ and $0 < \alpha_s < 1$.

A number of assumptions deserve comment.

1. I place no restrictions on the evolution of the endowments $h_{s,c,1}$ and $\theta_{s,c}$.

2. The human capital production function imposes the neutrality assumption that the exponents on h and l are the same. The main benefit is tractability: this allows the model to be solved in closed form. Section 2.5 shows that the neutral model fits observed age-wage profiles quite well.

3. Only time is used to produce human capital. One may think of additional inputs as maximized out. This does not change the conclusions as long the price of these inputs is proportional to the skill price.

4. I allow time endowments to vary across cohorts. This is motivated by the empirical observation that hours worked vary across cohorts. Cohorts that expect to work more hours later in life have stronger incentives to invest in human capital.

2.2 Solving the Model

For the special case where skill prices grow at constant rates and time endowments are constant, Huggett, Ventura, and Yaron (2006) show how to solve the model in closed form. Their argument generalizes to my model as follows. Assuming an interior solution, backward induction leads to the optimal investment rule

$$(l_{s,c,t} h_{s,c,t})^{1-\alpha_s} = \frac{\theta_{s,c} \alpha_s}{R w_{s,\tau(c,t)}} \sum_{j=1}^{D_s - t} w_{s,\tau(c,t)+j} \ell_{s,c,t+j} \left(\frac{1 - \delta_s}{R} \right)^{j-1} \tag{5}$$

(for $t < T$). The model has the remarkable property that training investment is independent of human capital. Longer work hours have the same effect on training as do higher skill prices; this motivates the inclusion of cohort specific time endowments into the model.

Solving the agent's problem is now easy. Given endowments $h_{s,c,1}$ and $\theta_{s,c}$, (5) solves for training time $l_{s,c,1}$. Iterating forward using the law of motion (4) yields $h_{s,c,2}$. Repeating these steps for each t unravels the agent's life history.

In principle, it is necessary to deal with corner solutions where $l_{s,c,t} = \ell_{s,c,t}$. However, this constraint does not to bind for any of the simulations reported in this paper because it would imply a measured wage of zero and an infinite deviation from the calibration targets.

2.3 Data

The data are taken from the March Current Population Survey (CPS) for 1964-2011 (King, Ruggles, Alexander, Flood, Genadek, Schroeder, Trampe, and Vick, 2010).[6] The sample contains men born between 1935 and 1968. To increase sample sizes, I divide the population into 12 equally spaced birth cohorts. Each cohort pools 3 adjacent birth years. For example, the cohort labeled 1950 contains persons born between 1949 and 1951.

For each cohort and school group, I construct age profiles of mean hours worked and of mean log wages. The age hours profiles measure time endowments. Specifically, I set $\ell_{s,c,t}$ to mean annual hours in each (s, c, t) cell divided by 2,000, which yields time endowments for middle aged men that are close to 1. The age-wage profiles form the calibration targets, as described in Section 2.4.

The measurement of wages closely follows Bowlus and Robinson (2012) and is described in Appendix (A). Each cohort's log wage profile is smoothed using an HP filter with parameter 5. Figure 2 displays selected cohort age-wage profiles, deflated by the consumer price index with base year 2000.[7] It is apparent that some of the profiles are not consistent with human capital theory and constant skill prices (Bowlus and Robinson 2012 make a similar observation). Some of the profiles are essentially flat past age 25. Many are not concave. Note also that the longitudinal wage profiles look very different from cross-sectional profiles that are sometimes used in their stead (see Thornton, Rodgers, and Brookshire 1997). They also look very different from the age-wage profiles that sometimes estimated using panel

[6] Relative to the Panel Study of Income Dynamics, the main advantage of CPS data is the much larger sample size. The drawback is that individuals cannot be followed over time.

[7] Bowlus and Robinson (2012) report log median wage profiles for a different set of cohorts. My data imply similar profiles for their cohorts, even though I use mean log wages instead. I am grateful to Chris Robinson for providing me with his data to perform this comparison.

Table 1: Fixed parameters

Parameter	Value
Cohort birth years	1935, 1938, ..., 1965, 1968
D_s: Career length	48, 47, 45, 43
R: Gross interest rate	1.04

data and imposing that a fixed age profile, combined with either year effects or cohort effects, characterizes all cohorts (e.g., Figure 3 in Huggett, Ventura, and Yaron 2006).

Age hours profiles are constructed as follows. For (s, c, t) cells inside the sample period, I compute mean hours. I then estimate a common age hours profile by regressing average hours on birth year dummies and a quartic in age. For each cohort, the age hours profile is smoothed using an HP filter with parameter 20. For ages that are past (before) the sample period, the quartic age profile is imposed, scaled to match the level of hours in the last (first) observed year. Figure 3 shows the age hours profiles for selected cohorts. For all school groups, the hours profiles get steeper over time. This implies that the incentives for investing in human capital around middle age get stronger over time.

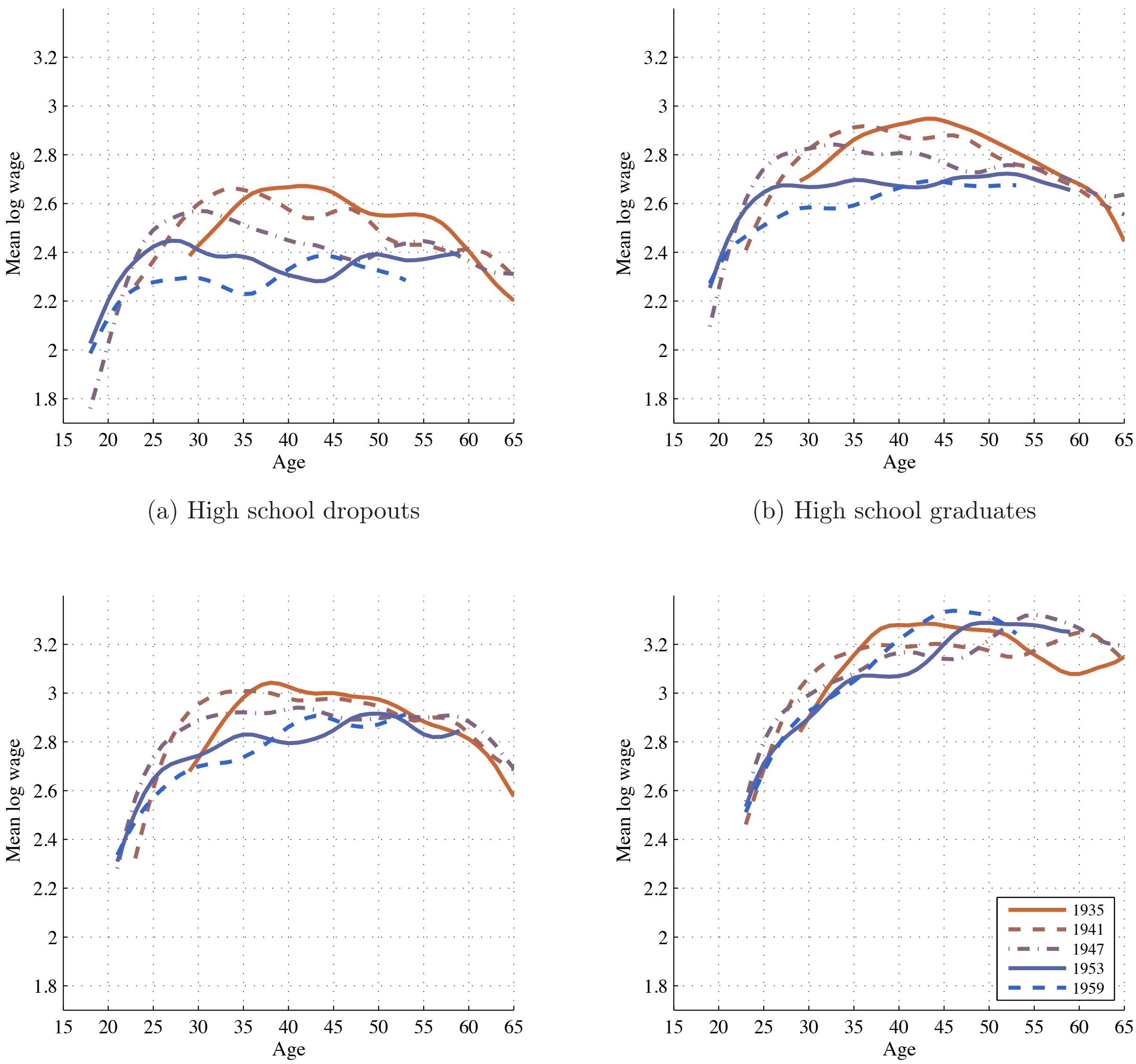
2.4 Calibration

Table 1 summarizes the parameter values that are fixed exogenously. Based on McGrattan and Prescott (2000) the gross interest rate is set to $R = 1.04$. In the data, I assume that agents start working at the ages of 18 (HSD), 19 (HSG), 21 (CD), and 23 (CG) and retire at age 65, which yields the values of D_s.

Calibrated parameters: The jointly calibrated parameters are

1. endowments: $h_{s,c,1}, \theta_{s,c}$;
2. human capital production functions: α_s, δ_s;
3. skill prices: $w_{s,\tau}$.

These parameters can be determined separately for each school group. Their values are chosen to minimize the sum of squared deviations between data and model mean log wages.



(a) High school dropouts

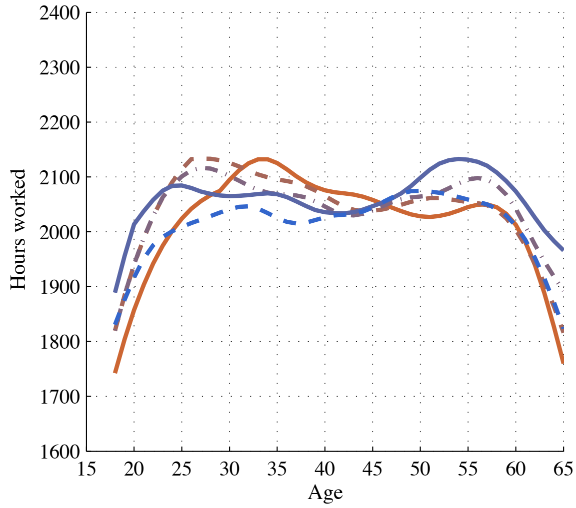
(b) High school graduates

(c) College dropouts

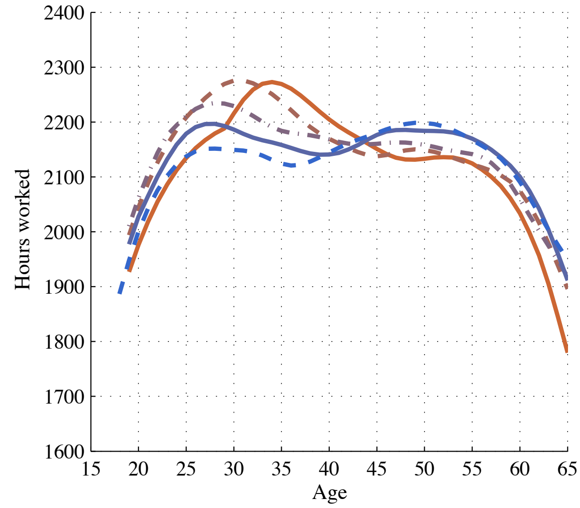
(d) College graduates

Figure 2: Age profiles of mean log hourly wages for selected cohorts

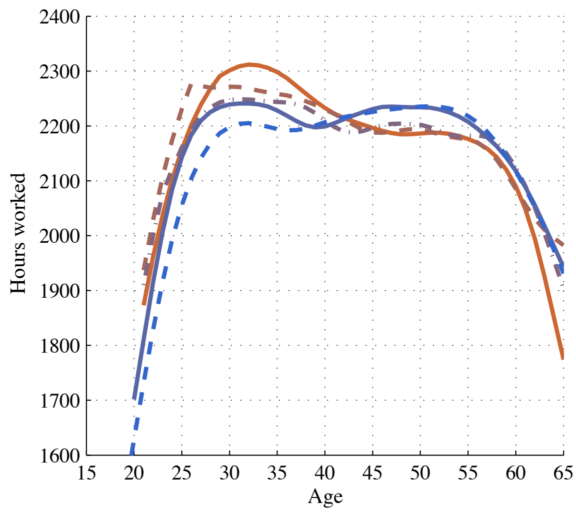
Notes: In year 2000 prices. CPS data.



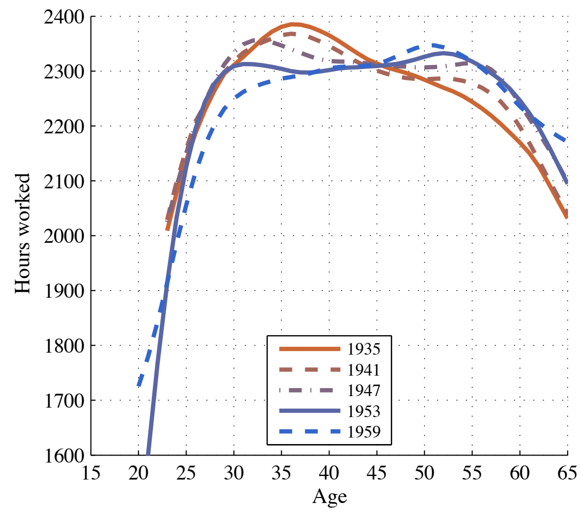
(a) High school dropouts



(b) High school graduates



(c) College dropouts



(d) College graduates

Figure 3: Age profiles of average annual hours worked for selected cohorts

Notes: CPS data.

As a scalar measure of how well the model fits the wage data, I use the pseudo-R^2 measure

$$\bar{R}_s^2 = 1 - \frac{\sum_c \sum_t (\ln y_{s,c,t} - \ln y^d_{s,c,t})^2}{\sum_c \sum_t (\ln y^d_{s,c,t} - \bar{y}^d_{s,c})^2} \tag{6}$$

where $y^d_{s,c,t}$ denotes data wages and $\bar{y}^d_{s,c}$ denotes the mean log wage in cell (s,c). The numerator is the sum of squared deviations between model and data log wages. The denominator is the total sum of squares for data log wages. If the model fits the data perfectly, then $\bar{R}_s^2 = 1$. If the model wages for each cohort are flat and equal to $\bar{y}^d_{s,c}$, then $\bar{R}_s^2 = 0$. Minimizing the sum of squared wage deviations is equivalent to maximizing $\bar{R}_s$.

Calibration details: To keep the dimensionality of skill prices low, I compute $\ln w_{s,\tau}$ from a cubic spline with 8 nodes between 1964 and 2011, which is the period for which data wages are available. Skill prices in 1964 are normalized to 1.

Since the model requires skill prices for all working ages of all cohorts, it is necessary to extend the skill price series beyond the period for which wage data are available. This is done without introducing additional calibrated parameters. I assume that the growth rate of $w_{s,\tau}$ smoothly changes from the value observed in the last data year towards the average growth rate over the data period at a rate of 0.20% per year. It is important to keep the skill price growth rate smooth, so as to avoid implausible out-of-sample wage profiles.

Similarly, to keep the dimensionality of cohort endowments low, I exploit a scaling feature of the model. Multiplying $h_{s,c,1}$ by $\lambda > 0$ and multiplying $\theta_{s,c}$ by $\lambda^{1-\alpha_s}$ multiplies the wage profile by λ while leaving training times unchanged. The calibration algorithm first computes $\ln(\theta_{s,c}/h^{1-\alpha_s}_{s,c,1})$ from a cubic spline with 6 nodes. This determines the household solution up to a scale factor. It then solves the household problem for $h_{s,c,1} = 1$. Finally, the algorithm sets each $h_{s,c,1}$ so that the model matches observed mean log wages.

The model is relatively parsimonious. 13 calibrated parameters are chosen to fit 12 cohort wage profiles, where the model fits the mean log wages of all cohorts by construction. To ensure that a global minimum is found, the calibration is run for 20 random starting values.

Penalties: For some parameter values, the model implies unobservables that could be considered implausible. With high average skill price growth rates, the model sometimes implies that human capital declines with age even for young workers. For high values of

α_s, the model sometimes implies that agents spend most of their first several work years on training rather than working.

To avoid such outcomes, the algorithm imposes quadratic penalties whenever a cohort's maximum human capital after 5 years of experience falls below $h_{s,c,1}$ or when training time exceeds an upper bound that declines linearly from 0.7 in year 1 to 0.35 in year 20 and stays constant thereafter. These restrictions are arbitrary. Dropping them would further widen the range of parameters for which the Ben-Porath model closely fits observed wage profiles.

2.5 Results

This section studies the implications of the calibrated model. I focus on the idenfication of two key model parameters: the curvature of the human capital technology α_s and the average growth rate of skill prices over the sample period, 1964-2011, denoted by Δw_s. As pointed out in the Introduction and as further explored in Section 4, the value of α_s is central for a number of empirical applications. The average growth rate of skill prices is of interest in itself. Understanding the evolution of skill prices has been the focus of a large literature, much of it focused on skill-biased technical change (e.g., Berman, Bound, and Griliches 1994; Berman, Bound, and Machin 1998). Estimating Δw_s using the Ben-Porath model is the purpose of the flat spot method proposed by Heckman, Lochner, and Taber (1998).

The first question addressed in this paper is: how far can one go towards identifying α_s and Δw_s without imposing strong restrictions on skill prices or cohort endowments? In short, the answer is: not very far. The paper's first main result may be summarized as follows. The Ben-Porath model can generate nearly identical age-wage profiles over a wide range of values taken by α_s and Δw_s. However, the model is not consistent with values of α_s above 0.8. This finding is essentially the opposite of the emerging consensus which argues that values of α_s above 0.8 are consistent with the data while much lower values are not. The model is also not consistent with rising skill prices. A wide range of negative values of Δw_s yield essentially identical wage profiles, but the model fit deteriorates significantly when Δw_s is positive. Section 3 argues that further progress towards identifying the model parameters can be made by modeling cohort endowments. Section 4 re-examines some recent empirical applications of the Ben-Porath model in light of these findings.

Examples: The paper's main result is perhaps best established by example. Figure 4 shows the age-wage profiles for high school graduates that belong to 4 cohorts with large numbers of wage observations in CPS data. Their empirical wage profiles vary greatly. The early cohorts' wage profiles peak around age 35, while the latest cohort's profile peaks near age 50. The early cohorts experience much faster wage growth early in life compared with the later cohorts.

Figure 4 also shows the wage profiles implied by the model when fixed values of α_s and Δw_s are imposed during the calibration. Even though α_s varies between 0.3 and 0.6 and Δw_s varies between -2% and 0% per year, all of the model wage profiles are visually hard to distinguish.[8] Moreover, the $\bar{R}_s^2$ measures of fit vary by no more than 0.01. Figure 1 in the Introduction shows similar examples for college graduates. In order to conserve space, I do not show the age profiles for the other cohorts and school groups. They yield similar results.

A more systematic look at the model fit is provided in Figure 5. For each school group, the figure shows $\bar{R}_s^2$ as α_s and Δw_s are varied over a grid of values. Three observations stand out:

1. The model fit, as measured by $\bar{R}_s^2$, is excellent. The maximum values of $\bar{R}_s^2$ exceed 0.97 for all school groups. The reader may refer back to Figures 1 and 4 to gain an impression of the model fit associated with these values of $\bar{R}_s^2$.

2. The model fit is very similar over a wide range of α_s and Δw_s values. However, the fit deteriorates for α_s close to 1 and for Δw_s above 0.

3. Even holding α_s constant, the model achieves very similar wage profiles as Δw_s is varied, and vice versa.

These results suggest that estimating α_s and Δw_s solely from mean age-wage profiles is not a promising approach.

2.6 Curvature of the Job-Training Technology

Figure 6 shows how the model's fit varies with the curvature parameter α_s. I calibrate the model while fixing the value of α_s on a grid that varies between 0.20 and 0.90. Figure 6

[8] There are notably discrepancies for the last cohort, but they occur at older ages, which are not observed in the data and can therefore not be used to estimate parameters.

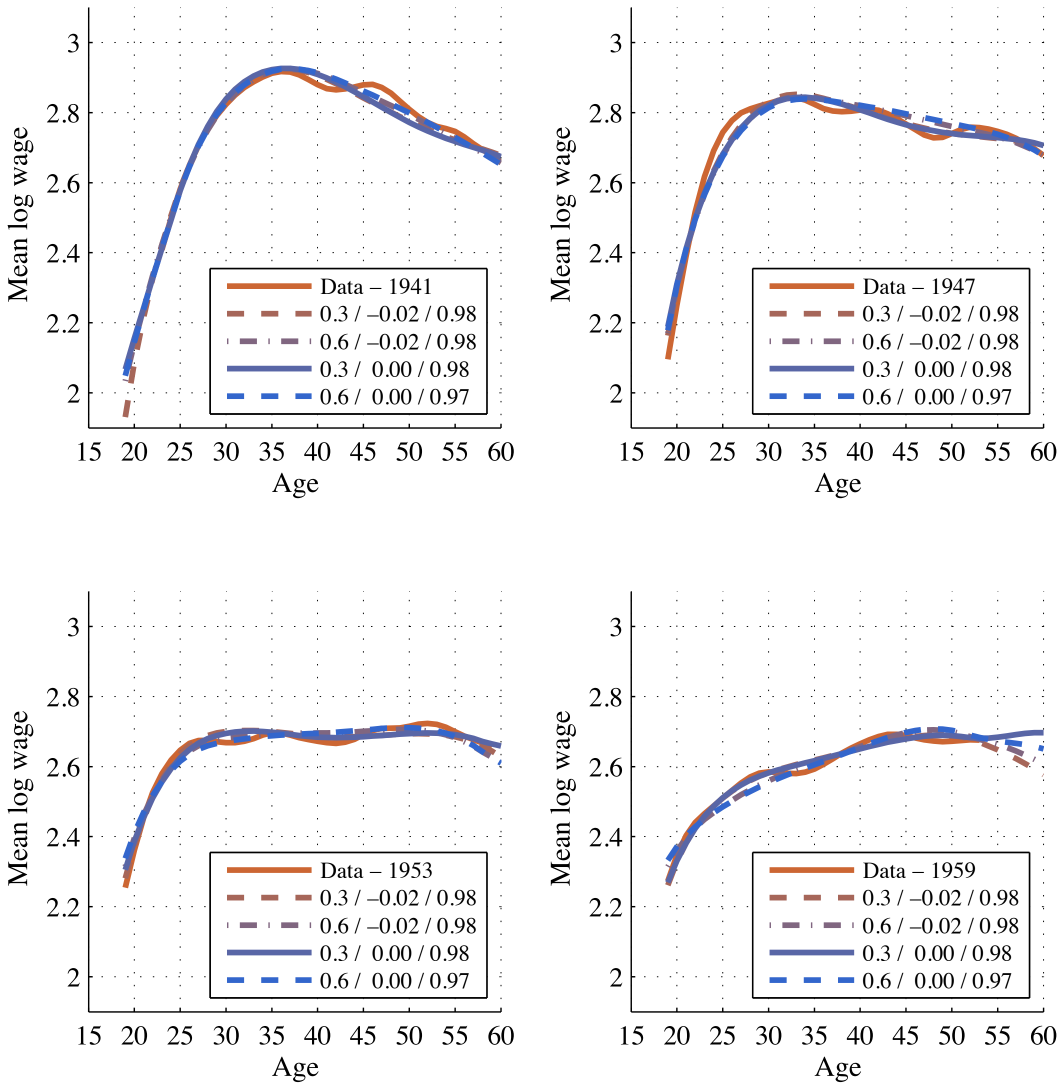
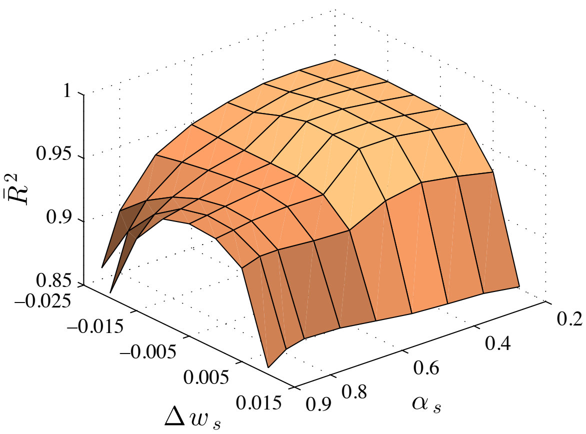
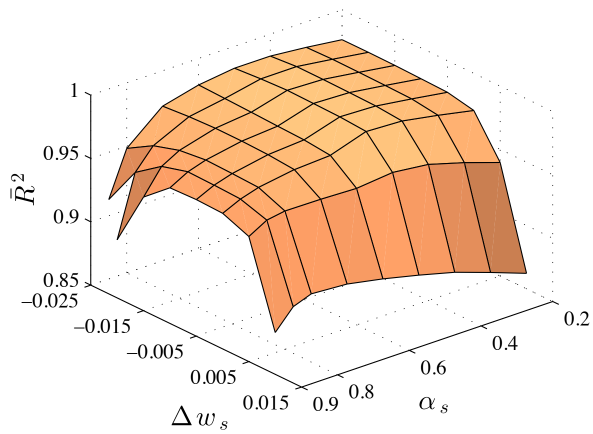


Figure 4: age-wage profiles for high school graduates

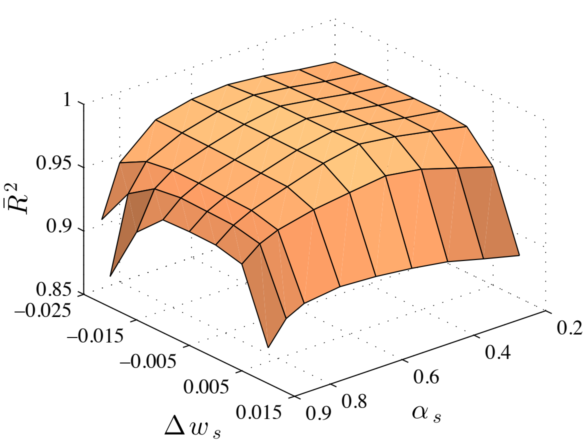
Notes: Each panel shows mean log wages for one cohort. Each legend entry shows the fixed values of $\alpha_s/\Delta w_s$ used in the calibration and the resulting measure of fit, $\bar{R}^2$.



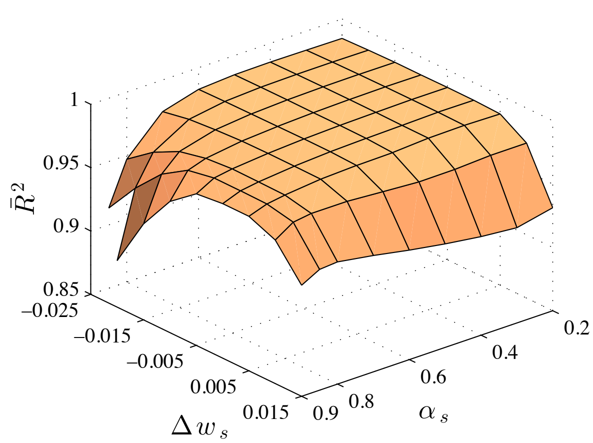
(a) High school dropouts



(b) High school graduates



(c) College dropouts



(d) College graduates

Figure 5: Model fit for fixed values of α_s and Δw_s

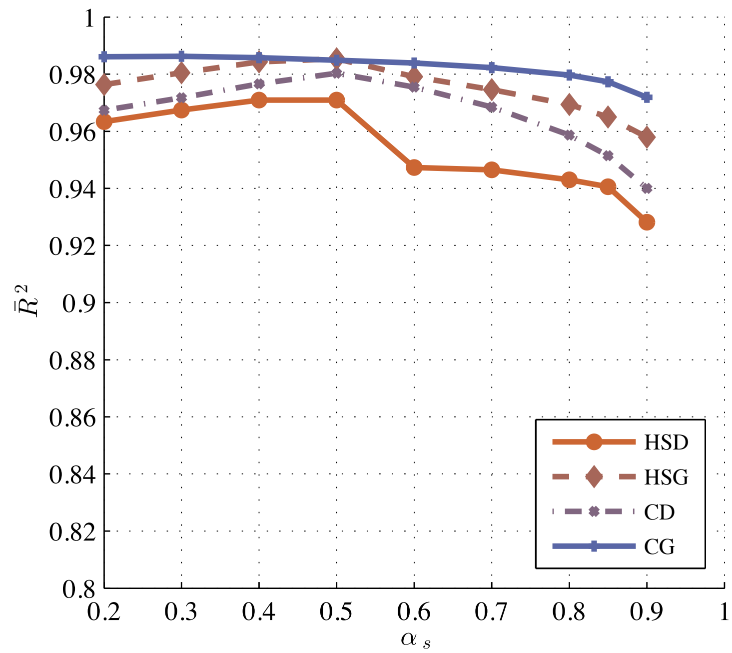


Figure 6: Model fit for fixed values of α_s

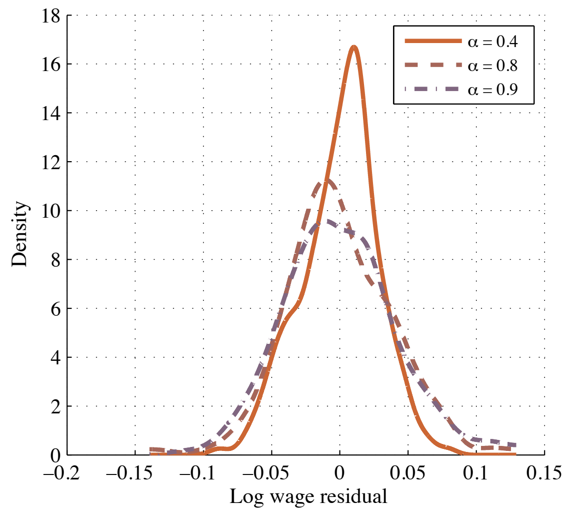
shows the resulting values of $\bar{R}_s^2$ for all school groups.

Except for high school dropouts, where the model never fits as well as for the other groups, $\bar{R}_s^2$ varies little over a wide range of curvature parameters. For values above $\alpha_s = 0.5$, the model fit deteriorates.

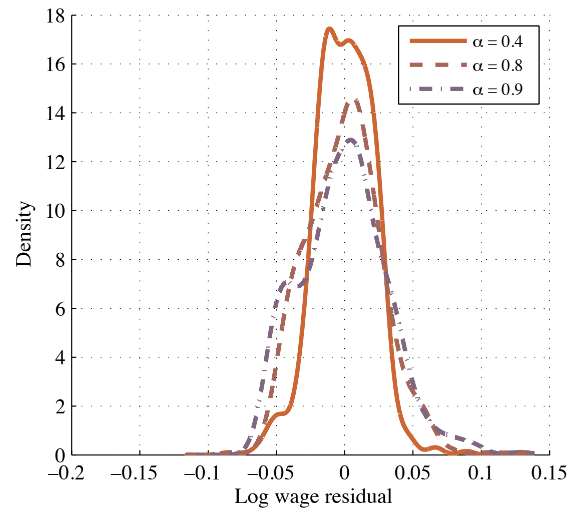
To illustrate the deterioration of model fit for high values of α_s, Figure 7 shows the density of residual log wages, defined as the difference between model and data mean log wages in each (s, c, t) cell. Except for college graduates, the dispersion of residuals increases markedly for values of α_s above 0.8.

These results are essentially the opposite of the "emerging consensus." Recent studies have argued that curvature parameters above 0.8 are consistent with the data, while lower values are not (Guvenen and Kuruscu, 2010; Manuelli and Seshadri, 2010). The results shown in Figure 7 suggest the converse: values of α_s as low as 0.3 fit the data well while values around 0.9 do not.

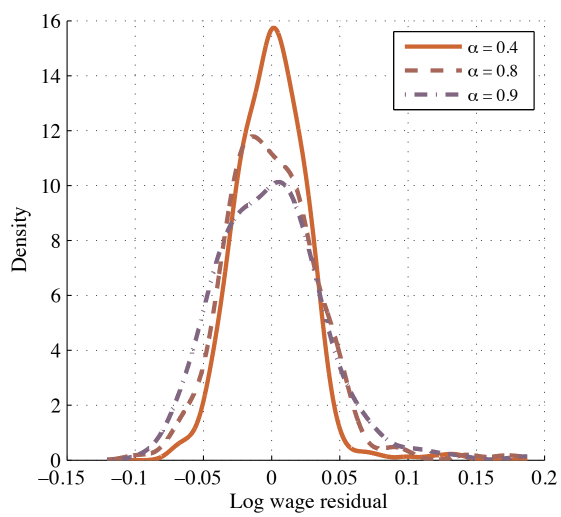
Why do near linear models fail? Models with near linear technologies have strong implications for the wages of older workers. Training investment diminishes rapidly after



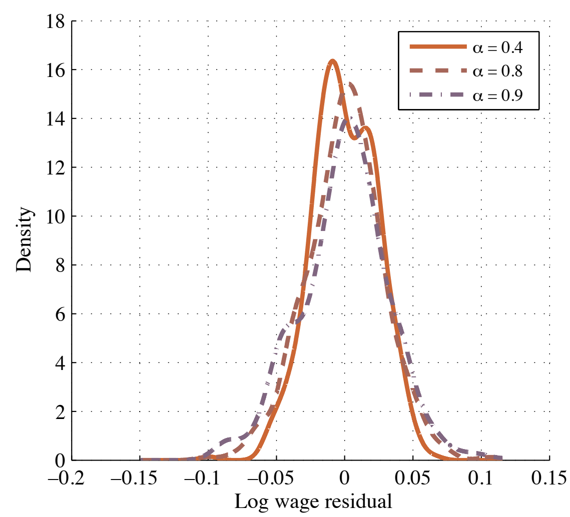
(a) High school dropouts



(b) High school graduates

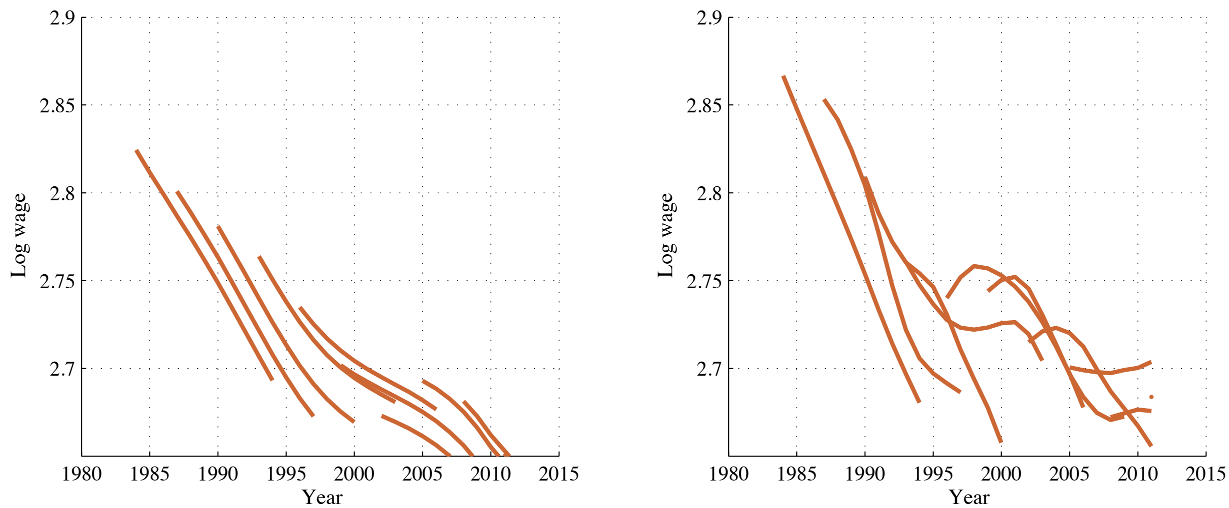


(c) College dropouts



(d) College graduates

Figure 7: Density of log wage residuals



(a) Model

(b) Data

Figure 8: Age-wage profiles for high school graduates aged 50-60

Notes: Each line represents the age profile of log wages for one cohort between the ages of 50 and 60. The model is calibrated with α_s fixed at 0.9.

job-market entry and effectively ceases around age 45. Assuming that human capital does not depreciate, each cohort's time input per hour worked, $h_{s,c,t}(1 - l_{s,c,t}/\ell_{s,c,t})$, remains constant over time. As a result, skill prices can be measured, up to a scale factor, from observed wages. This feature is the basis for the flat spot method proposed by Heckman, Lochner, and Taber (1998) and Bowlus and Robinson (2012).

One implication is that the measured wages of older cohorts comove almost perfectly. Figure 8 illustrates this point in the model when α_s is fixed at 0.9. Panel (a) shows the log wage profiles generated by the model for high school graduates. Each line follows one cohort between the ages of 50 and 60. The model wage profiles are almost exactly parallel. Panel (b) shows the corresponding data wages. There are substantial movements in the relative wages of different cohorts, which the model cannot match, unless α_s is lower and training continues until later in life.

In a sense, the model fails for high values of α_s because its implications are then consistent with the assumptions underlying the flat spot method. I return to this point in Section 4.1.

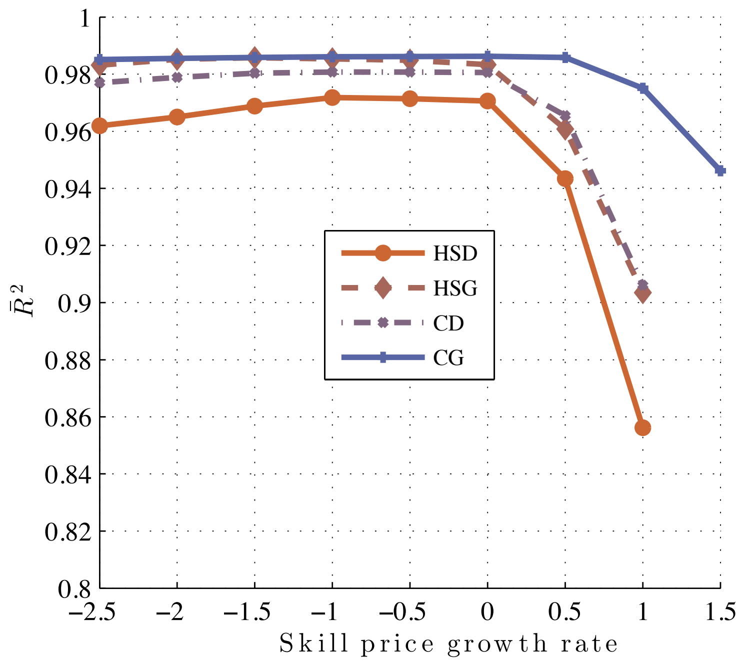


Figure 9: Model fit for fixed values of skill price growth, Δw_s.

2.7 Skill Price Growth Rates

Figure 9 examines how well the model replicates observed wage profiles for a range of average skill price growth rates, Δw_s. The model is calibrated for fixed values of Δw_s that range from -2.5% to 1.5%.

Once again, the model fit as measured by $\bar{R}_s^2$, is very similar over a wide range of skill price growth rates. In fact, $\bar{R}_s^2$ is essentially constant between $\Delta w_s = -0.025$ and 0 (with the exception of high school dropouts). However, it deteriorates dramatically as Δw_s turns positive. This finding suggests that attempts to extract skill prices from age-wage profiles alone, such as the flat spot method, are not promising. Section 4.1 examines this point in more detail.

2.8 Human Capital Depreciation

An important part of the emerging consensus regarding Ben-Porath parameters holds that human capital does not depreciate. This has two implications. First, it permits identification of other model parameters with limited wage data. Notably, the NLSY does not contain older workers, which makes it difficult to estimate the depreciation rate. Second, zero depreciation makes the flat spot method feasible. Depreciation drives an unknown wedge between skill prices and measured wages, even if human capital investment is zero for older workers.

The basis for the zero depreciation assumption is stated most clearly by Manuelli and Seshadri (2010, p. 14): "Available evidence suggests that in the United States, wage rates do not fall at the end of the life-cycle... This implies that the depreciation rate $\delta_h = 0$." In a similar spirit, Heckman, Lochner, and Taber (1998) point to U.K. wage profiles reported in Meghir and Whitehouse (1996) which lack a clear peak. Kuruscu (2006b) points to the cross-sectional wage profile of college graduates in the 1997 CPS which is roughly flat after age 40.

For high values of α_s the logic linking flat age-wage profiles with zero depreciation is consistent with the Ben-Porath model. All of the authors cited in the previous paragraph estimate that α_s is close to 0.9. Calibrating the model subject to this restriction implies that training time falls to near zero by age 50. Effective time input per hour worked, $h_{s,c,t}(1 - l_{s,c,t}/\ell_{s,c,t})$, then declines at rate δ_s. In order to account for age-wage profiles that are roughly flat past age 50, the model sets $\delta_s = 0$.

However, for lower values of α_s the model generally implies positive depreciation rates. As an example, consider high school graduates when α_s is fixed at 0.5 and Δw_s is fixed at -1%. Figure 10 summarizes the life-histories of selected model cohorts. Panel (a) shows that 50 year old workers still spend about 20% of their time endowment on job training. With zero depreciation, effective time inputs would continue to grow until retirement, mainly as a result of falling training times. In order to offset this, the model sets the depreciation rate to 6.2%. Now the falling $h_{s,c,t}$ roughly offsets the growing time input, $l_{s,c,t}/\ell_{s,c,t}$, which keeps effective time inputs roughly constant between the ages of 50 and 60.

An additional complication arises because the growth rate of skill prices is not known. Roughly flat wage profiles for men in their 50s are consistent with rising time inputs, as long as their growth is offset by falling skill prices. Consistent with this logic, the calibrated depreciation rates vary substantially, depending on the assumed values of α_s and Δw_s. Even if attention is restricted to parameter combinations that yield $\bar{R}_s^2$ within 0.010 of the best fitting model, the calibrated depreciation rates range from 0.00 to 0.07. In contrast to the logic expressed by Manuelli and Seshadri (2010), the model can generate roughly flat age-wage profiles even with substantial depreciation. I conclude that inferring the depreciation rate of human capital from the shape of age-wage profiles is not promising without an explicit model of wage determination.

2.9 Robustness

In this section I briefly summarize a number of robustness checks.

1. Using log median wages instead of mean log wages avoids issues related to top coding that are explained in detail in Bowlus and Robinson (2012).

2. Using weekly rather than hourly wages avoids issues related to the coding of hours worked per year that are explained in detail in Bowlus and Robinson (2012).

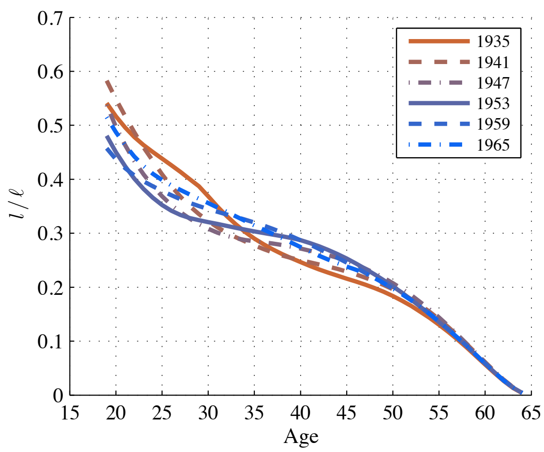
3. Restricting the sample to full-time/full-year workers addresses potential selection issues, especially among older workers.

4. Defining a model cohort as a single birth year instead of a 3 year interval increases the number of data points.

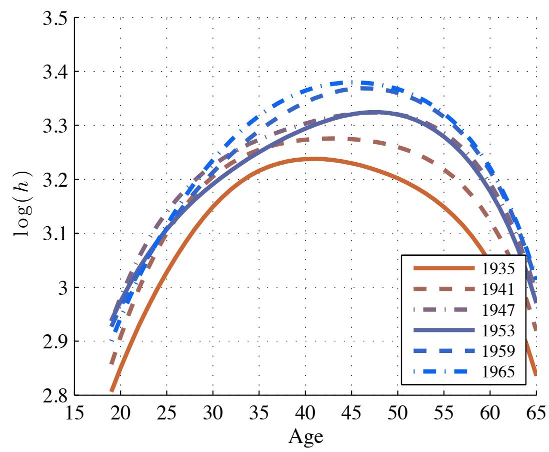
5. Setting all $\ell_{s,c,t} = 1$ brings the model closer to the literature which often abstracts from time-varying time endowments.

None of these extensions materially change the results. The only notable change occurs when weekly wages are used. This worsens the model fit as measured by $\bar{R}_s^2$ for high values of α_s.

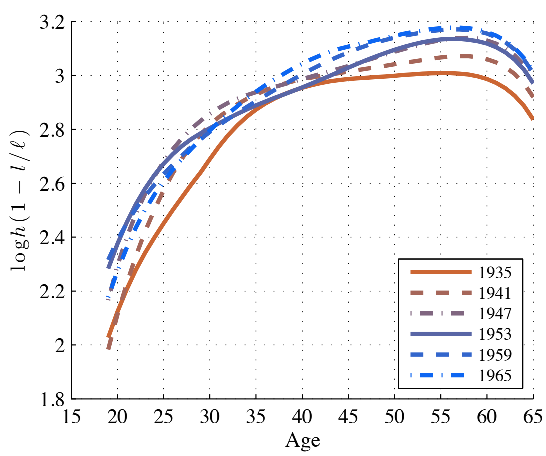
Given how similar the model wage profiles are across a broad range of parameters, it is tempting to conclude that some of the model parameters are not identified in the sense that multiple values of $(\alpha_s, \delta_s, \Delta w_s)$ yield the same age-wage profiles. However, it is easy to construct numerical examples where the model cannot fit the same wage profiles for multiple skill price growth rates or α_s, even though it fits them perfectly, by construction, for one particular parameter combination. The finding that Δw_s and α_s are not identified is not a generic feature of the model; it is a feature of the model applied to U.S. wage data.



(a) Training time



(b) Human capital



(c) Time input per hour worked

Figure 10: Life-cycle profiles for high school graduates.

Notes: Each line represents one birth cohort. The model is calibrated subject to the restrictions $\alpha_s = 0.5$ and $\Delta w_s = -0.01$.

3 A Model of School Choice

The results reported so far allowed the human capital endowments and learning abilities to vary arbitrarily across school groups and over time. Of course, in the data, agents in all school groups are drawn from the same population. While their endowments may differ due to self-selection, very large differences in endowments across groups or over time may be implausible.

This section explores how much progress can be made towards identifying the key model parameters, α_s and Δw_s, by imposing that workers in different school groups are drawn from the same population. To this end, I develop a model of school choice. At birth, agents are endowed with random human capital endowments and learning abilities. They choose from discrete schooling levels to maximize lifetime earnings. Once individuals enter the labor market, the model is identical to the one developed in Section 2.

The model imposes two restrictions on the evolution of human capital endowments and learning abilities across school groups:

1. It imposes limits on how much a school group's mean learning ability can change over time. The reason is that the mean learning ability across school groups must equal the population average.

2. It implies that the relative abilities and human capital endowments of school groups change over time as educational attainment grows. This is a powerful implication because U.S. educational attainment expanded until the 1950 birth cohort, but not between the 1950 and the 1970 cohorts.[9]

3.1 The Environment

Demographics: In each period, a unit measure of households is born. Each lives for T periods. Denote the year in which a person of cohort c is aged t by $\tau(c,t) = c + t - 1$.

Endowments: At birth, agents draw three independently distributed endowments: a, h_1, and p_s. Learning ability $a \sim N(0,1)$ determines how efficiently the agent produces

[9] Hendricks and Schoellman (2011) exploit this intuition to identify the changing selection of U.S. college students over the post-war period.

human capital in school or on the job. $\ln(h_1) \sim N(\bar{h}_{c,1}, \sigma_{h1})$ denotes the age 1 endowment of human capital, which I think of as produced during childhood prior to age 1. Its mean grows at a constant rate: $\bar{h}_{c,1} = g_{h1}(c-1)$. $\bar{h}_{1,1}$ is normalized to 0 by choosing units of h. p_s is a vector of "psychic costs" that determines how much the individual enjoys schooling level s. p_s is drawn from a Gumbel distribution, which simplifies the school choice problem. Allowing the endowments to be correlated has no material effect on the results.

In each period, a person works $\ell_{s,c,t}$ market hours. They can be used for work or study.

Preferences: Upon entering the labor market, individuals maximize the discounted present value of lifetime earnings. Equivalently, individuals maximize the present value of utility derived from consumption subject to a lifetime budget constraint with perfect credit markets. There is no need to specify the utility function. School choice is also affected by the psychic costs p_s (details below).

Technologies: Human capital is produced in school and on the job. Agents choose from S discrete school levels. Level s lasts T_s years and results in $h_{T_s+1} = F(h_1, a, s)$ units of type s human capital at the start of work (at age $1 + T_s$). On the job, human capital is produced from human capital and study time $l_{s,c,t}$ according to

$$h_{s,c,t+1} = (1-\delta_s)h_{s,c,t} + A(a,s)(h_{s,c,t}l_{s,c,t})^{\alpha_s} \tag{7}$$

where learning ability affects productivity according to $A(a,s) = e^{A_s + \theta a}$. I normalize $\mathbb{E}(a) = 0$ and $Var(a) = 1$ by choosing A_s and θ. Allowing A_s to grow over time has no material effect on the results.

3.2 Household Problem

I solve the household problem by backward induction, starting with the work phase.

Work phase: At the start of work, the agent is endowed with h_{T_s+1} units of type s human capital and with a productivity parameter $A(a,s)$. He maximizes the discounted present value of lifetime earnings

$$V(h_{T_s+1}, a, s, c) = \max_{\{l_t\}} \sum_{t=T_s+1}^{T} R^{-t} Y(l_t, h_t, s, c, t) \tag{8}$$

subject to the law of motion for h (7) and the time constraint $0 \le l_t \le \ell_{s,c,t}$. R denotes the exogenous gross interest rate. Period earnings are given by

$$Y(l_t, h_t, s, c, t) = w_{s,\tau(c,t)} h_t (\ell_{s,c,t} - l_t) \tag{9}$$

Wages are given by $y(l_t, h_t, s, c, t) = Y(l_t, h_t, s, c, t)/\ell_{s,c,t}$. The solution to this problem is again given by (5), where the productivity parameter $\theta_{s,c}$ is replaced by $A(a, s)$, which varies across individuals.

It is now necessary to derive a closed form solution for the case where agents differ in their endowments a and h_1. Assuming that study time is interior, optimal human capital investment is given by

$$(l_t h_t)^{1-\alpha_s} = A(a, s) d_t \tag{10}$$

where

$$d_t = \frac{\alpha_s}{R w_{s,\tau(c,t)}} \sum_{j=1}^{T-t} w_{s,\tau(c,t+j)+j} \ell_{s,c,t+j} \left(\frac{1-\delta_s}{R} \right)^{j-1} \tag{11}$$

with $d_T = 0$. Then human capital production is given by

$$\begin{aligned} A(a, s)(l_t h_t)^{\alpha} &= A(a, s)(A(a, s) d_t)^{\alpha_s/(1-\alpha_s)} && (12) \\ &= A(a, s)^{1/(1-\alpha_s)} d_t^{\alpha_s/(1-\alpha_s)} && (13) \end{aligned}$$

so that human capital is given by

$$h_t = h_1 (1-\delta_s)^{t-1} + A(a, s)^{1/(1-\alpha_s)} x_t \tag{14}$$

where

$$x_t = \sum_{j=1}^{t-1} d_j^{\alpha_s/(1-\alpha_s)} (1-\delta_s)^{t-1-j} \tag{15}$$

The fact that x_t is common to all agents in a (school, cohort) cell dramatically simplifies the numerical solution of the model. Corner solutions need not be considered, given that the model will be calibrated to match observed age-wage profiles. Setting $l = \ell$ would imply a wage rate of zero and an infinite deviation from the calibration targets.

School phase: At the start of life, the agent chooses schooling to maximize

$$W_s(p_s, h_1, a, s, c) = \ln V(F[h_1, a, s], a, s, c) + \mu_{s,c} + \pi_c p_s \tag{16}$$

In addition to lifetime earnings V the agent enjoys an idiosyncratic "psychic" utility $\pi_c p_s$ and a common, school specific utility $\mu_{s,c}$. The common utility $\mu_{s,c}$ allows the model to match the fraction of persons choosing each school level in each cohort. The psychic utility plays its usual role as a stand-in friction that generates imperfect school sorting by ability and h_1 (see Heckman, Lochner, and Todd 2006 for a discussion). $\pi_c > 0$ is a scale factor that defines the units of the psychic cost. It grows at a constant rate: $\pi_c = \pi_1(1+g_\pi)^{c-1}$.

Since p_s is drawn from a Gumbel distribution, the probability of choosing school level s is given by $X_s / \sum_j X_j$ where

$$X_s = \exp\left(\frac{V(F[h_1, a, s], a, s, c) + \mu_{s,c}}{\pi_c}\right) \tag{17}$$

(e.g., Greene 2012, ch. 18.2).

3.3 Calibration

The fixed model parameters are set as described in Section 2.4. The following parameters are calibrated jointly:

- endowment parameters: θ, (σ_{h1}, g_{h1}), (π_1, g_π);
- human capital technologies: α_s, δ_s, A_s;
- school costs $\mu_{s,c}$, where $\mu_{1,c}$ may be normalized to 0;
- skill prices: $w_{s,\tau}$.

To limit the number of parameters, skill prices are computed from a cubic spline with 9 nodes. These parameters are jointly calibrated for all school groups using a simulated method of moments. I search over the parameter space. For each parameter guess, I solve the model and simulate $1,000$ individuals. The values of $\mu_{s,c}$ are chosen so that the model exactly matches the fraction of persons choosing each school level in each cohort. The algorithm minimizes the sum of squared deviations between the mean log wages generated by the model versus the data in each (school, cohort, age) cell. Equivalently, the model maximizes $\bar{R}^2$, defined as the average of $\bar{R}_s^2$ (defined in equation 6) across school groups.

Table 2 shows the values of the calibrated parameters. All α_s are set to values near 0.55 with depreciation rates near 5.5%. All skill price growth rates are small but poorly identified, as

Parameter	Description	Value
On-the-job training		
A_s	Productivity	0.11, 0.10, 0.11, 0.12
α_s	Curvature	0.51, 0.55, 0.62, 0.55
δ_s	Depreciation rate	0.056, 0.053, 0.057, 0.061
Endowments		
σ_{h1}	Dispersion of h_1	0.789
g_{h1}	Growth rate of h_1 [pct]	-1.08
θ	Ability scale factor	0.051
π_1	Psychic cost scale factor	0.140
g_π	Growth rate of π [pct]	2.68
Other		
Δw_s	Skill price growth rate, 1964-2010 [pct]	-0.13, 0.07, 0.21, 0.28
$\bar{R}^2_s$	Goodness of fit.	0.87, 0.95, 0.96, 0.95

Table 2: Calibrated parameters. Model with school choice.

discussed in Section 3.4. To achieve falling average wages, as observed in the data, human capital endowments decline over time at -1.1% per year.

Model fit. The $\bar{R}^2$ values indicate that the model fits the observed wage profiles nearly as well as the baseline model, except for high school dropouts. A visual assessment of fit is provided in Figure 11, which shows model generated and observed mean log age-wage profiles for high school graduates of selected cohorts. Figure 12 shows the same information for college graduates.

3.4 Results

This section studies to what extent the additional restrictions implied by the school choice model help identify the key model parameters, α_s and Δw_s. I follow the approach used in Section 2.5 and calibrate the model for fixed values of either α_s or Δw_s.[10] To conserve space, I focus on high school graduates and college graduates.

[10]Recall that Δw_s denotes the average growth rate of w_s over the period 1964-2010. To fix this in the calibration, I impose a restriction on the spline nodes for $\ln(w_{s,\tau})$ for 1964 and 2010.

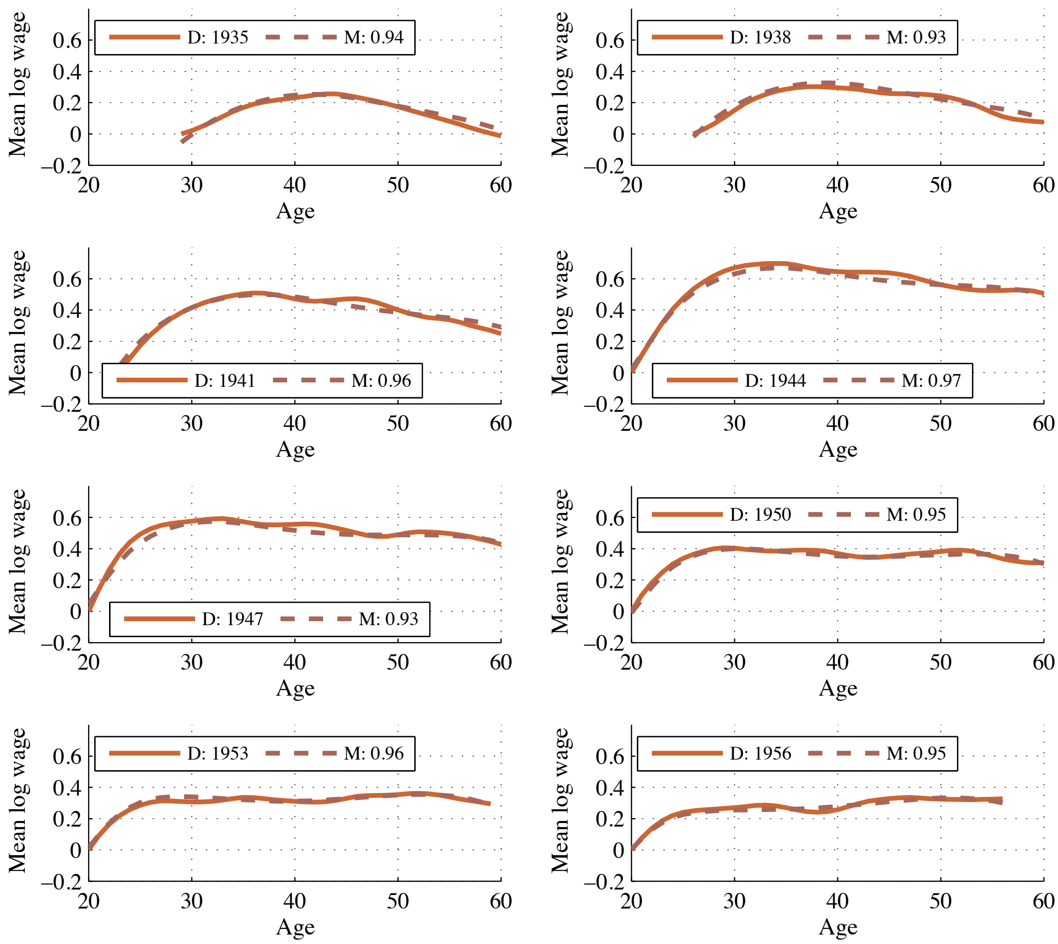


Figure 11: Model fit: high school graduates

Notes: The figure shows mean log wage profiles for selected cohorts. "D" followed by the cohort's birth year indicates the data profile. "M" followed by the $\bar{R}^2$ value for this cohort indicates a model profile.

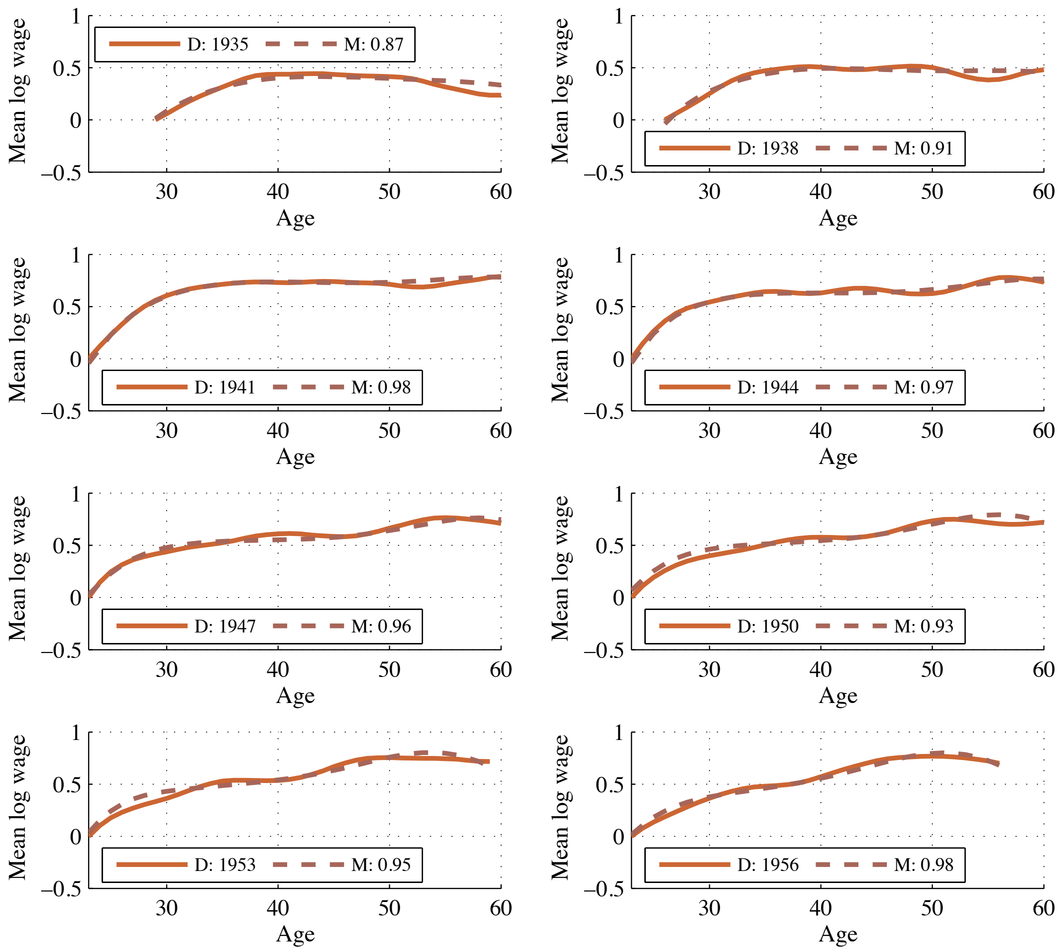
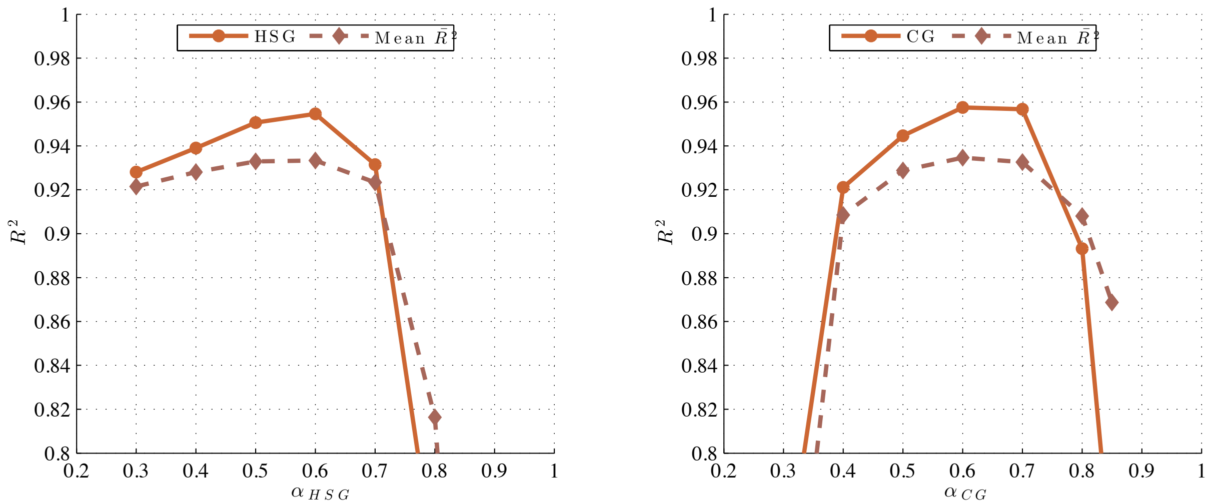


Figure 12: Model fit: college graduates

Notes: See Figure 11.



(a) High school graduates

(b) College graduates

Figure 13: Model fit for fixed values of α_s

Notes: Each panel shows $\bar{R}_s^2$ and the mean of $\bar{R}_s^2$ across school groups.

3.4.1 Curvature of the Job-training Technology

Figure 13 shows how the model fit varies with the curvature parameters α_s. Panel (a) focuses on high school graduates. Each data point represents a model economy that is calibrated for a fixed value of α_{HSG}. In addition, I impose the restriction $\alpha_{HSD} \leq \alpha_{HSG} \leq (\alpha_{CD}, \alpha_{CG})$. Without this restriction, the calibration algorithm yields implausible parameter values. For example, when α_{HSG} is fixed at 0.8, the algorithm chooses all other α_s to be much lower, roughly in line with the baseline calibration. Panel (b) fixes α_{CG} and imposes that $\alpha_s \leq \alpha_{CG}$ for all s.

As suggested by the baseline model with unrestricted α_s, the preferred values of α_{HSG} and α_{CG} are near 0.6. For both groups, $\bar{R}^2$ deteriorates rapidly for $\alpha_s > 0.7$, but remains reasonable for values as low as 0.4 for high school graduates and 0.5 for college graduates. Relative to the results of Section 2.5, modeling cohort endowments does not pin down α_s precisely, but it identifies them more tightly. In contrast to the "emerging consensus," models with near linear technologies ($\alpha_s \geq 0.8$) fit the data poorly.

3.4.2 Skill Price Growth Rates

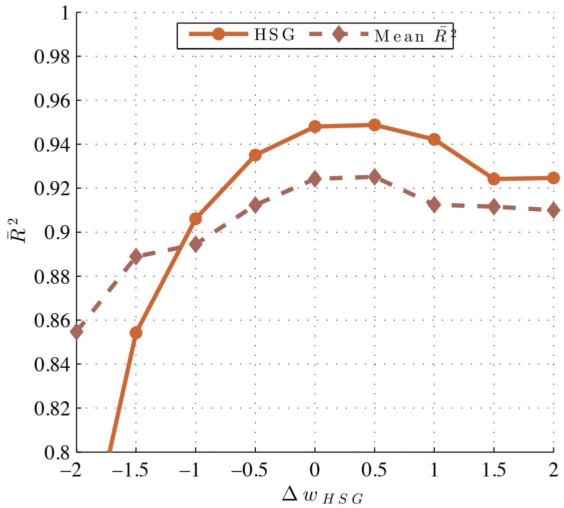
Figure 14 shows how the model fit varies with the skill price growth rates, Δw_s. Panel (a) fixes Δw_{HSG}, while panel (b) fixes Δw_{CG}. In both cases, a wide range of growth rates, roughly from -0.5% to $+1\%$, yields similar values of $\bar{R}^2$. As skill price growth rates increase, the model offsets this by reducing the growth rate of human capital endowments. Together with minor adjustments in α_s and δ_s, this is sufficient to keep the model age-wage profiles nearly unchanged.

Also not identified are the growth rates of relative skill prices. Notably, across the cases shown in Figure 14, the growth rate of the college premium, $\Delta w_{CG} - \Delta w_{HSG}$, can be negative or positive. When higher values of Δw_{CG} are imposed, the calibrated values of Δw_{HSG} rise only mildly. The model keeps the wage profiles of both groups roughly unchanged by varying the correlation between schooling and human capital endowments over time.

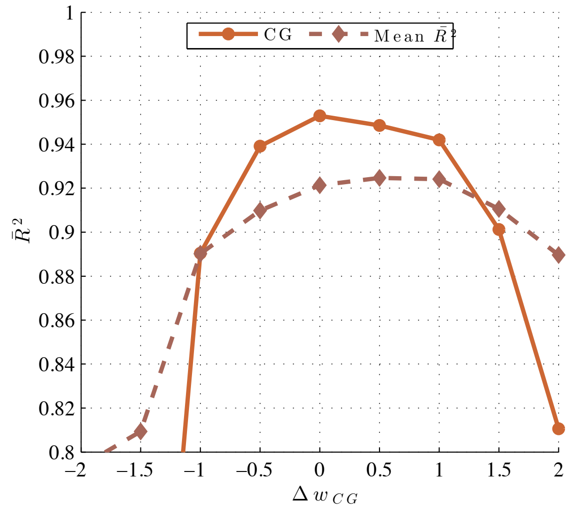
For example, if the college skill price grows rapidly, the model must generate falling human capital endowments of college graduates relative to high school graduates: $g(h_1|CG) < g(h_1|HSG)$. Otherwise, the wages of college graduates would rise too fast. To achieve this, the model sets a negative growth rate of π_c, which weakens the role of the psychic cost over time. For early cohorts, large psychic costs imply a small gap between the human capital endowments of high school graduates relative to college graduates.[11] Over time, as the psychic cost becomes less important, this gap opens up, so that $g(h_1|CG) < g(h_1|HSG)$. If high school skill prices grow rapidly, the model achieves the reverse by setting a positive growth rate of π_c.

An interesting implication is that the model can account for the entire increase of the college wage premium over the sample period, even if both skill prices grow at the same rates. This is consistent with the findings of Hendricks and Schoellman (2011). The mechanism is also similar: with constant relative skill prices, the human capital endowments of college graduates grow faster than those of high school graduates.

[11] The model implies that agents with high a and low h_1 are more likely to attend college. Unless a and h_1 are highly correlated, high school graduates have, on average, higher human capital endowments (but lower abilities) compared with college graduates.



(a) High school graduates



(b) College graduates

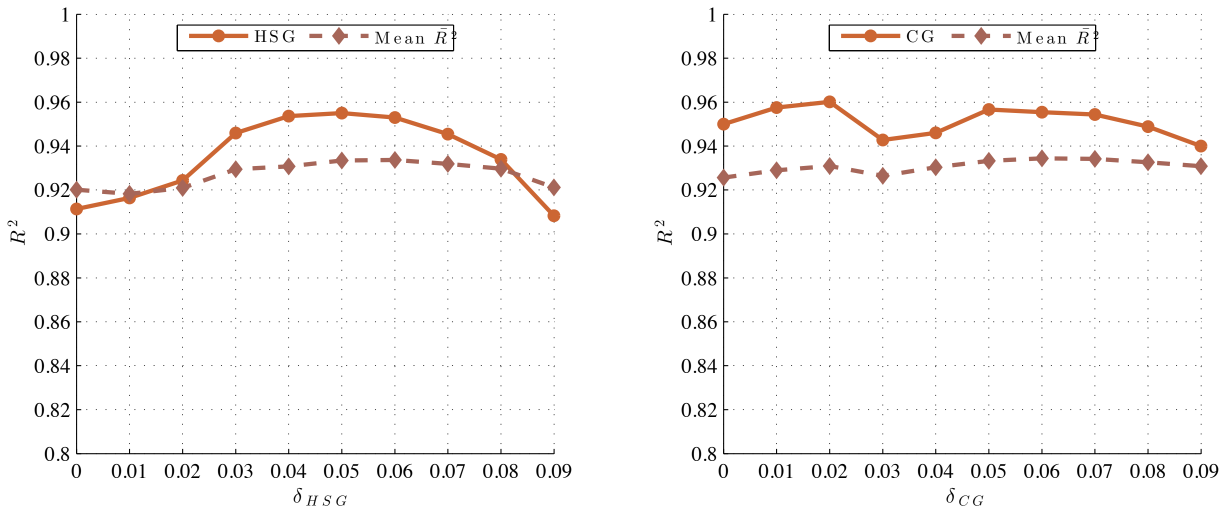
Figure 14: Model fit for fixed values of skill price growth

3.4.3 Human capital depreciation

Figure 15 shows how the model fit varies with the depreciation rate of human capital. Panel (a) shows fixed values of δ_{HSG}, while panel (b) shows the same for δ_{CG}. In both cases, depreciation rates are poorly identified. For high school graduates, depreciation rates between 3% and 7% fits the data well. For college graduates, the model fits the data about equally well over the entire range of depreciation rates considered.

4 Applications

As pointed out in the Introduction, a number of recent empirical applications of the Ben-Porath model have assumed near linear technologies and no human capital depreciation. By contrast, the estimates presented in this paper suggest that α_s takes on values below 0.7 and that human capital depreciates at around 5% per year. This section reexamines three applications of the Ben-Porath model in light of these findings. In all cases, I find that previously published results are sensitive to the assumption of near linear human capital production functions.



(a) High school graduates

(b) College graduates

Figure 15: Model fit for fixed depreciation rates

4.1 The Flat Spot Method

The flat spot method was proposed by Heckman, Lochner, and Taber (1998) as a method for estimating skill prices from data on empirical age-wage profiles. It is motivated by an implication of the Ben-Porath model that holds true regardless of parameter values: past some age, human capital investment becomes negligible ($l_{s,c,t} \approx 0$). Assuming that human capital does not depreciate, the growth rate of measured wages, following a cohort over time, equals the growth rate of skill prices.

Bowlus and Robinson (2012) apply this method to CPS wage data and arrive at a striking finding: the average growth rate of skill prices, 1964-2009, is very similar for all school groups. Almost the entire increase in the college wage premium is attributed to the rising relative quality of college graduates relative to high school graduates.[12]

The results presented in Sections 2.5 and 3.4 suggest that the flat spot method may be biased. Very different growth rates of skill prices are consistent with the same model generated age-wage profiles.

[12] Using a very different approach Hendricks and Schoellman (2011) arrive at a similar conclusion.

Approach. To assess the potential bias of the flat-spot method, I estimate skill prices from model generated wage profiles, using the flat spot age ranges proposed by Bowlus and Robinson (2012).[13] Specifically, for each year t I estimate the change in log skill prices as the change in log wages between t and $t+1$, averaged over all cohorts that are in the flat-spot age range in t. Since none of the model cohorts are in the flat spot age range early on, the flat spot wage series begins only around 1980. To assess the relationship between flat spot estimates and "true" skill price growth, I calibrate the model for a range of fixed values of Δw_s.

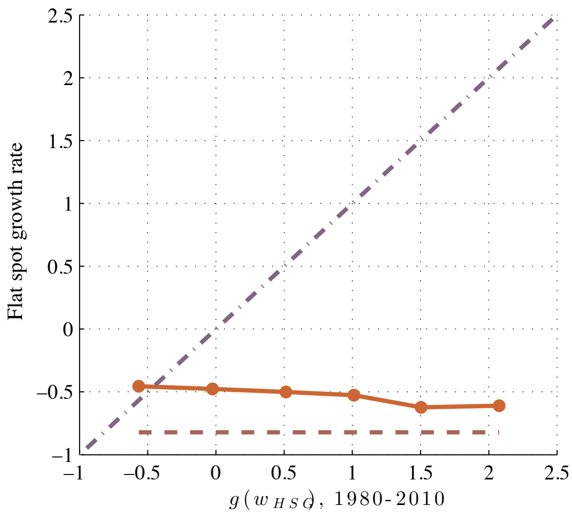
Results. Figure 16 compares the skill price growth rates estimated using the flat-spot method with the "true" skill price growth rate according to the model. Both growth rates are averaged over the time period for which flat-spot wages can be calculated. Each data point represents one fixed value of Δw_s in the model. Only models that fit the data "well," in the sense that $\bar{R}^2$ is within 0.04 of the best fitting model, are shown.

For high school graduates, shown in panel (a), the flat spot method yields essentially the same skill price growth rate even as the true growth rate varies by more than 2 percentage points. This is due to the fact that the model can generate almost identical age-wage profiles for a wide range of skill price growth rates. The flat spot method does better for college graduates, shown in panel (b), where faster true skill price growth is associated with higher flat spot estimates. Still, the flat spot estimates are too high when skill prices decline over time and too low when skill prices rise over time.

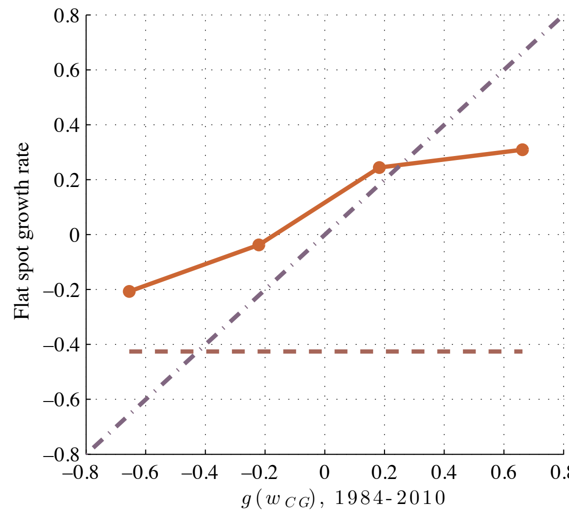
Intuition. The flat spot method relies on two assumptions, $l_{s,c,t} \approx 0$ and $\delta_s = 0$, which ensure that effective time input is age-invariant. As pointed out in Section 2.8, both assumptions are violated for values of α_s that are not close to 1.

When α_s is low, training time remains high until near retirement. Its decline then contributes to wage growth, which causes the flat spot method to overstate the growth rate of skill prices. However, when skill prices grow rapidly, the model must generate declining time inputs, $h_{s,c,t}(1 - l_{s,c,t}/\ell_{s,c,t})$, near retirement in order to match observed wage profiles. It does so by choosing a high depreciation rate δ_s, which causes the flat spot method to understate the growth rate of skill prices. The overall bias of the flat spot method is therefore ambiguous.

[13] The flat spot age ranges are 44-52 for HSD, 46-54 for HSG, 48-56 for CD, and 50-58 for CG.



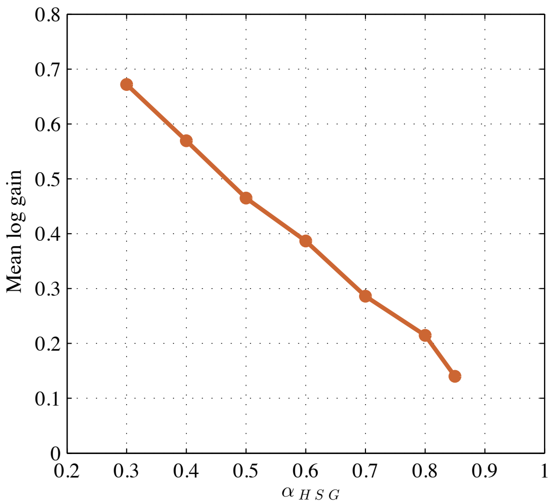
(a) High school graduates



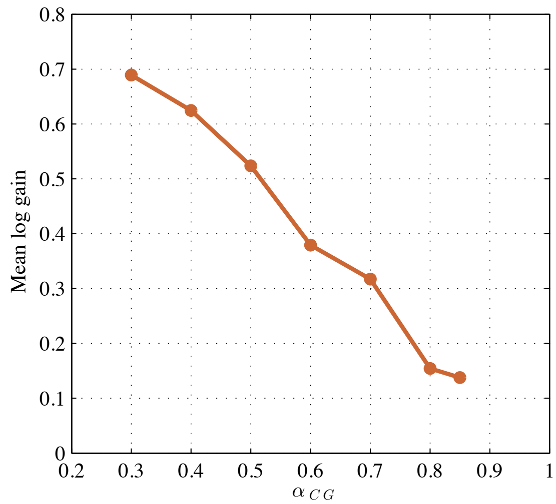
(b) College graduates

Figure 16: Flat spot versus true skill price growth rate

Notes: Each data point represents the average skill price growth rate estimated from model generated wage profiles using the flat spot method. The model is calibrated for fixed values of Δw_s.



(a) High school graduates



(b) College graduates

Figure 17: Gains from training and α_s

Notes: For each value of α_s the figure shows the lifetime earnings gain from training, measured by the change in log lifetime earnings relative to zero training time, averaged across birth cohorts.

4.2 Training and Lifetime Earnings

Kuruscu (2006a) studies how job training affects lifetime earnings. His model features a general job-training technology that nests the Ben-Porath model as a special case. It is calibrated to match mean age-wage profiles (in levels, not logs) from NLSY79 data, assuming constant skill prices, no depreciation, and no human capital investment past age 40. This approach follows Heckman, Lochner, and Taber (1998). Kuruscu arrives at a striking conclusion: job training has essentially no effect on lifetime earnings.

Figure 17 revisits Kuruscu's finding in the model developed in Section 3, though similar results are obtained when cohort effects are unrestricted as in Section 2. Relative to Kuruscu, I use wage data for more cohorts and for older workers, which allows me to relax his maintained assumptions. The gains from training are measured as the change in log lifetime earnings relative to the case when training is set to zero. Gains are averaged across members of all birth cohorts within a given school group.

The main finding is that the gains from training decline with the curvature parameter α_s. To show this, panel (a) of Figure 17 shows the average training gains for high school graduates for a grid of α_{HSG} values. Panel (b) shows the same for college graduates for a grid of α_{CG} values.

For high values of α_s, training gains fall as low as 10 log points, which is at the upper end of Kuruscu's estimates. However, for values of α_s where the model fits the data well (below 0.6 for high school graduates and below 0.7 for college graduates), the training gains rise to 30-65 log points. While fully reconciling my findings with Kuruscu's is beyond the scope of this paper, two points are worth noting. First, Kuruscu 2006b points out that stronger curvature of the human capital technology implies larger gains from training. However, he argues that a model with α_s substantially below 1 cannot fit the NLSY wage data, given the maintained assumptions of constant skill prices, no depreciation, and no investment past age 40. Second, when Kuruscu 2006b estimates his model with a Ben-Porath technology he finds values of α_s above 0.9. This renders his small estimated gains from training consistent with my model. However, with high values of α_s the model cannot account for the age-wage profiles observed in CPS data.

4.3 Sensitivity to Shocks

In this section, I present a final implication of models with near linear technologies that could potentially be used to estimate α_s: human capital investment is extremely sensitive to interest rates and skill prices.

Consider a worker who lives for T periods, faces skill prices that grow at the constant rate g_w, and is initially endowed with h units of human capital. With constant skill price growth, the optimal investment condition (5) becomes

$$\log(l) = \frac{1}{1-\alpha} \log \left[(1+g_w) \frac{\theta\alpha}{R} \frac{D^T - 1}{D - 1} \right] - \log(h) \tag{18}$$

where $D = (1+g_w)(1-\delta)/R$. An increase in g_w or a reduction in R increase D. The elasticity of training time with respect to D is proportional to $1/(1-\alpha)$, which is large when the Ben-Porath technology is close to linear.

To assess the sensitivity of study time, I fix several parameters at values that are consistent with the models calibrated in this paper. The interest rate is set to $R = 1.04$. Skill prices are constant. The length of work life is set to $T = 40$. The values of θ and h do not affect the results. Human capital depreciation is fixed either at $\delta = 0$, in line with the "emerging consensus," or at $\delta = 0.05$, which is close to the calibrated values reported above. The shock is a permanent reduction of R by by -0.005. From the optimal investment condition (5), it is apparent that reducing R has the same effects as increasing the growth rate g_w.

Figure 18 shows the percentage change in training time for a range of curvature parameters α.[14] For values of α near 0.6, which is close to the calibrated value in the model with school choice, training time increases by roughly 20 log points. However, the elasticity of training time with respect to R increases dramatically for high values of α, especially when $\delta = 0$. For $\alpha = 0.9$, training time rises by 100 log points in response to a fairly moderate change in investment incentives.

It is useful to consider what such models would imply for wage data in a world with small, persistent shocks to interest rates or wage growth rates. Upon impact, a -0.005 shock to the interest rate would reduce the wages of young workers to zero as they devote all their time to training. The response of older workers, with smaller T, would be much smaller and their wages would be far smoother.

In a sense, this is not a new result. Manuelli and Seshadri (2010) develop a Ben-Porath model with school choice to study the sources of large cross-country income differences. The model gives rise to a closed form solution, which implies that exogenous variation in total factor productivity is amplified by human capital accumulation. The size of this multiplier, i.e. the elasticity of human capital at the end of schooling with respect to total factor productivity, is proportional to $1/(1-\alpha)$, as it is in the optimal investment condition (5). Manuelli and Seshadri (2010) calibrate $\alpha = 0.936$, so that $1/(1-\alpha) = 15.6$. A 50% gap in total factor productivity is then sufficient to account for the roughly 30-fold income gaps observed in the data.

The results presented here suggest that α is no larger than 0.7, in which case $1/(1-\alpha)$ shrinks to 3.3. Far smaller values of α are also consistent with the data, in which case the multiplier shrinks further. While it is not possible to calculate exactly how α affects Manuelli & Seshadri's results without recalibrating all of the model parameters, their main conclusion is clearly sensitive to the assumed value of α.

[14] This does not impose the constraint $l \leq \ell$ as the level of training time depends on the unspecified parameters θ and h.

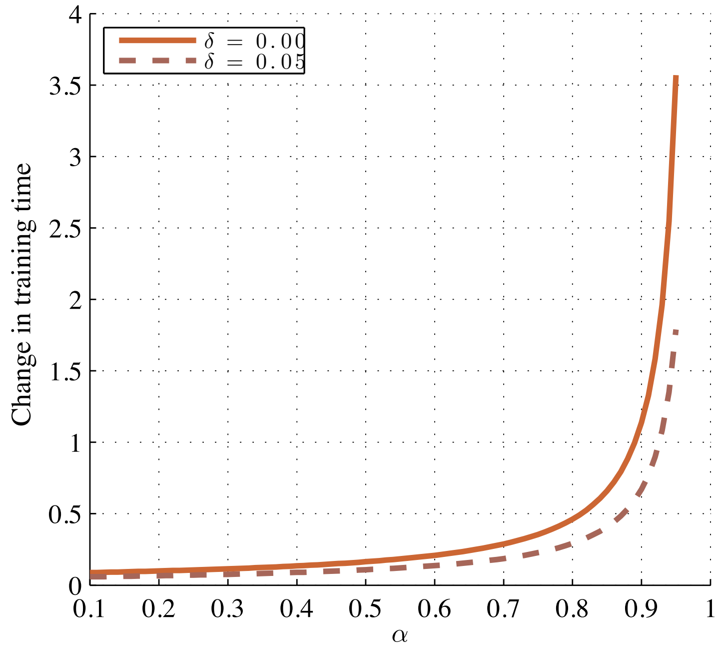


Figure 18: Response of training time to changing R

Notes: The figure shows the change in log training time in response to a permanent reduction of R by -0.005.

5 Conclusion

The results presented in this paper raise a number of questions. If mean age-wage profiles are not enough to identify all parameters of the Ben-Porath model, how can progress be made? One approach, pursued in Hendricks (2012), is to consider higher moments of the wage distribution and direct measures of cognitive abilities. Another potentially fruitful approach would model the determination of skill prices. Heckman, Lochner, and Taber (1998) have taken this approach in a model of skill-biased technical change. Guvenen and Kuruscu (2010) draw on data on the dispersion of wages to estimate cohort endowments and skill prices. Both approaches take the Ben-Porath parameters from separate sources and assume near linear technologies. Both could be extended to jointly estimate the Ben-Porath parameters and skill prices.

The finding that a wide range of skill price growth rates is consistent with nearly identical age-wage profiles calls into question previous results about the evolution of skill prices. To single out one example, Card and Lemieux (2001) show that the rise in the college wage premium during the 1980s was largely confined to young workers. Their interpretation is that young and old workers are imperfect substitutes. Figure 19 shows how the model of Section 2 interprets their result. I calibrate the model, fixing $\alpha_s = 0.40$ and $\Delta w_s = 0.0\%$ (of course, these values do not matter). In each year, I calculate the mean log wage of young persons, aged 26-30, and of older workers, aged 46-60, for both the data and for model generated wage histories. Since the model covers only a limited set of cohorts, I cannot compute the wages for older workers before 1980 and the wages for younger workers after 1997.

As expected, the model closely matches what Card and Lemieux (2001) report in their Figure 1. However, the interpretation is entirely different. By construction, model workers of all ages are perfect substitutes. What drives the rise in the relative college wage premium among young workers is not an increase in relative skill prices, but an increase in effective time inputs. This example illustrates how difficult it is to draw inferences about skill prices in a world where human capital investment decouples them from observed wages.

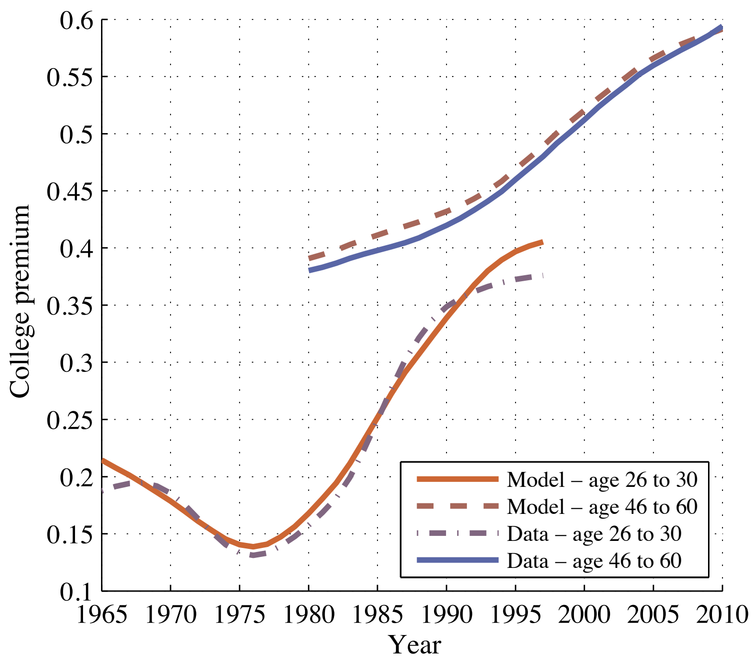


Figure 19: The college wage premium for young and old workers

References

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): “Trends in U.S. Wage Inequality: Revising the Revisionists,” *Review of Economics and Statistics*, 90(2), 300–323.

BERMAN, E., J. BOUND, AND Z. GRILICHES (1994): “Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers,” *The Quarterly Journal of Economics*, 109(2), 367–397.

BERMAN, E., J. BOUND, AND S. MACHIN (1998): “Implications of Skill-Biased Technological Change: International Evidence,” *The Quarterly Journal of Economics*, 113(4), 1245–1279.

BOWLUS, A. J., AND C. ROBINSON (2012): “Human Capital Prices, Productivity, and Growth,” *The American Economic Review*, 102(7), 3483–3515.

BROWN, C. (1976): “A Model of Optimal Human-capital Accumulation and the Wages of Young High school Graduates,” *The Journal of Political Economy*, 84(2), 299–316.

BROWNING, M., L. P. HANSEN, AND J. J. HECKMAN (1999): “Micro Data and General Equilibrium Models,” vol. 1, Part A of *Handbook of Macroeconomics*, chap. 8, pp. 543 – 633. Elsevier.

CARD, D., AND T. LEMIEUX (2001): “Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis,” *The Quarterly Journal of Economics*, 116(2), pp. 705–746.

GOLDIN, C., AND L. F. KATZ (2008): *The Race between Education and Technology*. Harvard University Press.

GREENE, W. (2012): *Econometric Analysis*. Prentice Hall, 7 edn.

GUVENEN, F., AND B. KURUSCU (2010): “A Quantitative Analysis of the Evolution of the US Wage Distribution: 1970-2000,” *NBER Macroeconomics Annual*, 24(1), 227–276.

HALEY, W. (1976): “Estimation of the Earnings Profile from Optimal Human Capital Accumulation,” *Econometrica*, 44(6), 1223–1238.

HECKMAN, J. J. (1976): “A Life-Cycle Model of Earnings, Learning, and Consumption,” *Journal of Political Economy*, 84(4), pp. S11–S44.

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), 1–58.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. Hanushek, and F. Welch, vol. 1, chap. 7, pp. 307–458. Elsevier.

Hendricks, L. (2012): "The Evolution of U.S. Wages: Skill Prices versus Human Capital," Manuscript. University of North Carolina.

Hendricks, L., and T. Schoellman (2011): "Student Abilities During the Expansion of U.S. Education, 1950-2000," Mimeo. University of North Carolina.

Huggett, M., G. Ventura, and A. Yaron (2006): "Human Capital and Earnings Distribution Dynamics," *Journal of Monetary Economics*, 53(2), 265–290.

——— (2011): "Sources of Lifetime Inequality," 101, 2923–54, Mimeo. Georgetown University.

Jaeger, D. A. (1997): "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers," *Journal of Business and Economic Statistics*, 15(3), pp. 300–309.

King, M., S. Ruggles, J. T. Alexander, S. Flood, K. Genadek, M. B. Schroeder, B. Trampe, and R. Vick (2010): "Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]," Minneapolis: University of Minnesota.

Kuruscu, B. (2006a): "Training and lifetime income," *The American Economic Review*, 96(3), 832–846.

——— (2006b): "Training and lifetime income," Manuscript. University of Toronto.

Laitner, J. (2000): "Earnings within Education Groups and Overall Productivity Growth," *The Journal of Political Economy*, 108(4), pp. 807–832.

MANUELLI, R., AND A. SESHADRI (2010): "Human Capital and the Wealth of Nations," Mimeo. University of Wisconsin.

MCGRATTAN, E. R., AND E. C. PRESCOTT (2000): "Is the Stock Market Overvalued?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 24(4).

MEGHIR, C., AND E. WHITEHOUSE (1996): "The Evolution of Wages in the United Kingdom: Evidence from Micro Data," *Journal of Labor Economics*, 14(1), 1–25.

TABER, C. (2002): "Tax reform and human capital accumulation: Evidence from an empirical general equilibrium model of skill formation," *Advances in Economic Analysis & Policy*, 2(1), 1–36.

THORNTON, R., J. RODGERS, AND M. BROOKSHIRE (1997): "On the interpretation of age-earnings profiles," *Journal of Labor Research*, 18(2), 351–365.

Table 3: Summary statistics for CPS data

Year	N	Avg N per cell	N range
1965	26226	212	41 - 500
1970	23751	192	52 - 383
1975	21716	175	63 - 324
1980	30071	243	90 - 446
1985	27672	223	77 - 461
1990	29250	236	90 - 553
1995	25313	204	85 - 396
2000	43699	352	112 - 657
2005	41778	337	94 - 552
2010	41055	331	102 - 544

Notes: N is the number of observations. Avg N per cell refers to the average number of observations in each (age, school) cell. N range shows the minimum and maximum number of observations in each cell. Cells cover age range 30-60.

A Appendix: CPS Data

A.1 Sample

The sample contains all men between the ages of 18 and 65 observed in the 1964-2011 waves of the March Current Population Survey. The data are obtained from King, Ruggles, Alexander, Flood, Genadek, Schroeder, Trampe, and Vick (2010). Summary information is provided in Table

A.2 Individual Variables

The construction of individual variables is based on Bowlus and Robinson (2012). As discussed in Jaeger (1997), the coding of schooling changes in 1991. I use the coding scheme proposed in his tables 2 and 7 to recode HIGRADE and EDUC99 into the highest degree completed and the highest grade completed.

Hours worked per year are defined as the product of hours worked last week and weeks worked last year. Weeks worked per year are intervalled until 1975. Each interval is assigned the average of weeks worked in years after 1975 in the same interval. Until 1975, hours worked per week are only available for the previous week (HRSWORK). I regress hours worked on years of schooling and a quadratic in experience to impute hours worked. After 1975, I use usual hours worked per week (UHRSWRK).

Income variables: Labor earnings are defined as the sum of wage and salary incomes (INCWAGE). Wages are defined as labor earnings divided by hours worked. Wages are set to missing if weeks worked are below 20 or hours worked per week are below 20. Outliers with less than 5% or more than 100 times the median wage are dropped.

Income variables are top-coded. As discussed in Bowlus and Robinson (2012), the frequency of top-coding and the top-coded amounts vary substantially over time. In addition, top-coding flags contain obvious errors. In most years, fewer than 2% of labor earnings observations appear to be top-coded. Following Autor, Katz, and Kearney (2008), I multiply top-coded amounts by 1.5 in years before 1988. From 1996 onwards, top-coded amounts are set to the average of all values above to top code. I leave these value unchanged. Between 1988 and 1995 there is no clear way of identifying top-coded values in IPUMS data because INCWAGE is the sum of two variables with different top codes. In these years I leave top-coded values unchanged.

To avoid top-coding issues, I drop the top 2% of wage observations from the data and from the simulated model data in each year when computing wage statistics (such as mean log wages). Since Bowlus and Robinson (2012) find that allocated values have little effect on the constructed wage series, I do not exclude them. Dollar values are deflated using the Consumer Price Index (all items, U.S. city average, series Id: CUUR0000SA0; see bls.gov).