"Golden Ages": A Tale of Two Labor Markets*

Hanming Fang† Xincheng Qiu‡

July 30, 2020

Abstract

We document stark differences in the labor market outcomes between the U.S. and China, the largest two economies in the world, during the past 30 years. (1) The peak age in cross-sectional age-earnings profiles stays constant at around 45-50 years old in the U.S. but decreases sharply from 50 to 35 years old in China. (2) Age-specific labor earnings grow drastically in China but almost stagnate in the U.S. (3) The cross-sectional and life-cycle age-earnings profiles look remarkably similar in the U.S. but differ substantially in China. We provide a unified decomposition framework to infer life cycle human capital accumulation, inter-cohort human capital growth, and human capital price changes from repeated cross-sectional data and address the above facts. We apply the framework to several applications.

Keywords:
JEL Codes:

*We would like to thank Xiaodong Zhu and the seminar participants at Fudan University, Peking University, Penn, Nankai University and University of Pennsylvania for useful comments and suggestions. All remaining error are ours.

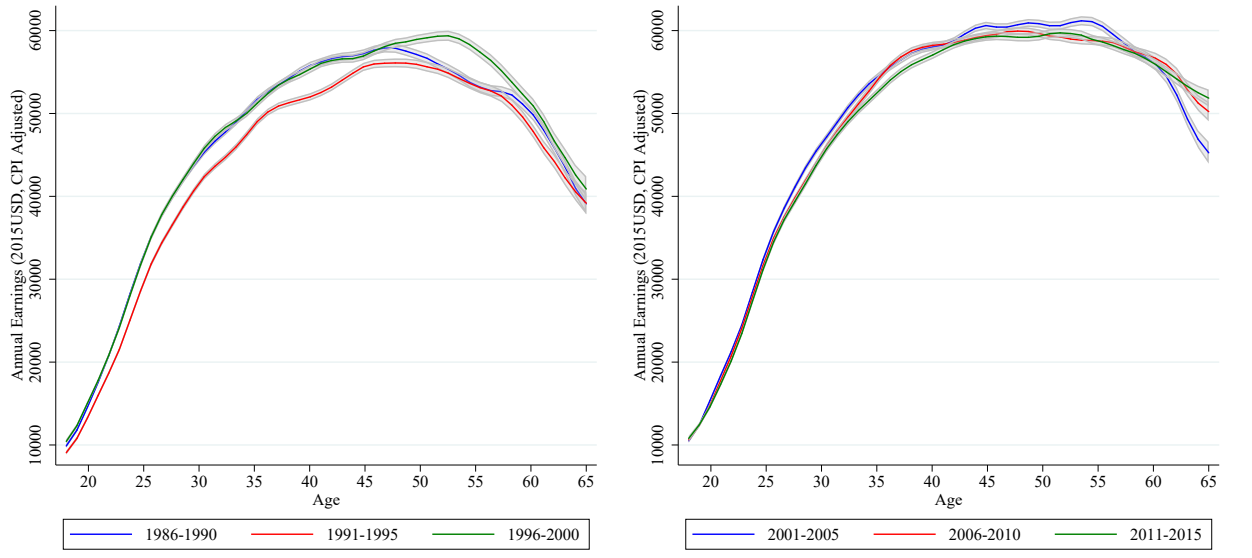
†Department of Economics, University of Pennsylvania, Ronald O. Perelman Center for Political Science and Economics, 133 S. 36th Street, Philadelphia, PA, United States, 19104; and the NBER. Email: hanming.fang@econ.upenn.edu.

‡Department of Economics, University of Pennsylvania, Ronald O. Perelman Center for Political Science and Economics, 133 S. 36th Street, Philadelphia, PA, United States, 19104. Email: qiux@sas.upenn.edu.

1 Introduction

The cross-sectional age-earnings profile is one of the most empirically examined objects in labor economics, dating back at least to Mincer (1974). It exhibits a robust regularity of a hump shape – earnings are low for young workers who have just entered the labor market, then rise with age, but at some point stop growing, and eventually decline after reaching the peak earnings age. Figure 1, constructed using data from the March Current Population Survey (CPS) obtains from IPUMS, depicts the typical cross-sectional age-earnings profiles of male workers in the U.S. from 1986 to 2015, pooling five years of CPS in the construction of each curve.[1] In this paper, we define "*golden age*" as the age group that achieves the highest mean earning in a cross-sectional age-earnings profile. For example, a "golden age" of 50 in 2010 means that in the 2010 cross-sectional age-earning profile, 50-year-old workers have the highest average earnings among all age groups. Figure 1 reveals that, first, the "golden age"in the U.S. stays almost constant over time at around 50 years old; second, the U.S. has witnessed little growth in age-specific average real earnings during the past three decades.

Figure 1: Cross-Sectional Age-Earnings Profiles of U.S. Male Workers



Source: Calculated from the Current Population Survey (CPS)

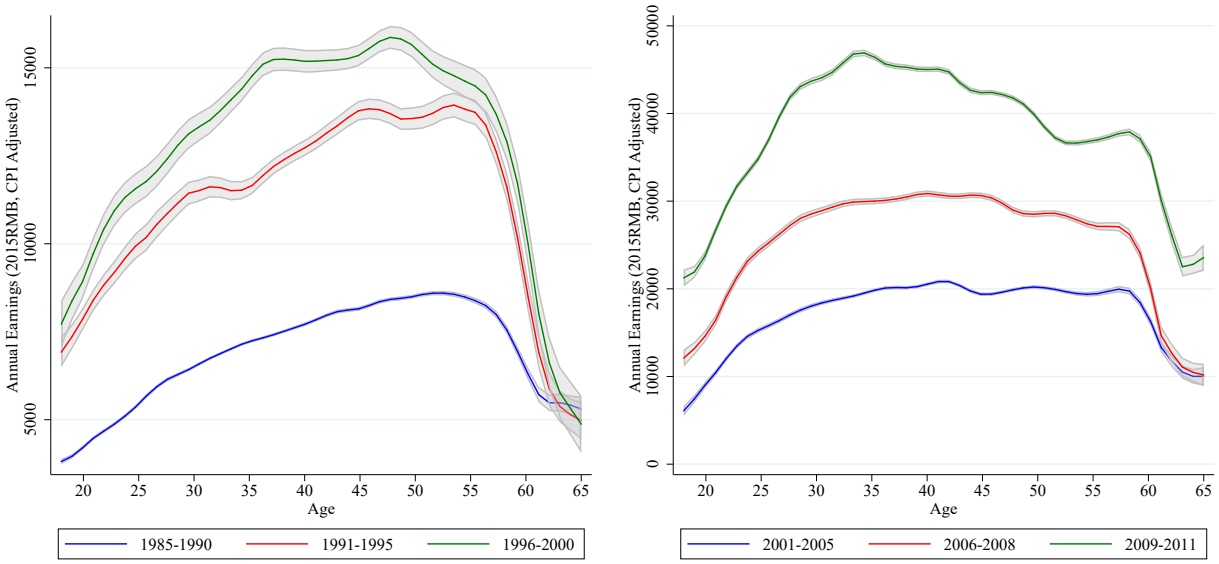
In Figure 2, we plot the cross-sectional age-earnings profile for Chinese urban male workers, using repeated cross-section data from the Urban Household Survey (UHS) of China.[2] There are several striking contrasts between Figures 1 and 2. First, Chinese

[1] See Flood et al. (2018) for a description of the IPUMS data.

[2] Both Figure 1 and 2 display the predicted values, together with the (shaded) 95% confidence interval

workers have experienced dramatic increase in real earnings in the last thirty years for all age groups. In the last thirty years, the real wages of Chinese urban male workers increased by nearly 6 folds.[3] This is in marked contrast to the real wage stagnation in the U.S. Second, while in the U.S., the shape of the cross-sectional age-earnings profiles and thus the corresponding "golden ages" have stayed more or less constant in the last thirty years, the "golden age" of China is continuously evolving to younger ages. Before 2000, the age-earnings profiles of China had a "standard" hump-shape with the "golden age" at around 50. There is, however, already some evidence of a declining "golden age" that appeared in 1996-2000. Between 2001 and 2005, the age-earnings profile is almost flat.[4] The profile humps at around age 40 in 2006-2008 and at around 33 in 2009-2011.[5]

Figure 2: Cross-Sectional Age-Earnings Profiles of Chinese Urban Male Workers



Source: Calculated from the Urban Household Survey (UHS)

In Figure 3, we plot the evolution of the cross-sectional "golden age" in the U.S. and China during 1985-2011. We run kernel regressions of log earnings on age separately for each country and each year, and calculate the "golden age" as the age achieving the maximal predicted values. The 95% confidence intervals constructed as the 2.5th and

bands. The predicted values are smoothed from a Kernel regression of labor earnings on age, where the Epanechnikov kernel function is applied, and a rule-of-thumb (ROT) bandwidth estimator is calculated. Earnings in the top 2% or bottom 2% are excluded to avoid the potential effects of extreme values.

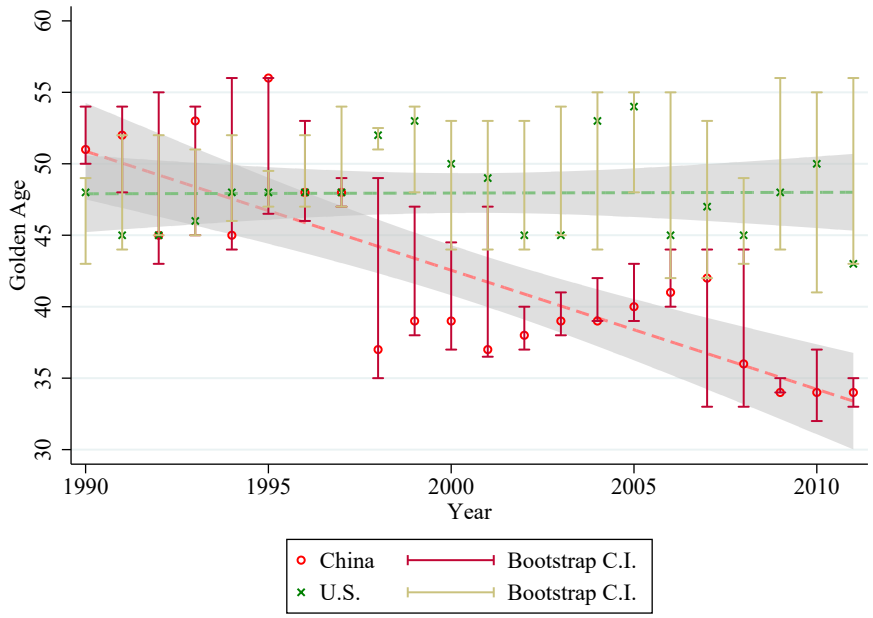
[3]Note that the scale of the left and right panels of Figure 2 differ.

[4]Cai et al. (2014) also plotted the income profile in 2002 and 2007 using data from the Chinese Household Income Project (CHIP), and they notice an earlier arrival of peak earning age.

[5]Throughout this paper, a year refers to the year to which the income variable corresponds, not the period in which the survey was conducted. For example, year 1999 corresponds to the income information in CPS 2000 and UHS 2000.

97.5th percentiles of the bootstrapped estimates are plotted. In the last thirty years, the "golden age" stays around 48 years old in the U.S., but decreases from 50 years old to 35 years old in China during the same period. Linear time trends of the golden ages with 95% confidence intervals (with point estimates taken as data) are also presented. They show clearly that in the U.S. the golden ages stayed constant, while in China it showed clear downward trend from 1985-2011. [6]

Figure 3: Evolution of Cross-Sectional "Golden Age" in the U.S. and China



Cross-Sectional Age-Earnings Profile vs. Life-Cycle Earnings Profile. Conceptually, a cross-sectional age-earnings profile, which summarizes earnings of workers of different ages at a given point of time, does not coincide with the life-cycle earnings profile, which tracks earnings of a typical person over his life course. On the right panel of Figures 4 and 5, we respectively report the life-cycle earnings path for *median* American and Chinese male workers of various birth cohorts.[7] On the left panels, we report the cross-sectional age earnings profile based on samples in the last thirty or so years using median wages (instead of mean wages as we showed in Figures 1-2). Surprisingly, in the

[6]The same patterns hold if we focus on age-specific medians –which are less sensitive to changing top labor income inequality or fluctuations in non-employment–instead of age-specific means. See the left panels of Figure 4 and 5, which plots the median earnings by two-years age bins. Again, in the U.S., workers at around 50 years old earn the highest median labor income. In China, it was the same case back to the 1990s, but now it is the 35-year-old men who make the most median annual earnings.

[7]The labels are for 10-year bins. For instance, cohort 1930s represents those who were born between 1930 and 1939.

Figure 4: Cross-Sectional and Life-Cycle Age-(Median)Earnings Profiles (U.S. Male)

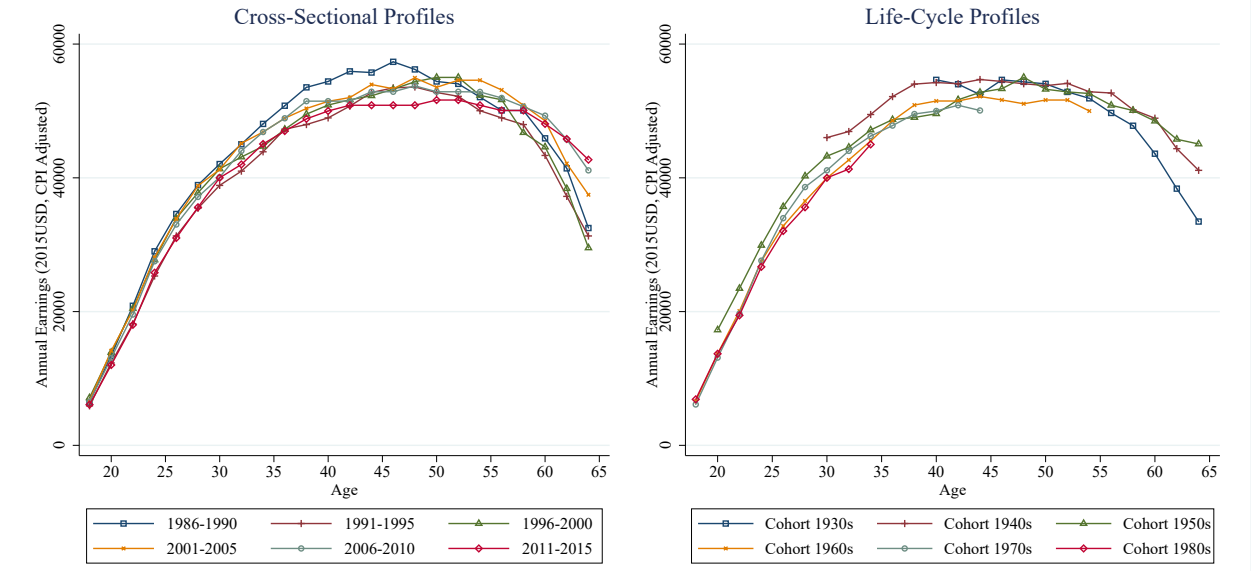
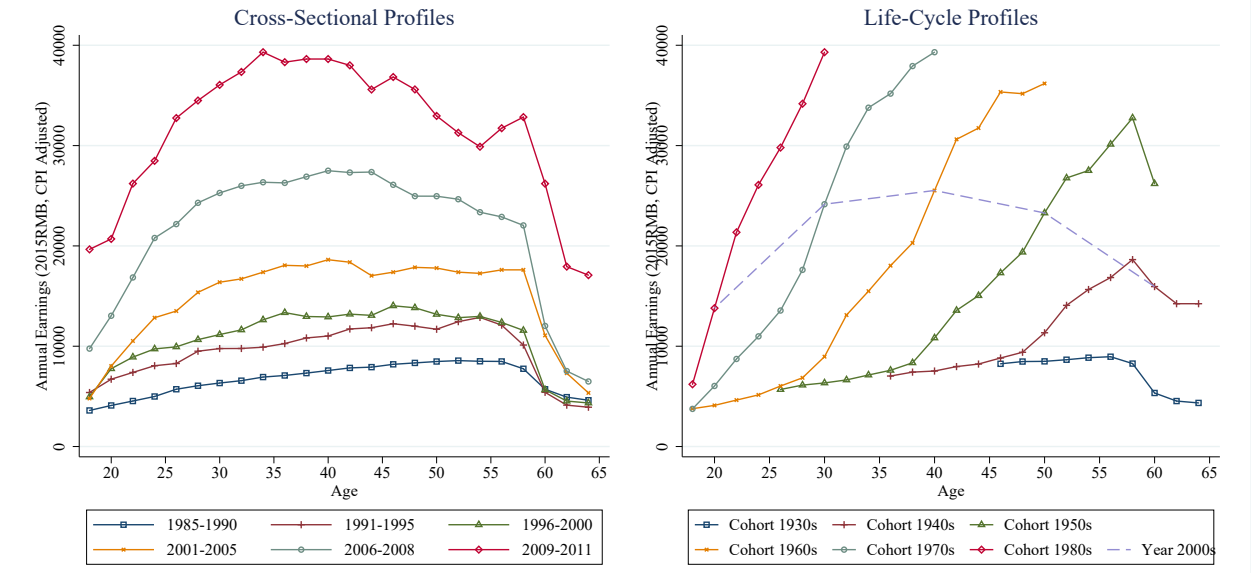


Figure 5: Cross-Sectional and Life-Cycle Age-(Median)Earnings Profiles (Chinese Male)



U.S., since cohorts expanding half a century share remarkably similar life-cycle earnings paths, the life-cycle earning profiles on the right panel of Figure 4 resemble the cross-sectional profiles on the left panel. In a stationary environment where the life-cycle profile does not vary across cohorts, the cross-sectional profiles and the life-cycle profiles essentially coincide with each other. In this economy, a median 20-year-old worker who wants to predict his earnings 20 years later can simply take a look at the contemporary

earnings of a median 40-year-old worker.[8]

However, as shown in Figure 5, the life-cycle patterns of different cohorts and the cross-sectional age-earnings profiles demonstrates no resemblance at all for China, which as a fast-changing economy stationarity is not a valid approximation. The right panel of Figure 5 shows that, more recent cohorts enjoy both much higher earnings and steeper life-cycle earnings paths. Notice that, for example, cohort born in the 1980s is around 20 years old in year 2000s, cohort of 1970s around 30 years old in year 2000s, and cohort 1960s, 1950s, 1940s are about 40, 50, 60 years old in year 2000s, respectively. The dashed line connects these corresponding points in the life-cycle profiles of various cohorts and yields a similar picture to the cross-sectional profile in 2000s, as expected. In addition, a common feature for all cohorts from cohort 1950s to cohort 1980s is that the fastest wage growth over their life-time happened during the same calendar period, around year 2000-2010. This suggests a substantial (common) time effect acting on workers of all ages.

We summarize the striking differences between the US and the Chinese labor market in the last thirty years, as the following three stylized facts:

1. The cross-sectional "golden age" stays stable at around 45-50 years old in the U.S. but continuously decreases from 50 to 35 years old in China.

2. Age-specific labor earnings almost stagnate in the U.S. but grow drastically in China.

3. The cross-sectional and life-cycle age-earnings profiles look remarkably similar in the U.S. but differ substantially in China.

In this paper, we seek to uncover the causes for the above differences between the two labor markets. To this end, we provide a framework to decompose the repeated cross-sectional age-earnings data into *experience*, *cohort*, and *time* effects, where:

- the experience effects capture how individuals' human capital accumulates with experience through his life-cycle;

- the cohort effects capture inter-cohort human capital growth, or, the relative human capital level of a cohort of workers at the time when they enter the labor market;

- the time effects capture the human capital rental prices at a given time, which of course, may change over time.

[8]Figure 4 thus provides a justification for voluminous previous works that use cross-sectional profiles as approximations to life-cycle patterns, although conceptually it is not correct to read cross-sectional profiles as life-cycle patterns, in practice they are close enough to each other in the U.S. and hence stationarity is an acceptable assumption.

As is well-known (and we will show below), without further restrictions, these three factors cannot be separately identified due to perfect collinearity. The identifying assumption we adopt is that there is no human capital accumulation (or experience effect) at the end of working life, as implied by the standard human capital investment theory (Ben-Porath (1967)). This identification idea is exploited originally by Heckman, Lochner and Taber (1998), and more recently also by Lagakos et al. (2018), Bowlus and Robinson (2012), and Huggett, Ventura and Yaron (2011). We show that under the above-mentioned identifying assumption, we can use repeated cross-sectional age-earnings profiles to separately identify these three effects, which in turn allows us to simultaneously account for the three stylized facts regarding the differences in the evolutions of the U.S. and Chinese labor markets in the last thirty years.

First, the"golden age' in a cross-sectional age-earnings profile' is essentially determined by the race between returns from the experience accumulation (the experience effect) and the inter-cohort human capital growth (the cohort effect). When the experience effect dominates, the "golden age" tends to be old. When the inter-cohort human capital growth prevails, the "golden age" tends to be young. We find that, in China the rapid inter-cohort human capital growth wins the race against the experience effect, and as a result, the golden ages in China has experienced a gradual decline in the last thirty years. In contrast, in the U.S. the two forces form a stalemate, resulting in very stable golden ages.

Second, we find that the rental rate to human capital (the time effect) is increasing much faster over the last thirty years in China than the U.S. (In fact, we find that the real price to human capital is declining by about 1% per year in the U.S.) Moreover, China experienced much higher inter-cohort human capital growth (the cohort effect) than the U.S. Both contribute to the much faster growth in age-specific labor earnings in China.

Lastly, we find that the cohort effect and the time effect are negligible in the U.S. compared to the experience effects. That is, life cycle human capital accumulation is the main driving force of both cross-sectional and life-cycle age-earnings profiles for U.S. As a result, they are close to experience effects and hence look similar to each other. In China, however, both the cohort and the time effects are substantial in the last thirty years. Thus, stationarity is lost and the life-cycle earnings profiles are very different from cross-sectional age-earnings profiles.

We also use our decomposition to fine tune some important accounting exercises in macroeconomics and labor economics. First, our decomposition separately estimates experience and cohort effects, which together constitute a notion of effective human capital quantities. This allows us to conduct a growth accounting exercise that properly accounts for the evolution of human capital. Adjusting for the changes of effective hu-

man capital, we obtain a series of estimated total factor productivity (TFP) growth lower than previous estimates. Second, we further decompose human capital rental prices into the marginal product of human capital, which is affected by both TFP and capital intensity, and a wedge which we interpret as a proxy measure of labor market frictions. Our results show that most of the decreases in U.S. human capital rental prices are due to the increase in the wedge; while in China, half of rise in the human capital price can be accounted for by increasing capital intensity, and one-third by the growth of its TFP. Third, we also implement the same decomposition separately for college and non-college workers. There we obtain an estimated series for skill-biased technical change where relative human capital quantities between high-skilled and low-skilled workers are allowed to change over time.

Related Literature. There is a long tradition in the economics literature that documents the age-earnings profiles. This line of research is so huge that we do not attempt to provide a thorough survey here. The age-earnings profile is an important input for many topics, including human capital, incentive contracts, labor supply, retirement, consumption and saving. More recent work includes Rupert and Zanella (2015) who document the life-cycle profiles separately for different cohorts, and find that the tracking between hour and wage profiles over the life cycle is broken for cohorts born between 1937 and 1946. Specifically, for cohorts born between 1937 and 1946, hours start to decline after age 55, while wages are still increasing until very late into one's 60s. Their finding calls for modifications to the standard life cycle model where hours profile tracks the wage profile closely. Throughout this paper, we focus on annual earnings and do not speak to implications on hours worked, mainly due to the lack of data on hours worked in UHS. Casanova (2013) who focuses on wage and earnings profiles for older workers over age 50 facing retirement decisions. She finds that once the employment status of full-time or part-time is controlled, wage does not decline as one ages. There is indeed a one-time drop in wage when one transitions from full-time work to part-time work. The same pattern also holds for hours worked and hence earnings. The observed downward sloping part in a typical age-wage profile or age-earning profile is thus a result of increasing composition of part-time workers. Without proper information on hours worked in UHS, we cannot directly address the partial retirement issue. Instead, we use median regression techniques, which could minimize the potential bias under the assumption that a median worker is working full time.

The remainder of the paper is structured as follows. Section 2 lays out the theoretical framework and discusses issues on identification. Section 3 describes the main results from the decomposition and the applications. Section 4 extends the benchmark frame-

work and presents additional results. Section 5 concludes and discusses future research directions.

2 Framework

2.1 Setup

A cohort of workers is labeled by the year when they enter the labor market. Observed earnings reflect the product of the prices of human capital and the quantity of human capital supplied, as well as a residual component. Denote $w_{i,c,t}$ the earnings of a worker i of cohort c at time t. Let the price of human capital at year t be p_t and the amount of human capital supplied by a representative worker of cohort c at time t be $h_{c,t}$. Thus we have

$$w_{i,c,t} = p_t \cdot h_{c,t} \cdot \epsilon_{i,c,t}, \tag{1}$$

where $\epsilon_{i,c,t}$ reflects the idiosyncratic component the framework is abstracting from.

Since neither price nor quantity of human capital is observed, this specification leads to a non-identification problem. A normalization alone cannot solve it, because $\{p_t, h_{c,t}\}$ are not only observationally equivalent to $\{\lambda p_t, h_{c,t}/\lambda\}$ for a constant λ (normalization), but also observationally equivalent to $\{\lambda_t p_t, h_{c,t}/\lambda_t\}$ for any arbitrary series of $\{\lambda_t\}$ (non-identification). As a consequence, we cannot say anything about how human capital price evolves over time or how workers accumulate human capital over the life-cycle, without imposing further restrictions.

For notational convenience, we can rewrite $h_{c,t} = s_c \cdot r^c_{t-c}$, where $s_c = h_{c,c}$ is the level of human capital of cohort c when they enter the labor market, and $r^c_k = h_{c,c+k}/s_c$ is the return to k years of experience for cohort c. This notation is without loss of generality. Obviously, $r^c_0 = h_{c,c}/s_c = 1$ by definition. We do not model the labor market entry decision and abstract from the difference between age and experience. In other words, workers enter the labor market at the same age[9]. Using this notation, we can decompose (log) earnings into time effects, cohort effects, and experience effects,

$$\ln w_{i,c,t} = \ln p_t + \ln s_c + \ln r^c_{t-c} + \ln \epsilon_{i,c,t},$$

where time effects p_t reflect the human capital prices, cohort effects s_c reveal cohort-specific human capital, and experience effects r^c_k are associated with the life-cycle human

[9]Later on, we allow for difference between age and experience by introducing different levels of education. But we still assume that workers with the same level education enter the labor market at the same age. That is, conditional on education, we abstract away from any other potential difference between age and experience.

capital accumulation.

A popular practice in the literature is to impose the returns to experience to be the same across cohorts, i.e. to restrict $r_k^c = r_k, \forall c$. The real benefit of restricting $\{r_k^c\}$ not to vary across cohorts is that we can get a complete estimated age profile even though every cohort may be observed for only part of their life. Without this restriction, we can only obtain few estimates and thus the exercise is not interesting, since results will be missing once a particular experience group of a given cohort is absent in the data. The loss compared to allowing the experience accumulation paths to differ across cohorts is that it does not fully capture all possible interactions between cohort and experience. It is worth noting that this restriction is not solving the non-identification problem mentioned above. We cannot isolate year effects, cohort effects, and experience effects without imposing additional assumptions, due to the perfect collinearity among year, cohort, and experience that $k = t - c$. We review the literature on experience-cohort-time identification in Subsection 2.3. In this paper, we will first report the baseline results under this restriction, but will discuss in detail the pros and cons of this restriction in Section 4.2.

2.2 Characterization

Consider a cross-sectional age-earnings profile for time t, $\{\ln w(k,t)\}_{k=0}^{R}$, where $\ln w(k,t)$ denotes the log earnings of a representative worker of experience k at time t, and k goes from 0 to the retirement experience R (set exogenously). Under the above framework, we have

$$\ln w(k,t) = \ln p(t) + \ln s(t-k) + \ln r(k),$$

where human capital price p is a function of time t, cohort-specific human capital s is a function of cohort $c = t - k$, and returns to experience is a function of experience k. The following proposition describes the monotonicity property of a cross-sectional age-earnings profile. We present the results in continuous time. The logic applies to discrete time version as well, *mutatis mutandis*.

Proposition 1. *The cross-sectional age-earnings profile $\{\ln w(k,t)\}_{k=0}^{R}$ is increasing (decreasing) in k if the rate of experience accumulation exceeds (falls behind) the rate of inter-cohort human capital growth.*

Proof. Fixing t, take the derivative of $\ln w(k,t)$ with respect to k,

$$\frac{\partial \ln w(k,t)}{\partial k} = -\frac{s'(t-k)}{s(t-k)} + \frac{r'(k)}{r(k)},$$

which is positive if $\frac{r'(k)}{r(k)} > \frac{s'(t-k)}{s(t-k)}$ and negative if $\frac{r'(k)}{r(k)} < \frac{s'(t-k)}{s(t-k)}$. □

This proposition states that the slope of a cross-sectional age-earnings profile is determined by the race between experience accumulation and inter-cohort human capital growth: at a given point of time, the relative earnings of an older cohort to that of a younger cohort is determined by two competing effects – the returns to experience and the inter-cohort growth of initial human capital. If the cohort-to-cohort human capital growth is vast, then the older cohorts tend to earn less relative to more recent cohorts. If experience acquisition dominates, then the older cohorts tend to have higher relative earnings. It also clarifies the difference between cross-sectional profiles and life-cycle profiles: the cross-sectional profiles confound the life-cycle age profiles with inter-cohort differences. The bias due to cohort effects may come from that later cohorts receive more or higher-quality education, stay in better health conditions, or be equipped with the skills to perform newer technologies.

The framework thus far does not distinguish age and experience. The cross-sectional "golden age" at time t is defined as

$$k^*(t) = \arg\max_{k \in [0,R]} w(k,t).$$

The following property, which is a straightforward corollary of Proposition 1, provides a characterization for the cross-sectional "golden age".

Corollary 1. *Suppose the cross-sectional age-earnings profile* $\{\ln w(k,t)\}_{k=0}^{R}$ *is a unimodal function in k. Then the "golden age" at time t happens at experience* k^**, such that*

$$\frac{s'(t-k^*)}{s(t-k^*)} = \frac{r'(k^*)}{r(k^*)}.$$

In other words, the cross-sectional "golden age" is when the speed of inter-cohort human capital growth exactly cancels with the rate of life-cycle experience accumulation.

2.3 Identification

Suppose one has access to a repeated cross-sectional dataset on earnings

$$\{\ln w_{i,c,t}\}_{c=t-R}^{t}, \qquad i = 1,2,\ldots,N; t = 1,2,\ldots,T,$$

where i refers to an individual observation, c cohort and t time. In line with the above decomposition framework that $\ln w_{i,c,t} = \ln p_t + \ln s_c + \ln r_{t-c}^{c} + \ln \epsilon_{i,c,t}$, we rewrite the

underlying wage process in the form of a regression,

$$\ln w_{i,c,t} = \text{cons} + \alpha_c + \beta_k + \gamma_t + \varepsilon_{i,c,t}, \tag{2}$$

where $\alpha_c, \beta_k, \gamma_t$ are vectors of cohort dummies, experience dummies and time dummies with $k = t - c$, and residuals are assumed to satisfy conditional mean independence $\mathbb{E}_i(\varepsilon_{i,c,t}|c,t) = 0, \forall c, t$. There are two issues that are worth mentioning.

First, for each indicator vector, we have to omit one category as the baseline group. All estimates for the indicator vectors are relative terms to the baseline group. For example, in the main analysis, we set the baseline group to be cohort 1935, year 1985 and experience 0. That is, the cohort effect for cohort 1970 means the initial human capital relative to that of cohort 1935, time effect for year 2000 is interpreted as the relative human capital price compared to that of year 1985, and experience effect for experience 30 refers to the experience accumulation since experience 0. The log earning of the baseline group is loaded on the constant term.

Second, $k = t - c$ holds by definition. Due to the perfect collinearity among them, cohort, experience, and time effects cannot be separately identified without further restrictions. In practice, there are cases where they are not exactly collinear. For instance, potential experience is often imputed as $\min\{\text{age} - \text{edu} - 6, \text{age} - 18\}$. Alternatively, some surveys may provide information on the year when one entered the labor market and there will be variation in experience for a given cohort at a given time, even conditional on education. In either case, however, cohort, experience, and time are still highly correlated. We are facing an issue of near multicollinearity anyway and the OLS will generate imprecise estimates. The main purpose of this section is to review possible ways to facilitate identification.

2.3.1 McKenzie (2006)

Hall (1968) and McKenzie (2006) show that second differences (and higher order differences) of these effects can be identified without any assumption, while first differences of these effects can be identified with one restriction. To see this, suppose the variable of interest is a linearly additive model of cohort, time and age effects with

$$y_{c,t} = \alpha_c + \beta_k + \gamma_t + \varepsilon_{c,t},$$

where $k := t - c$. Consider cohort c_1 at time periods t_1 and $t_2 = t_1 + 1$ and take a first difference:

$$\Delta_t y_{c_1,t_2} \equiv (y_{c_1,t_2} - y_{c_1,t_1}) = (\beta_{k_2} - \beta_{k_1}) + (\gamma_{t_2} - \gamma_{t_1}) + \Delta_t \varepsilon_{c_1,t_2},$$

where $k_1 = t_1 - c_1$ and $k_2 = t_2 - c_1 = k_1 + 1$. Similarly, consider cohort $c_0 = c_1 - 1$ at the same time periods t_1 and t_2:

$$\Delta_t y_{c_0,t_2} \equiv (y_{c_0,t_2} - y_{c_0,t_1}) = (\beta_{k_3} - \beta_{k_2}) + (\gamma_{t_2} - \gamma_{t_1}) + \Delta_t \varepsilon_{c_0,t_2},$$

where $k_3 = t_2 - c_0$. Taking a second difference of the above two first differences we have

$$\Delta_c \Delta_t y_{c_0,t_2} \equiv (\Delta_t y_{c_0,t_2} - \Delta_t y_{c_1,t_2}) = (\beta_{k_3} - \beta_{k_2}) - (\beta_{k_2} - \beta_{k_1}) + \Delta_c \Delta_t \varepsilon_{c_0,t_2}.$$

Thus the change in the slope of the age-effect profile is identified. Second differences of time and cohort effects are also identified in the same fashion.

Furthermore, by normalizing one first difference, one can recover all remaining slopes. To illustrate, say, we normalize one first difference of experience effects. Then, we can obtain all other first differences of experience effects from the identified second differences. With first differences of experience effects at hand, we can identify first differences of time effects, using the fact that the time differences of the outcome variable for a given cohort are the sum of first differences of experience effects and first differences of time effects. Similarly, we can identify first differences of cohorts, too. Hence one normalization on a first difference suffices for identification of all first differences.

In addition, normalizing one level each of two effects, one can recover all levels. But throughout this paper, what we care most are the slopes, which is the relative effects up to a benchmark group, not the levels. Hence we load the level of the benchmark group to a constant term, and aim at identifying first differences (i.e., slopes).

2.3.2 Deaton (1997)

As argued in the previous section, one normalization suffices for identification. Many papers thus proceed in this way and adopt one normalization. The consumption literature, though studies a different topic, offers one popular approach to deal with the collinearity issue. Deaton and Paxson (1994) and later Deaton (1997), in the section "Decompositions by age, cohort, and year" (page 123) of his book "*The Analysis of Household Surveys*," view year dummies as a device to capture cyclical fluctuation, with the restriction that time effects are orthogonal to a linear time trend. Aguiar and Hurst (2013) is a recent example that follows the same practice to study life cycle expenditures.

Suppose again

$$y_{i,c,t} = \text{cons} + \alpha_c + \beta_k + \gamma_t + \varepsilon_{i,c,t}.$$

In matrix form, we have

$$y = C + A\alpha + B\beta + \Gamma\gamma + \varepsilon,$$

where each row is an observation, A, B, Γ are matrices of cohort dummies, experience dummies, and time dummies, respectively, and α, β, γ are vectors of cohort effects, experience effects, and time effects, respectively. Note that the collinearity across time, cohort, and age $t = c + k$ implies

$$\Gamma s_t = A s_c + B s_k,$$

where the s vectors are arithmetic sequences $\{0, 1, 2, 3, \dots\}$ of the length given by the number of columns of the matrix that premultiplies them. Replace the parameter vectors by

$$\tilde{\alpha} = \alpha + \kappa s_c, \quad \tilde{\beta} = \beta + \kappa s_k, \quad \tilde{\gamma} = \gamma - \kappa s_t.$$

Thus an arbitrary time-trend can be added to the age dummies and cohort dummies by subtracting it from the year dummies, which sheds light on the non-identification problem. Deaton assumes that the year effects capture cyclical fluctuations or business-cycle effects. Formally, in addition to $\sum_t \gamma_t = 0$ (which is an innocuous normalization as it only adjusts the constant term), he restricts that $s'\gamma = 0$ to capture the idea that time effects are orthogonal to a linear trend. Notice that the label of years is without loss of generality, for any chronological relabel of years will still satisfy this relation.

To implement Deaton's idea, one can regress y on a set of dummies for each cohort excluding (say) the first, a set of dummies for each age excluding (say) the first, and a set of $T - 2$ year dummies defined as follows for $t = 3, .., T$,

$$d_t^* = d_t - \left[(t-1)d_2 - (t-2)d_1\right].$$

The coefficients of d_t^*'s thus give the third through final year coefficients. Then one can recover the first and second coefficients γ_1, γ_2 by solving the system of equations $\sum_t \gamma_t = 0$ and $s'\gamma = 0$.

This approach assumes that secular trends appear only in cohort effects and time effects simply reflect fluctuations. Alternatively, one could also take an opposite restriction that cohort dummies are orthogonal to the time trend. Some papers may investigate both, or a mixture of the two (e.g., Lagakos et al. (2018)), to examine the sensitivity of their results to the identifying assumption.

Another related but different approach is even simpler – instead of imposing a normalization, it directly uses observable measures as a proxy for time effects. For instance, in Gourinchas and Parker (2002) studying age-consumption profiles and Kambourov and Manovskii (2009) studying age-earnings profiles, they use unemployment rates to capture the time effects arising from booms and recessions.

2.3.3 Schulhofer-Wohl (2018)

Schulhofer-Wohl (2018) proposes an alternative method that does not require the somewhat arbitrary normalization, but virtually shifts focus from directly estimating the age effects to estimating the parameters in age effects implied by a structural model. That is, now the aim is to estimate θ in the following equation

$$y_{c,t} = \text{cons} + \alpha_c + \beta(k, \theta) + \gamma_t + \varepsilon_{c,t},$$

where $\beta(k, \theta)$ is derived from an economic model and θ is a vector of model fundamentals. To achieve identification, this approach requires the function $\beta(k, \theta)$ to be sufficiently nonlinear in k. Under this condition, θ can be estimated consistently via a minimum distance procedure. Essentially, the structural parameters are identified from second and higher derivatives of the age effects. This approach ultimately facilitates identification of structural parameters associated with age effects without first identifying the age effects.

2.3.4 Heckman, Lochner and Taber (1998)

Heckman, Lochner and Taber (1998) (HLT hereafter) deal with the non-identification issue using economic theory. The HLT identifying assumption is that there is no human capital accumulation at the end of working life. This assumption could be justified in a Ben-Porath (1967) framework, where zero on-the-job investment in that stage is the optimal choice. HLT's approach is a perfect combination of the previous two approaches. On the one hand, the identifying assumption is essentially a normalization on the first difference of experience effects. On the other hand, this restriction is coming from economic theory and can be derived in a structural model of human capital investment.

To see how this assumption facilitates separate identification of human capital prices (time effects), cohort-specific initial human capital (cohort effects), and human capital accumulation paths (experience effects), consider a cohort and a time such that this cohort is right at the end of their working life. For this cohort, the change in wages in that stage of life identifies the change in human capital prices (time effects) solely, as there is no human capital accumulation due to the identifying assumption, and there is no cohort effect since we are fixing a particular cohort. Following the same procedure repeatedly for many cohorts at the end of their working life will generate a time series of human capital prices. With the time effects at hand, we can readily identify cohort effects and experience effects. A concrete example may help. Suppose one assumes there is no human capital accumulation (say) from 49 to 50 years old. First, comparing the wages of 49-year-old workers in 2009 and 50-year-old workers in 2010 gives the year effect from

2009 to 2010 because (1) they are the same cohort so there is no cohort effect, and (2) by assumption there is no experience effect. Second, comparing the wages of a-year-old workers in 2009 and $(a+1)$-year-old workers in 2010 gives the experience effect from a to $(a+1)$ because (1) they are the same cohort so there is no cohort effect, and (2) we have already backed out the year effect from 2009 to 2010. Finally, comparing the wages of a-year-old workers and $(a+1)$-year-old workers in 2010 gives the cohort effect for $(2010-a)$ because (1) they are in the same year so there is no year effect, and (2) we have already backed out the experience effect from a to $a+1$.

Lagakos et al. (2018) adopt the HLT method as their preferred estimates for returns to experience and set the flat spot phase as from 25 years of experience to 35 years of experience. We generally follow Lagakos et al. (2018), which in turn combines the identification assumption proposed by Heckman, Lochner and Taber (1998) with the procedure laid out by Deaton (1997). The basic idea is to recast the regression of interest (2) as

$$\ln w_{i,c,t} = \text{cons} + \alpha_c + \beta_k + gt + \tilde{\gamma}_t + \varepsilon_{i,c,t},$$

where $\tilde{\gamma}$ satisfies $\sum_t \tilde{\gamma}_t = 0$ and $s'\tilde{\gamma} = 0$. That is, they rewrite an arbitrary time series γ_t as a linear trend gt plus fluctuations $\tilde{\gamma}_t$. The benefit of this manipulation is that once a value of g is guessed, we can run Deaton's procedure on the deflated wage $\ln \tilde{w}_{i,c,t} = \ln w_{i,c,t} - gt$ and get estimates for cohort, experience, time effects under this particular guess g. The time trend is thus pinned down by the HLT assumption: we update the guess of g until the associated experience effects are the same for the two experience groups late in life presumed by the HLT assumption.

They obtain the estimates using an iterative procedure. First, start with a guess for the growth rate of a linear time trend. In practice, the guess is picked as the coefficient on the linear time trend term by regressing log wage on the set of dummies for experience groups and a linear time trend. Second, deflate the wage data using the guess of growth rate. Following Deaton's procedure laid out in Section 2.3.2 but using log deflated wage as the dependent variable, we regress log deflated wage on a set of dummies for experience groups, cohort groups and d^*'s define in the previous subsection 2.3.2. Now check if the experience effects are sufficiently close between experience group 25 (or experience group 30) and experience group 35, according to a preset precision. If so, the convergence condition is satisfied. Otherwise, we update the guess for the growth rate by adding the growth rate of the currently estimated experience effects from experience group 25 (or experience group 30) to experience group 35 with a damping factor.

3 Decomposition

We will follow the insight in Heckman, Lochner and Taber (1998) to facilitate the age-cohort-time identification using implications from economic theory along the lines of Ben-Porath (1967). To start with, we perform an HLT decomposition of earnings among experience, cohort, and time effects. Specifically, we estimate the experience effects (relative to labor market entry) in 5-year bins, cohort effects (relative to the 1935-1939 birth cohorts) in 5-year bins, and year effects (relative to 1985) year by year. We generally follow Lagakos et al. (2018), but our implementation differs from theirs in several regards.

First, we focus on medians in the analysis instead of means. The benefit of working with medians is that the results are immune to various concerns. The most standard advantage of the median over the mean is that the results are not sensitive to outliers nor missing values, both of which are quite common in earnings data. Besides, median earnings are unlikely to be affected by the evolving inequality in the top or bottom. Furthermore, average annual earnings may confound the trends in part-time versus full-time works, whereas the median will not, provided that the median man is working full time and a full year. Hence in the estimation algorithm, we perform a quantile regression estimating how conditional median earnings in response to age, cohort, and time, instead of the standard OLS concerning the conditional means.

Second, they restrict the experience profiles to be the same at 25 and 35 years of experience bins in their baseline results. Although this is meant to capture the assumption that there are no experience effects over the last 10 years of working life, the return to 30 years of experience is effectively left to be whatever number generated by the algorithm and allowed to differ from returns to 25 or 35 years of experience (and indeed so in the results). They also report results based on the assumption that there are no experience effects over the last 5 years, which are obtained under the restriction that the experience profiles are the same at 30 and 35 years of experience bins. We prefer the second choice, as it is more consistent with the identifying assumption that there is no experience effect at the end of working life.

Third, potential experience is imputed as $\min\{\text{age} - \text{edu} - 6, \text{age} - 18\}$ in Lagakos et al. (2018). That is, workers with more than 12 years of schooling are assumed to start schooling at 6 years old and enter the labor market after they finish schooling, and workers with fewer than 12 years of schooling are assumed to enter the labor market at 18 years old. This assumption, together with the previous one, means the end of working life is set as 48-57 years old for high school graduates but 52-61 years old for 4-year college graduates. We consider the end of working life to be the same age range for everyone in the benchmark results by defining experience as $(\text{age} - 20)$. We do not

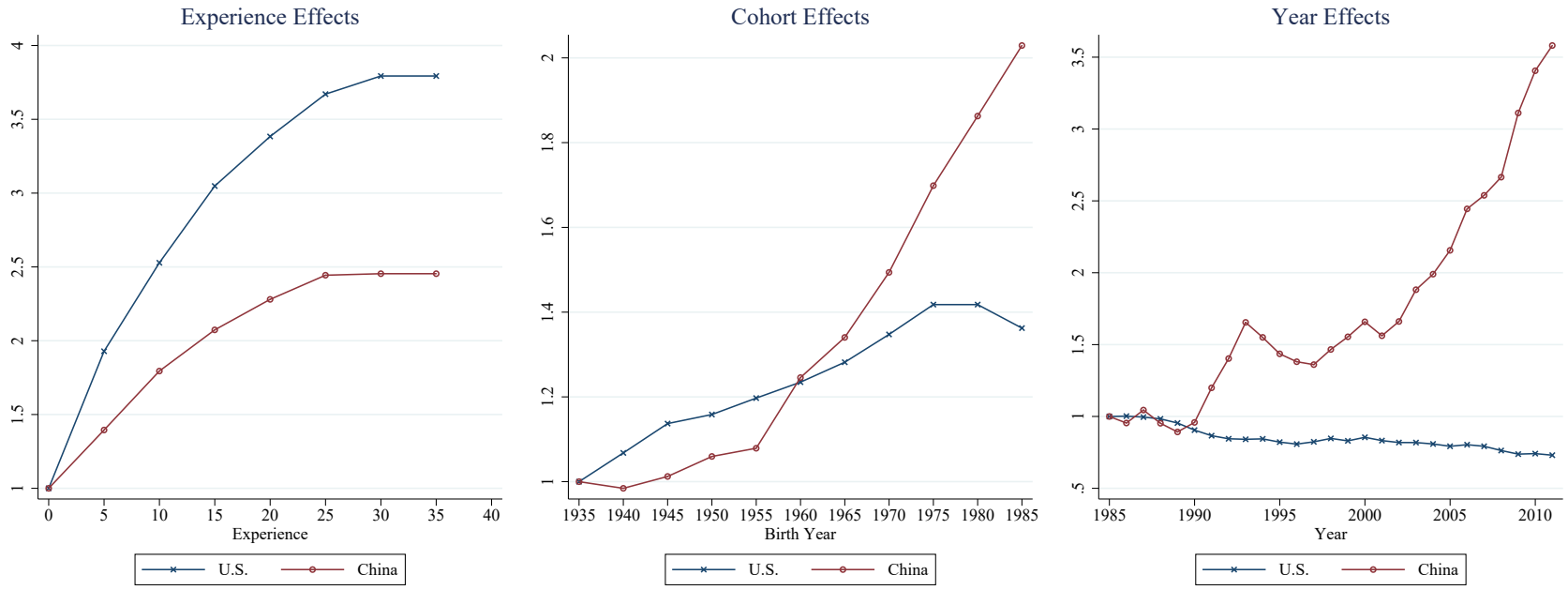
distinguish experience effects from age effects.

Finally, our goal here is not to identify the causal effect on experience on earnings but rather an accounting exercise with similar flavors in the growth accounting literature. As a first step, we do not control for years of schooling. But we do separately consider college and non-college groups in Section 4.1, which considers the case that college workers and non-college workers are not perfect substitutes.

Apart from the above differences, we adopt the same iterative procedure laid out in the previous section. Our estimates are shown in Figure 6. Figure 7 shows the model fit, where the solid lines are the data, and dashed lines are model predictions.

In general, the HLT approach could be extended to allow for two sets of preset parameters that one could pick her preferable values: (1) a human capital depreciation rate, and (2) a "flat spot" area for which workers do not invest in human capital. In particular, Heckman, Lochner and Taber (1998) assume a zero human capital depreciation rate, which is consistent with evidence documented by Browning, Hansen and Heckman (1999). We follow this assumption, as many other papers studying life-cycle human capital accumulation do (e.g., Kuruscu (2006)). As for the second preset input for the algorithm, Bowlus and Robinson (2012) attempt to determine the flat spot regions more carefully and prefer the flat spot age ranges to be 50–59 for college graduates and 46–55 for high school graduates. Note that our choice of flat spot areas are largely overlapped with theirs (52-61 for college graduates and 48-57 for high school graduates in our case).

Figure 6: Decomposition



3.1 Life Cycle Human Capital Accumulation

Experience effects are interpreted as life cycle human capital accumulation. Consistent with the finding in Lagakos et al. (2018); Islam et al. (2018) that developed countries have higher returns to experience than developing countries, we find that the U.S. exhibits higher experience effects than China, as shown in the left panel of Figure 6. For a median American male worker, the human capital supplied at the end of his working life will be nearly 4-folds of his initial human capital supplied upon entry into the labor market. In China, the accumulated return to experience for the most experienced male workers is about 2.5.

The levels are not directly comparable to the result of the U.S. reported by Lagakos et al. (2018), however. Apart from the several discrepancies in the estimation procedure mentioned above, the outcome variable they are concerning is hourly wage, constructed as labor earnings divided by the number of hours worked, while we are looking at annual earnings. In Ben-Porath (1967)'s framework, time devoted to working and learning are distinct concepts within the model, but a usual dataset cannot distinguish them. One has to take a stand on how much of the measured hours worked reflects time spent on working and learning. For example, Huggett, Ventura and Yaron (2011) assume that the measured hours worked is only work time and does not include learning time. Focusing on annual earnings here thus avoids the potential mismeasurement on time allocation.

In the classical life-cycle human capital accumulation literature, pioneered by Ben-Porath (1967); Mincer (1974), life-cycle earnings are interpreted as the amount of human capital supplied to the employers. In those models, earnings are increasing over the life cycle because (1) workers accumulate human capital to enlarge their human capital capacity, and (2) workers will invest less and hence contribute a larger fraction of their capacity to work when it approaches the end of their career. An implicit assumption, when people interpret wage changes over the life cycle as changes in the quantity of human capital supplied, is that the price of human capital is constant in different periods over the life cycle. Formally, when assuming $p_t \equiv p, \forall t$, we have

$$\frac{w_{c,t_1}}{w_{c,t_2}} = \frac{p_{t_1} \cdot h_{c,t_1}}{p_{t_2} \cdot h_{c,t_2}} = \frac{h_{c,t_1}}{h_{c,t_2}}.$$

The considerable time effects detected in Figure 2 and 5 suggest that it cannot be an innocuous assumption for the case of fast-growing economies like China. The price per efficiency unit of human capital evolves hugely instead.

How can we explain the steeper returns to experience in U.S. than in China (or more generally, in developed countries than in developing countries)? The Ben-Porath (1967) interpretation is that workers in rich countries invest more in human capital. Manuelli

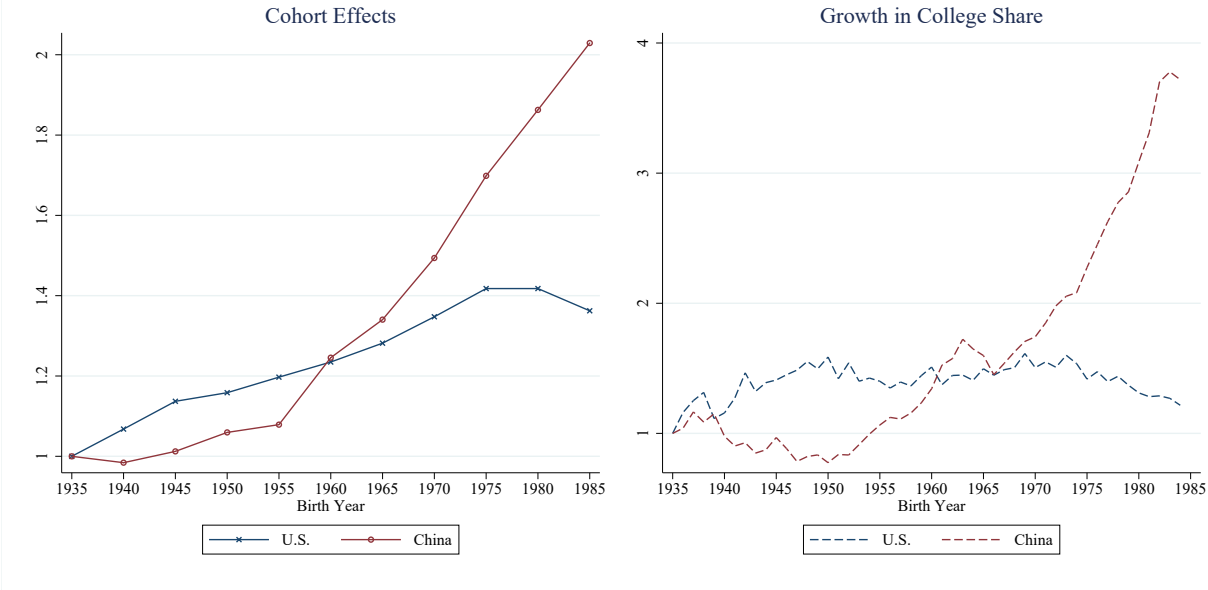
and Seshadri (2014) builds an augmented Ben-Porath (1967) model to incorporate an early childhood phase and to have both time and goods as inputs in the production of human capital. In their model, human capital investment is endogenous to TFP. In particular, higher TFP results in lower prices of market good inputs to human capital production and encourages human capital investment in rich countries. Alternative explanations include search frictions. In on-the-job search theories (e.g., Burdett and Mortensen (1998)), wage growth is a result of the arrival of new job offers and workers climbing up the job ladder. If there are more search frictions in poor countries where workers make job-to-job transitions less frequently, flatter age-earnings profiles arise in poor countries.

We propose another new, potential explanation for why experience effect is higher in U.S. than in China that relies on the multidimensionality of skills. Lise and Postel-Vinay (2019) find that the half-life of skill accumulation varies quite a lot across different types of skills. According to their estimates, the half-year is 7.5 for cognitive skills, 1.7 for manual skills, and 55.8 for interpersonal skills. The message is that it is super hard to accumulate one's interpersonal skills. If, say, job requirements in China emphasize more on interpersonal skills than the U.S., it may result in a lower experience effect. The investigation of this hypothesis is beyond the scope of the current paper, and we leave this direction for future research when there is suitable data to study heterogeneous distributions of multidimensional skills and skill requirements across countries.

3.2 Inter-Cohort Human Capital Growth

Cohort effects represent estimates for the inter-cohort growth of initial human capital upon entry into the labor market. Since the life cycle human capital accumulation is imposed to be the same across cohorts, the same numbers also reflect the inter-cohort growth of human capital at any given age, or the life-time human capital. The middle panel of Figure 6 shows that China has experienced rapid human capital growth for subsequent cohorts. While U.S. workers' human capital increase by only 40% in half a century of cohorts, the most recent cohort in China more than doubles the human capital as their counterparts 50 years ago.

Figure 8: College Share by Birth Cohort



One possible source of inter-cohort human capital growth comes from increases in educational attainment for subsequent cohorts. The right panel Figure 8 plots the cohort-by-cohort growth of the share of college graduates in the U.S. and China, and the left panel reproduces the cohort effects for comparison. The estimated cohort effects replicate to a large extent the evolution pattern of the fraction of college workers in these two countries.

Despite the rapidness of inter-cohort growth in China, the growth is unequally shared among different cohorts. Most of the growth is reaped by workers born after 1960, while a whole generation prior to that witnessed very little human capital growth. During the Cultural Revolution (1966-1976), higher education in China was suspended for 10 years. This corresponds to the meager college share among cohorts born in the late 1940s and early 1950s. An opposite narrative is that the Chinese government expand college enrollment at a large scale in 1999. In the following years, the expansion was unprecedentedly massive. There were 1.08 million students admitted by colleges in 1998. The number doubled after only two years, with 2.21 million students admitted in 2000. This is reflected in the even faster increase of college share for cohorts after around 1980.

3.3 Human Capital Price Changes

We interpret the year effects in the right panel of Figure 6 as changes in the rental price to human capital. Human capital price has increased to more than 3.5 folds its level 25

years ago in China, while there is little change in human capital prices in the U.S. If anything, the human capital price in the U.S. decreases by 25% from 1985 to 2011, which is around a 1% decrease per year.

3.3.1 Growth Accounting

One application of the estimated human capital price change series is to perform a growth accounting. Consider an aggregate production function $Y_t = A_t K_t^{\alpha} H_t^{1-\alpha}$, where Y_t is the aggregate output, K_t the aggregate physical capital, H_t the aggregate human capital, and A_t the total factor productivity (TFP) at time t. Denote lower case letters the corresponding variables in per worker terms. The output per worker can be expressed as $y_t = A_t k_t^{\alpha} h_t^{1-\alpha}$ and hence

$$\frac{\dot{y}}{y} = \frac{\dot{A}}{A} + \alpha \frac{\dot{k}}{k} + (1-\alpha)\frac{\dot{h}}{h},$$

where the parameter α is assumed to be constant[10].

The output per worker y_t and capital stock per worker k_t for each year are observed in the data. To construct these two measures, We divide the aggregate output and aggregate capital stock by the number of workers, all of which are obtained from Penn World Table 9.0 provided by the Federal Reserve Bank of St. Louis website. Our growth accounting exercise differs from standard ones only in the measure for human capital h_t. Assume everyone faces the same rental price to human capital, including both males and females. This assumption implies that our estimated series of human capital price changes from male earnings data also apply to females. As a result, the average human capital supplied by each worker in a year can be calculated as the average annual earning divided by human capital price. Finally, the distribution parameter α is calibrated to 0.4, as is standard in the literature. Therefore, TFP changes are measured as a residual.

Our method only delivers growth accounting and cannot perform levels accounting. This is because our decomposition gives an estimated series of human capital price changes, instead of human capital prices. As a result, we can only back out relative changes of human capital quantity, but not the absolute levels of human capital quantity. Our model is all about relative changes and silent on levels.

There are several attempts to measure the human capital for the input in the aggregate production function. Hall and Jones (1999) presume that aggregate human capital

[10] Elsby, Hobijn and Şahin (2013) redo the standard growth accounting exercise but take into account the change in labor share, and find that the result barely changes. Thus, they argue that approximate constancy of share parameter is still a reasonable assumption for growth accounting, albeit the declining of labor share.

is $H_i = \exp\{\phi(E_i)\} L_i$, where $\phi(E_i)$ reflects the efficiency units of labor with E years of education (relative to labor with no schooling). In practice, they estimate $\phi'(E)$ as the returns to schooling from a standard Mincerian wage regression. Bils and Klenow (2000) also exploit Mincerian returns, but extend the focus on education to include experience as well. In addition, they introduce interdependence on the human capital of older cohorts to capture the impact of teachers. The standard approach inspired by Mincer (1974) that using Mincerian rate of return to schooling to measure human capital typically finds cross-country differences in output per worker are largely driven by differences in TFP. Manuelli and Seshadri (2014), instead of relying on the Mincer equation, calibrate a model of human capital acquisition with early childhood development, schooling, and on-the-job training, and compute human capital stocks by evaluating the human capital production at the individual optimum under equilibrium prices. They find a larger role for human capital as well as a smaller role for TFP in explaining the cross-country differences in output per worker. The closest to our paper is Bowlus and Robinson (2012), which is also based on the HLT identifying assumption. The difference is they are decomposing hourly wage but we are decomposing annual earnings. The benefit of working with annual earnings, apart from data limitation of Chinese data, is that earnings changes net from human capital price changes (derived from the decomposition exercise) is directly an estimate for human capital changes as an input for the growth accounting, without any further adjustment of hours worked.

The growth accounting results suggest little TFP growth in the U.S during the past 25 years, which is consistent with the productivity slowdown (e.g. Fernald (2015)) in the same period. But our results provide even lower estimates for TFP. This is understandable because the notion for human capital here properly incorporates inter-cohort human capital improvements and life-cycle human capital accumulation, and TFP is a model-based concept. In other words, part of the improvements in human capital was attributed to TFP growth in previous estimates that do not properly adjust for human capital growth. For China, TFP increases by around 50% during the same period, almost all of which are gained after 2000. It reflects the success of the several economic reforms in the late 1990s and early 2000s. For example, state-owned enterprises (SOEs) were involved in waves of privatization and restructuring since mid-1990s under the slogan "grasp the large and let go of the small" to privatize small or medium-sized SOEs while retaining control of large ones. From 1995 to 2001, there were an estimated 34 million workers that were officially registered as laid off from the state sector. The privatization of SOEs contributes to the overall productivity of the economy (e.g. Chen et al. (2019)). Another milestone event is that China joined the WTO in 2002. Trade liberalization is often believed as one of the key drivers of firms' productivity and hence China's rapid

growth (see Brandt et al. (2017) for evidence).

The contributions of physical capital per worker, human capital per worker, and TFP to the GDP per worker growth in both countries are reported in Figure 9. For the U.S., almost all growth in output per worker comes from factor growth, where physical capital accounts for about one-third and human capital explain the other two-thrids. The story is quite different in China, where physical capital is responsible for about a half, and human capital and TFP each account for around one fourth. Our results suggest that most of the TFP growth since 1985 in China is coming from the post-2000 growth, during many prominent economic reforms took place, such as SOE privatization, trade liberalization, and massive internal migration. This is consistent with Zhu (2012), according to whom the average annual total factor productivity growth in nonagricultural sector is 2.17% and 0.27% for nonstate and state sectors during 1988-1998, but is 3.67% and 5.50% for nonstate and state sectors during 1998-2007. The numbers are not necessarily comparable as TFP are model-based notions, but we both find larger TFP growth arose after 2000.

3.3.2 Understanding Human Capital Prices

A natural question is how we should understand human capital prices. Consider the same production function as before. The marginal product of human capital is

$$MPH_t = (1-\alpha) A_t \left(K_t / H_t\right)^{\alpha},$$

and the marginal product of a worker with human capital h_{it} is $MPL_{it} = MPH_t \times h_{it}$. We further allow for a reduced-form wedge parameter μ so that wages do not necessarily equal marginal product but $w_{it} = \frac{1}{\mu_t} MPL_{it}$. This implies

$$p_t = \frac{1}{\mu_t} (1-\alpha) A_t \left(K_t / H_t\right)^{\alpha}.$$

Hence the observed changes in price to human capital are a combination of changes in capital intensity, TFP, and the wage wedge on marginal product of labor. To account for why the U.S. and China exhibit opposite patterns in human capital prices, we decompose the human capital price changes into these three factors:

$$\Delta \ln p_t = \underbrace{-\Delta \ln \mu_t}_{\text{wedge}} + \underbrace{\Delta \ln A_t}_{\text{TFP}} + \underbrace{\alpha \Delta \ln \left(\frac{k_t}{h_t}\right)}_{\text{capital intensity}}.$$

We calculate the contribution of improvements or deteriorations in capital intensity as $\alpha\Delta \ln\left(\frac{k_t}{h_t}\right)$, where k_t, h_t and α are the same as before. TFP changes, $\Delta \ln A_t$, are the residual measures from the growth accounting above. As a result, changes in wage wedge are the residuals from the estimated price changes subtracted by changes in capital intensity and TFP[11].

The decomposition is shown in Figure 10. Most of the decreases in U.S. human capital prices are due to the worsening wage wedge. The wedge is a reduced-form parameter capturing all frictions that preclude wages to be equal to marginal products. It is a mixture of effects from output market power, labor market power, as well as labor market institutions such as unions and occupational licensing. De Loecker, Eeckhout and Unger (2020) document that the average markups in the U.S. rise from 21% above marginal cost in 1980 to 61% nowadays and hence the rise in the output markup may be a significant source of the worsening wage wedge. Macaluso et al. (2019); Lipsius (2018); Berger, Herkenhoff and Mongey (2019) find that the labor market concentration in the U.S. is actually declining, which is the opposite of what is needed to account for the worsening estimated wage markdown. Kleiner and Krueger (2013) show that the share of workers holding an occupational license is increasing dramatically over time, from less than 5% in the 1950s, to around 29% nowadays, while union members account for more than 30% in the 1950s and only about 10% nowadays. To the extent that these labor market institutions pose frictions, they may also be related to changes in the wage wedge. But since these two prominent labor market institutions are changing in the opposite direction, it is possible that their effects cancel. To sum up, the existing evidence suggests that the rising output markup might be the most likely driving force of the worsening wage wedge, but more rigorous examinations are needed to make a firm conclusion. A cautionary note that we need to keep in mind is that the wedge may also come from the misrepresentation of a representative worker framework. The average marginal product of heterogeneous workers does not necessarily equal the marginal product of the average worker.

Although TFP and capital intensity do not explain much of the human capital price changes in the U.S., they together explain 85% of the human capital price increases in

[11]If one takes the model literally, the wedge μ_t is allowed to vary over time but the output elasticity α is hold fixed. As a result of this strong assumption, all changes in the labor share are attributed to wedges but not to technologies, because the model implies that the labor share of income is $s_t = \frac{1}{\mu_t}(1-\alpha)$. Although holding fixed the output elasticity is consistent with the pratice in papers that focus on, for instance, markups to explain labor share, where the potential changes in output elasticity are abstracted (De Loecker, Eeckhout and Unger (2020)), it will be at odds with papers that emphasize the role technological changes, for example, an automation process where tasks previously performed by labor are gradually taken by machines (Acemoglu and Restrepo (2018)). We thus view our model as only convenient and illustrative.

China, where half of the human capital price increases in China can be accounted for by increasing capital intensity, one-third can be explained by improving TFP.

4 Discussions

4.1 Heterogeneous Human Capital by Education

In the previous baseline estimation, we implicitly assume that human capital is homogeneous so that every worker's skill can simply be represented by a single index indicating the level of efficiency units. We now extend the framework to allow for human capital to have different types. For example, college graduates and non-college graduates may possess different types of skills that are not perfect substitutes. It is straightforward to extend our framework to allow for heterogeneous human capital across education groups so that $w_{it}^e = p_t^e \cdot h_{it}^e$, where $e \in \{\text{cl}, \text{hs}\}$ refers to an education group ("cl" stands for college and "hs" high school). We then perform the decomposition as discussed before but now separately for college workers and non-college workers. The only restriction is that for both college workers and non-college workers, there is no additional skill accumulation from experience in the flat spot region preset to be from 30-34 years of experience bin to 35-39 years of experience bin. Since we assume college graduates start to gain experience from 22 years old and high school graduates start to gain experience from 18 years old, this is essentially assuming the flat spot region is 52-61 years old for college graduates and 48-57 years old for high school graduates, which is largely overlapped with the flat spot areas determined by Bowlus and Robinson (2012). College workers and non-college workers are allowed to have different paths of life-cycle human capital growth (except for the flat spot areas where there is no additional experience accumulation for both groups), different inter-cohort human capital growth, and different time series of human capital price changes. The results are presented in Figure 11.

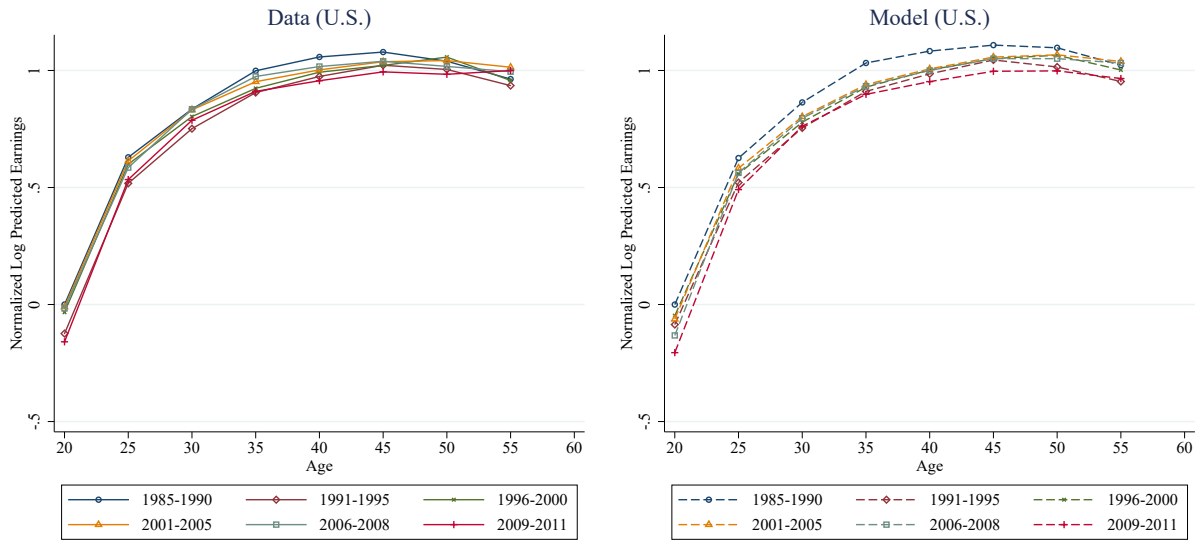
First, within an education group, the returns to experience are still higher in the U.S. than in China. Experience effects are larger for low-skilled workers than skilled workers in both countries. This is consistent with findings documented by a long literature that earnings growth tends to be lower for workers with more years of schooling (see Weiss (1986) for example). But the life-cycle paths of total quantity of human capital supplied in a year may confound the life-cycle paths of human capital supplied in a unit time with the life-cycle paths of time spent on working (because annual earnings confound hourly wages with annual hours worked).

Second, China and U.S. exhibit opposite patterns of cross-education comparisons in cohort effects. Specifically, the inter-cohort human capital growth is higher for college

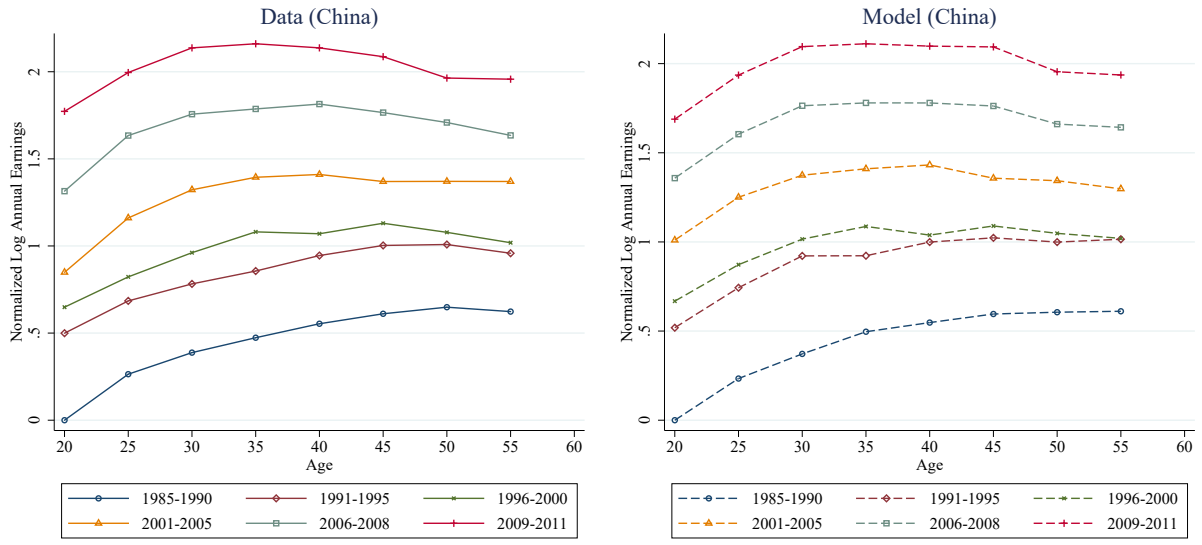
workers than non-college workers in the U.S. On the contrary, the cohort-to-cohort improvement in human capital of college workers is only a bit more than half that of non-college workers in China. It may be because although the higher educational system is getting improved fast, the improvements in basic education in China are even faster, which is reflected in the substantial increase of the initial human capital for non-college workers. It may also be a result of ability selection due to higher education expansion. As a larger fraction of later cohorts can enroll in college, which implies college is not as selective as before, the average innate ability of college students may be lower. Therefore, the initial human capital of an average college student after graduation may increase relatively more slowly from cohort to cohort.

Finally, the year effects are similar for both groups in the U.S., decreasing by around 1% per year. In China, the human capital price increases to 5 times its original level in 1985 for college workers, but less than 3 times for non-college workers.

Figure 7: Goodness of Fit



Source: Calculated from the authors' estimation



Source: Calculated from the authors' estimation

Figure 9: Growth Accounting

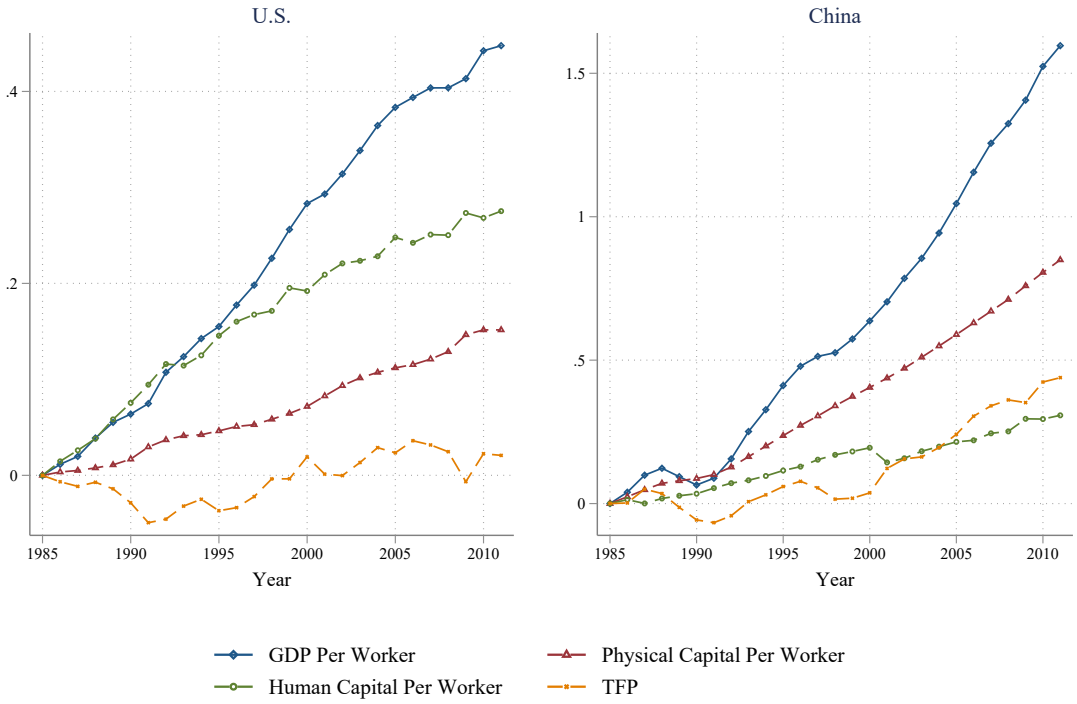


Figure 10: Decompose Human Capital Price Changes

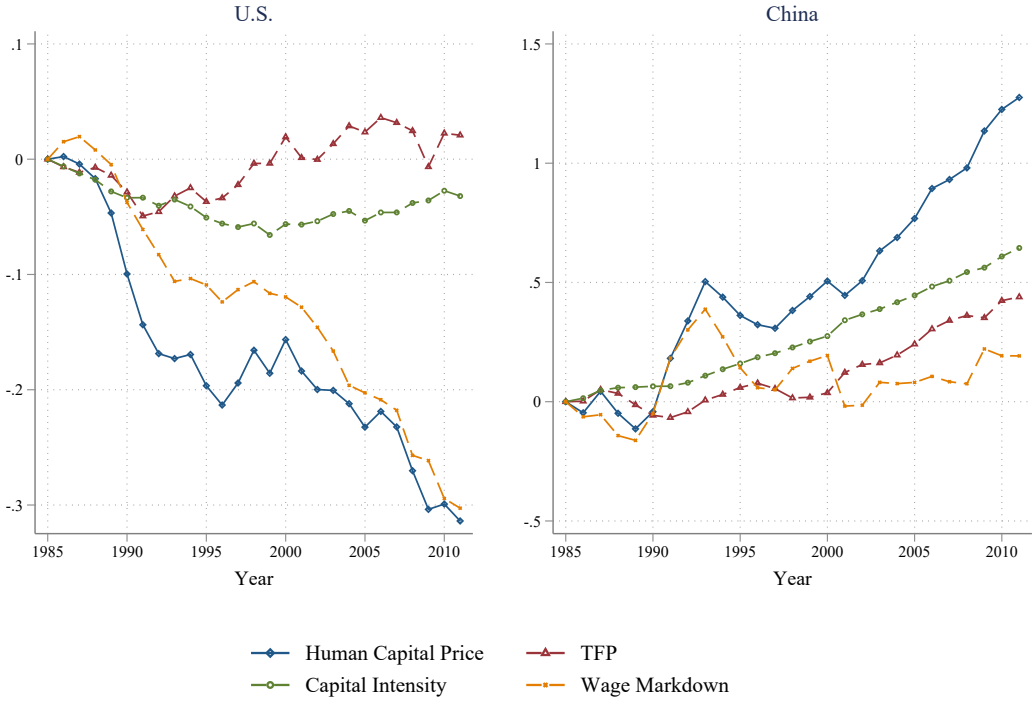
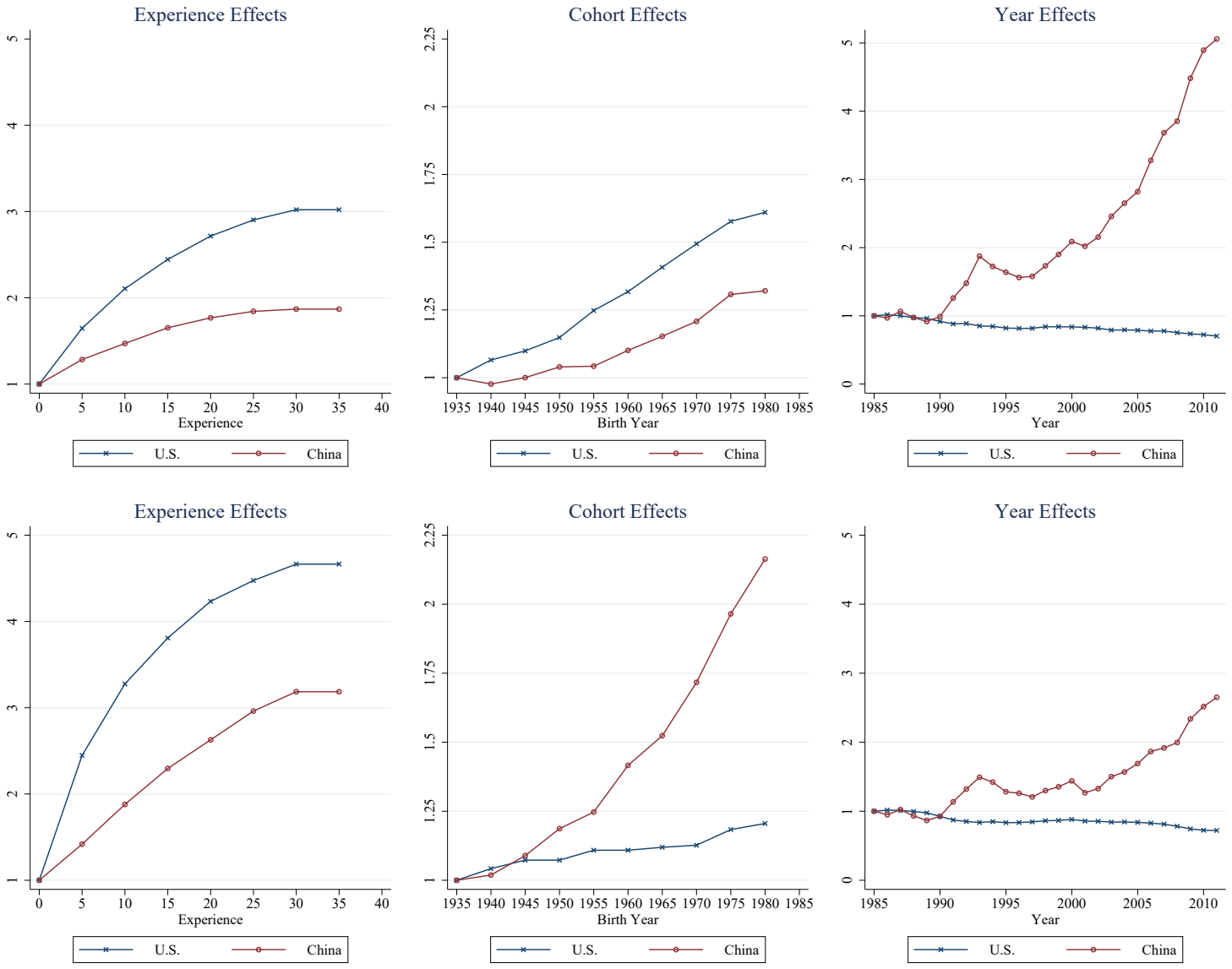


Figure 11: Decomposition for College (Upper Panel) and Non-College (Lower Panel) Male Workers



4.1.1 College Premium

The college premium literature typically interprets wage ratio of the college graduates and high school graduates as the relative price between high skill and low skill. Under this interpretation, the fact that an increase in the college wage premium is coming together with a remarkable increase in the supply of college workers in the U.S. becomes a puzzle and hence motivates the literature on the skill-biased technical change. The implicit assumption, in the above argument, is that the relative amount of human capital between education groups is constant. Consider, for example, the college premium (in annual earnings) for young workers within 5 years of experience. The average earnings of young workers in each education group $e \in \{\text{cl}, \text{hs}\}$ at time t is $w_t^e = p_t^e h_t^e$, where p_t^e is the rental price to human capital of type e at time t, and h_t^e is the average human capital for these young workers. Under the assumption of constant relative amount of human capital that

$$\frac{h_t^{\text{cl}}}{h_t^{\text{hs}}} \equiv \xi, \forall t,$$

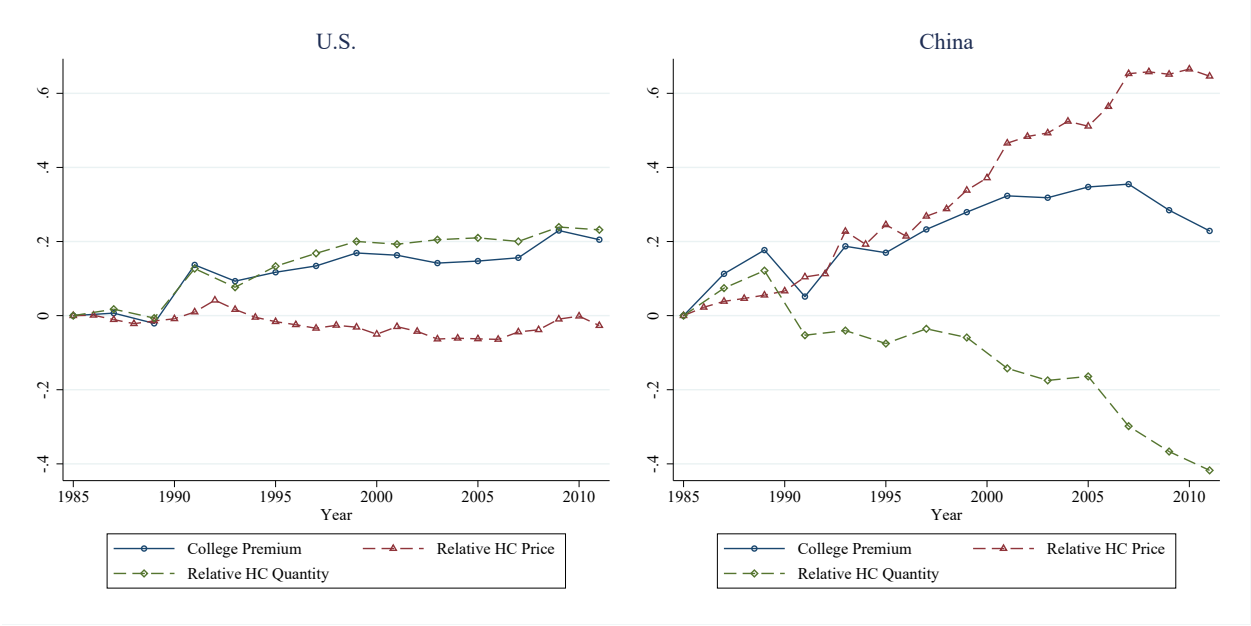
we have $w_t^{\text{cl}} / w_t^{\text{hs}} = \left(p_t^{\text{cl}} h_t^{\text{cl}}\right) / \left(p_t^{\text{hs}} h_t^{\text{hs}}\right) = \xi \left(p_t^{\text{cl}} / p_t^{\text{hs}}\right)$. Then changes in college premium over time reflect changes in the price ratio of college human capital and high school human capital, as

$$\frac{w_{t_1}^{\text{cl}} / w_{t_1}^{\text{hs}}}{w_{t_2}^{\text{cl}} / w_{t_2}^{\text{hs}}} = \frac{p_{t_1}^{\text{cl}} / p_{t_1}^{\text{hs}}}{p_{t_2}^{\text{cl}} / p_{t_2}^{\text{hs}}}.$$

If the relative amount of human capital changes over time, this equation no longer holds. Figure 12 reveals the difference between the path of college premium (for male workers within 5 years of experience) and changes in relative human capital prices. The difference lies in relative human capital quantities changes.

For the U.S., the relative price between college human capital and non-college human capital is virtually constant during the past 25 years. This means the rising college premium in the U.S. is mainly a result of the increase in the relative quantity between college human capital and non-college human capital. In China, the relative price increases even more than the increase in college premium, but the relative quantity of human capital is declining. Although human capital quantity of both college and non-college workers in China are growing, the non-college human capital quantity is growing even faster. The increasing relative human capital price dominates the decreasing relative human capital quantity, and hence we still see an increasing college premium in China.

Figure 12: Changes in College Wage Premium and Relative Human Capital Prices



4.1.2 Skill-Biased Technical Change

Although we are carefully distinguishing between the relative wage of college versus non-college worker and the relative price to college versus non-college human capital, the result turns out to still call for the role of skill-biased technical change. Without introducing skill-biased technical change, an increasing relative supply of college human capital with an almost constant relative price of college human capital in the U.S. is still at odds with a simple supply side story. In this section, we will re-estimate the magnitude of skill-biased technical change, taking into account the potential changes in the relative quantity of college versus non-college human capital.

Consider an aggregate production function that exhibits constant elasticity of substitution (CES) over skilled and unskilled human capital (college and non-college human capital):

$$Y(t) = \left[\left(A_s(t) H_s(t) \right)^{\frac{\sigma-1}{\sigma}} + \left(A_u(t) H_u(t) \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

where H_s (H_u) is the aggregate efficiency units of skilled (unskilled) labor, A_s (A_u) is the technology that augments skilled (unskilled) human capital, and $\sigma > 0$ is the elasticity of substitution between skilled and unskilled human capital.

Note that human capital price per efficiency unit under this production function is

$$p_s = \frac{\partial Y}{\partial H_s} = \left[\left(A_u H_u \right)^{\frac{\sigma-1}{\sigma}} + \left(A_s H_s \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}-1} \left(A_s H_s \right)^{\frac{\sigma-1}{\sigma}-1} A_s,$$

$$p_u = \frac{\partial Y}{\partial H_u} = \left[(A_u H_u)^{\frac{\sigma-1}{\sigma}} + (A_s H_s)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}-1} (A_u H_u)^{\frac{\sigma-1}{\sigma}-1} A_u,$$

with the time index t dropped for notational convenience but readers should bear in mind that all variables are allowed to change over time. As a result, the relative human capital price estimated above reflects $\frac{p_s}{p_u} = \left(\frac{A_s}{A_u}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{H_s}{H_u}\right)^{\frac{\sigma-1}{\sigma}-1}$, or in log terms,

$$\ln\left(\frac{p_s}{p_u}\right) = \frac{\sigma-1}{\sigma} \ln\left(\frac{A_s}{A_u}\right) - \frac{1}{\sigma} \ln\left(\frac{H_s}{H_u}\right).$$

The above expression shows that the evolution of relative human capital prices confound the skilled-bias technical change with the changes in relative amount of human capital efficiency units.

We calibrate $\sigma = 1.4$, which is the benchmark value estimated by Katz and Murphy (1992), and obtain an estimated series for changes in $\ln\left(\frac{A_s}{A_u}\right)$. Since $\sigma > 1$, an increase in A_s/A_u (i.e. skill-biased technical change) will increase p_s/p_u, while an increase in H_s/H_u (i.e. an increasing relative supply of skilled human capital) will decrease p_s/p_u. Denote h_s (h_u) the efficiency units per worker of skilled (unskilled) workers such that the aggregate supply is $H_u = h_u L_u$ ($H_s = h_s L_s$). Our decomposition gives the relative price p_s/p_u and the relative human capital quantities per worker h_s/h_u. Since the relative labor supply L_s/L_u is observed, we can obtain the relative supply of aggregate human capital H_s/H_u. Hence the skill-biased technical changes can be obtained as a residual. Since the production function here concerns the whole economy, we no longer restrict attention to males or young workers within 5 years of experience but extend analysis to all workers. An assumption that the male human capital price changes are the same as female human capital changes, however, is needed.

In the skill-biased technical change literature (see excellent reviews by Acemoglu (2002) and Violante (2008)), it is often assumed that

$$Y(t) = \left[(B_s(t) L_s(t))^{\frac{\sigma-1}{\sigma}} + (B_u(t) L_u(t))^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{4}$$

where L_s (L_u) is the labor supply of skilled (unskilled) workers, and the evolution in B_s/B_u is interpreted as the skill-biased technical change. Our formulation is consistent with it, and in fact generalizes it. Our formulation (3) distinguishes improvements in h_s/h_u with the technology that improves the productivity of skilled or unskilled human capital (i.e. A_s/A_u). These two forces together form the standard interpretation of skill-

biased technical change B_s/B_u. To see this, rewrite the production function as

$$Y(t) = \left[(\underbrace{A_s(t) h_s(t)}_{B_s(t)} L_s(t))^{\frac{\sigma-1}{\sigma}} + (\underbrace{A_u(t) h_u(t)}_{B_u(t)} L_u(t))^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}.$$

In a competitive labor market, we have

$$w_s = \frac{\partial Y}{\partial L_s} = \left[(A_u h_u L_u)^{\frac{\sigma-1}{\sigma}} + (A_s h_s L_s)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma-1}{\sigma}-1} (A_s h_s)^{\frac{\sigma-1}{\sigma}} L_s^{\frac{\sigma-1}{\sigma}-1},$$

$$w_u = \frac{\partial Y}{\partial L_u} = \left[(A_u h_u L_u)^{\frac{\sigma-1}{\sigma}} + (A_s h_s L_s)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma-1}{\sigma}-1} (A_u h_u)^{\frac{\sigma-1}{\sigma}} L_u^{\frac{\sigma-1}{\sigma}-1}.$$

Therefore the college earnings premium can be written as $\frac{w_s}{w_u} = \left(\frac{A_s h_s}{A_u h_u}\right)^{\frac{\sigma-1}{\sigma}} \left(\frac{L_s}{L_u}\right)^{\frac{\sigma-1}{\sigma}-1}$, or in log terms,

$$\ln\left(\frac{w_s}{w_u}\right) = \frac{\sigma-1}{\sigma}\ln\left(\frac{A_s}{A_u}\right) + \frac{\sigma-1}{\sigma}\ln\left(\frac{h_s}{h_u}\right) - \frac{1}{\sigma}\ln\left(\frac{L_s}{L_u}\right).$$

Note that with the typical formulation (4) such as in Katz and Murphy (1992), we will have

$$\ln\left(\frac{w_s}{w_u}\right) = \frac{\sigma-1}{\sigma}\ln\left(\frac{B_s}{B_u}\right) - \frac{1}{\sigma}\ln\left(\frac{L_s}{L_u}\right).$$

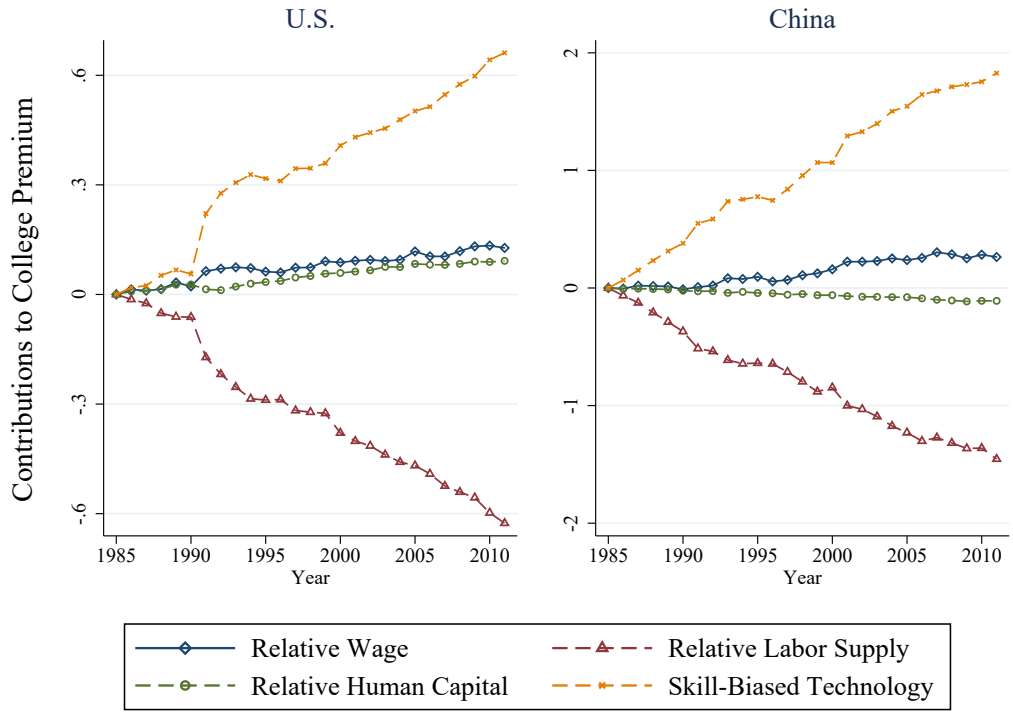
In other words, B_s/B_u in their model is equivalent to $(A_s h_s)\,/\,(A_u h_u)$ in our formulation, and it is really a combination of the skill-biased technical change and the changes in relative human capital per worker between skilled and unskilled workers. Katz and Murphy (1992) estimate this equation, using 40 years of U.S. data and their benchmark estimate is $\sigma = 1.4$. In log changes, we have

$$\Delta\ln\left(\frac{w_s}{w_u}\right) = \frac{\sigma-1}{\sigma}\Delta\ln\left(\frac{A_s}{A_u}\right) + \frac{\sigma-1}{\sigma}\Delta\ln\left(\frac{h_s}{h_u}\right) - \frac{1}{\sigma}\Delta\ln\left(\frac{L_s}{L_u}\right).$$

As we can see, there are three factors that affect the college premium. An increase in relative labor supply L_s/L_u, holding everything else fixed, decreases relative wage. The relative average human capital efficiency units h_s/h_u have two effects. First, it decreases the relative human capital prices p_s/p_u. Second, it also increases skilled-labor's relative earnings capacity. The overall effect is positive if $\sigma > 1$ when the second effect dominates the first. The effect of the skill-biased technical changes (increasing A_s/A_u) on college premium depends on σ, too. The contributions of relative labor supply, relative human capital per worker, and skill-biased technical change to the evolution of overall college

wage premium is shown in Figure 13. Both countries has witnessed an increasing fraction of college workers and continuing skill-biased technical changes during the same period.

Figure 13: Contributions to (Log) College Wage Premium ($\sigma = 1.4$)



4.2 Heterogeneous Experience Effects Across Cohorts

In the previous decomposition exercise, the life-cycle human capital accumulation path is restricted to be the same across cohorts. Suppose the path of returns to experience varies across cohorts. That is, suppose the true model is

$$\ln w_{i,c,t} = \text{cons} + \ln p_t + \ln s_c + \ln r_k^c + \varepsilon_{i,c,t},$$

where $\mathbb{E}_i\left(\varepsilon_{i,c,t}|c,t\right) = 0, \forall c,t$ and the experience effect r_k^c is allowed to depend on c.

In theory, this model can still be identified by the same identifying assumption that $r_{k_1}^c = r_{k_2}^c$ for some preset k_1, k_2 for each cohort c. In practice, however, it is infeasible because for many cohorts, we will not be able to observe in the data the actual earnings when they have k_1, k_2 years of experience. To illustrate, suppose k_1 and k_2 correspond to 50 and 55 years old. If we have repeated cross-sectional data covering continuously from 1980 to 2010, we can only observe the 50-year-old's earnings for cohorts born between

1930 and 1960, and we can only observe the 55-year-old's earnings for cohorts born between 1925 and 1955. Therefore, we can only apply the identifying assumption at most to the cohort born between 1930 and 1955. What is even worse, we can only observe the life cycle after 50 years old for the 1930 cohort. Only for the cohort with the best coverage (i.e. cohort 1955), we can only observe the whole meaningful working career from 25 years old to 55 years old. This reveals the benefit of restricting the returns to experience to be constant across cohorts as done in the previous decomposition – we can obtain an estimate for a complete life-cycle human capital accumulation path even if we typically cannot observe the whole life cycle in the data for most cohorts.

4.2.1 Econometric Issue

A natural algorithm for estimating the heterogeneous experience effects model is to first estimate the constant returns to experience model

$$\ln w_{i,c,t} = \text{cons} + \ln p_t + \ln s_c + \ln r_k + \tilde{\varepsilon}_{i,c,t},$$

and then back out the heterogeneous paths of returns to experience by a simple plug-in estimator

$$\widehat{\ln r_k^c} = \overline{\ln w_{c,t}} - \widehat{\text{cons}} - \widehat{\ln p_t} - \widehat{\ln s_c},$$

where $\overline{\ln w_{c,t}}$ is the average earning across workers in cohort c at time t, and $\widehat{\text{cons}}$, $\widehat{\ln p_t}$, $\widehat{\ln s_c}$ are the estimates from the specification of homogeneous returns to experience. However, we show that this algorithm does not generate consistent estimates. To see this, notice that $\tilde{\varepsilon}_{i,c,t} = \ln r_{t-c}^c - \ln r_{t-c} + \varepsilon_{i,c,t}$. We have the following lemma.

Lemma 1. $\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t=t'\}}\right] = 0, \forall t'$; $\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}}\right] \neq 0, \forall c'$; $\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t-c=k'\}}\right] = 0, \forall k'$.

Proof. (1) Fix any t'. Using the law of iterated expectation, we have

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t=t'\}}\right] = \mathbb{E}_{c,t}\left[\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t=t'\}} | c, t\right]\right].$$

Observe that

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t=t'\}} | c, t \neq t'\right] = 0,$$

and

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t=t'\}} | c, t = t'\right] = \mathbb{E}_i\left[\ln r_{t-c}^c - \ln r_{t-c} + \varepsilon_{i,c,t} | c, t = t'\right] = \ln r_{t'-c}^c - \ln r_{t'-c}.$$

Note that $\mathbb{E}_c\left\{\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t=t'\}} | c, t = t'\right]\right\} = \mathbb{E}_c\left[\ln r_{t'-c}^c - \ln r_{t'-c}\right] = 0$. Thus $\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t=t'\}}\right] =$

$0, \forall t'$.

(2) Fix any c'. Using the law of iterated expectation, we have

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}}\right] = \mathbb{E}_{c,t}\left[\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}} | c, t\right]\right].$$

Observe that

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}} | c \neq c', t\right] = 0,$$

and

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}} | c = c', t\right] = \mathbb{E}_i\left[\ln r^c_{t-c} - \ln r_{t-c} + \varepsilon_{i,c,t} | c = c', k\right] = \ln r^{c'}_{t-c'} - \ln r_{t-c'}.$$

Note that $\mathbb{E}_t\left\{\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}} | c = c', t\right]\right\} = \mathbb{E}_t\left[\ln r^{c'}_{t-c'} - \ln r_{t-c'}\right]$ does not equal o in general. Therefore, $\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}}\right] \neq 0, \forall c'$ in general.

(3) Fix any k'. Using the law of iterated expectation, we have

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t-c=k'\}}\right] = \mathbb{E}_{c,t}\left[\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t-c=k'\}} | c, t\right]\right].$$

Observe that

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t-c=k'\}} | t - c \neq k'\right] = 0,$$

and

$$\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t-c=k'\}} | t - c = k'\right] = \mathbb{E}_i\left[\ln r^c_{t-c} - \ln r_{t-c} + \varepsilon_{i,c,t} | t - c = k'\right] = \ln r^c_{k'} - \ln r_{k'}.$$

Note that $\mathbb{E}_c\left\{\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t-c=k'\}} | t - c = k'\right]\right\} = \mathbb{E}_t\left[\ln r^c_{k'} - \ln r_{k'}\right] = 0$. Thus $\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{t-c=k'\}}\right] = 0, \forall k'$. □

Lemma 1 tells us that the aforementioned naive plug-in estimator is not consistent, because $\mathbb{E}_i\left[\tilde{\varepsilon}_{i,c,t} \cdot \mathbb{I}_{\{c=c'\}}\right] \neq 0, \forall c'$. In general, at the presence of even one endogenous variable in a regression, all of the other coefficients will be inconsistent. But a special case is that when the endogenous variables are uncorrelated with other right-hand side variables, then the endogeneity will not contaminate the other coefficients. We make precise this result in Lemma 2.

Lemma 2. *Consider a regression $y_i = \beta w_i + \gamma z_i + \varepsilon_i$, where w_i and z_i are vectors of independent variables. Suppose $\mathbb{E}(z_i \varepsilon_i) = 0$, but $\mathbb{E}(w_i \varepsilon_i) \neq 0$. In general, both the OLS estimator $\hat{\beta}$ and $\hat{\gamma}$ are inconsistent. If in addition, w_i and z_i are uncorrelated in the sense that $W'Z = 0$, the OLS estimate $\hat{\gamma}$ is consistent.*

Proof. The OLS estimator is $(X'X)^{-1} X'Y$, where $X \coloneqq \begin{pmatrix} W & Z \end{pmatrix}$. In the special case that $W'Z = 0$, we have

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} W'W & W'Z \\ Z'W & Z'Z \end{pmatrix}^{-1} \begin{pmatrix} W'Y \\ Z'Y \end{pmatrix} = \begin{pmatrix} W'W & 0 \\ 0 & Z'Z \end{pmatrix}^{-1} \begin{pmatrix} W'Y \\ Z'Y \end{pmatrix} = \begin{pmatrix} (W'W)^{-1} W'Y \\ (Z'Z)^{-1} Z'Y \end{pmatrix}.$$

As a result,

$$\hat{\gamma} = \left(Z'Z\right)^{-1} Z' \left(\beta W + \gamma Z + \varepsilon\right) = \gamma + \left(Z'Z\right)^{-1} Z'\varepsilon \to 0.$$

□

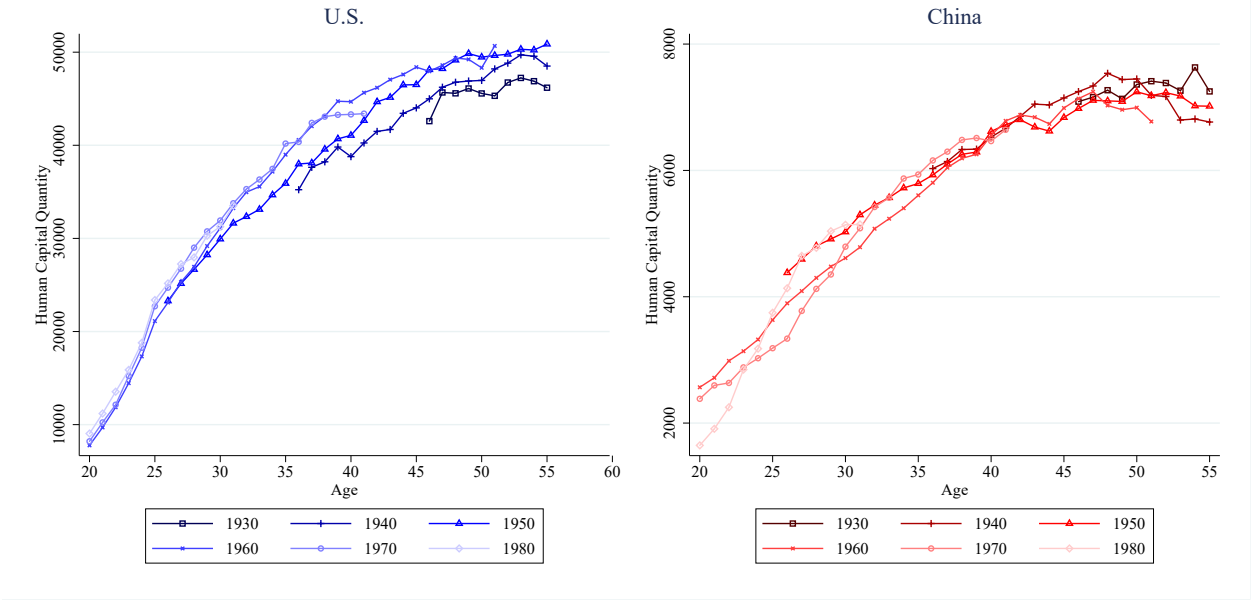
Conceptually, this condition should be satisfied, as cohort is completely a different concept than time and experience. For example, how many years of experience a worker has should not be systematically correlated with the birth year; after all, everyone has to be young and old. Presumably, either worker of cohort 1950, 1960, 1970, 1980, 1990 can be observed at year 2010. If the assumption that cohort dummies are uncorrelated with year dummies and experience dummies holds, we still get consistent estimates for time effects and average returns to experience. Using the correct estimates for price effects, we can back out the heterogeneous paths of life-cycle human capital accumulation for each cohort by

$$\widehat{\ln h_{c,t}} = \overline{\ln w_{c,t}} - \widehat{\ln p_t},$$

up to a normalization. The results are reported in Figure 14. But in practice, due to the finiteness of the life period, c and t are typically correlated in the data. For example, older cohorts tend to live in earlier periods, but younger cohorts tend to live in recent time. This may violate the assumption that cohort dummies are uncorrelated with year dummies.

Figure 14 reports the life-cycle human capital accumulation by cohorts estimated using the naive algorithm, where darker lines are for older cohorts and lighter lines for younger cohorts. The life-cycle human capital accumulation paths for different cohorts highly overlap with each other, suggesting that the homogeneous returns to experience assumption would be a reasonable one. If anything, the profiles are shifting outwards, meaning that returns to experience are getting a little bit higher for later cohorts.

Figure 14: Life-Cycle Human Capital Accumulation by Cohorts



A compromise is to focus on cohorts who are observed for the whole flat spot in the dataset. As discussed before, the benefit is that we can now allow for heterogeneous returns to experience across cohorts, and impose the assumption that there is no human capital accumulation in the flat spot for each cohort. The limitation is that we will lose too many samples and can only focus on too narrow a range of cohorts as discussed earlier in this section, which is no so interesting.

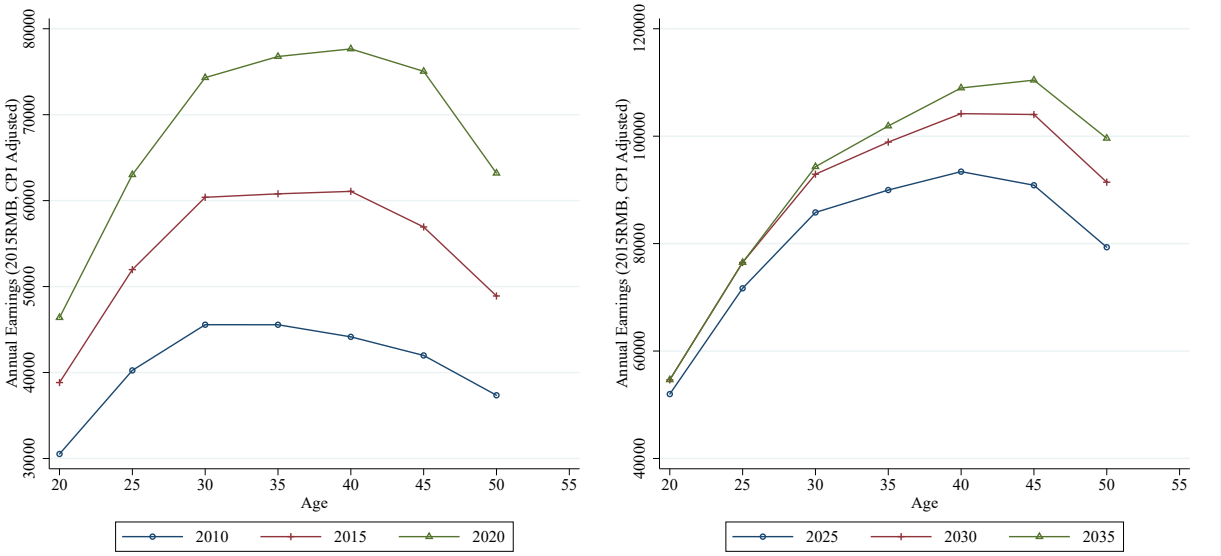
4.3 Counterfactual

The fast growth in China is expected to slow down in future. Between 1985 and 2011, the average inter-cohort human capital growth rate in China is 1.67% per year, and the average growth rate of human capital prices in China is 4.82% per year. This section performs an experiment that both the cohort effects and time effects uniformly decelerate from the average growth rate (around 1.5% and 5%, respectively) to zero growth (approximately the U.S. case) in 30 years, with the experience effects fixed at China's current level.

In this scenario, the vertical gaps between two cross-sectional earnings profiles are shrinking, showing the slowdown in the time effects. Notably, the "golden age" is around 30 in 2010, but the profile is getting flat between 30 to 40 years old in 2015, and the "golden age" would become 45 in 2035. Recall Proposition 1 and its corollary that the position of the "golden age" is essentially a race between experience effects and cohort effects. The "golden age" becoming older and older is a result of the slowdown in

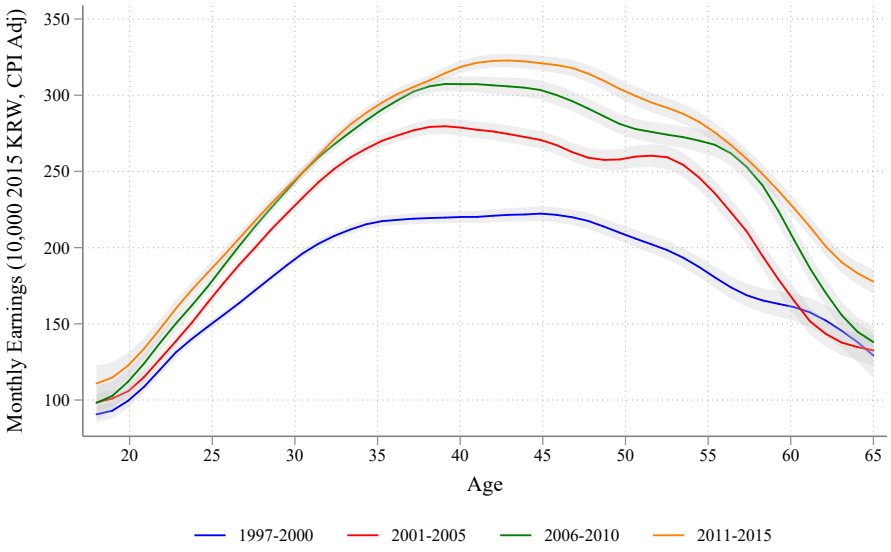
cohort effects. This scenario for the future 30 years reveals an opposite pattern to what happened in China during the past 30 years, but a similar pattern to what happened in Korea during the past 10 years. Korea experienced its fastest growth during 1960s to 1990s. After that, it begins to slowdown. Figure 17 depicts the decomposition for Korea, together with the previous decomposition for U.S. and China.

Figure 15: A Counterfactual Experiment for China's Earnings Profiles in 30 years



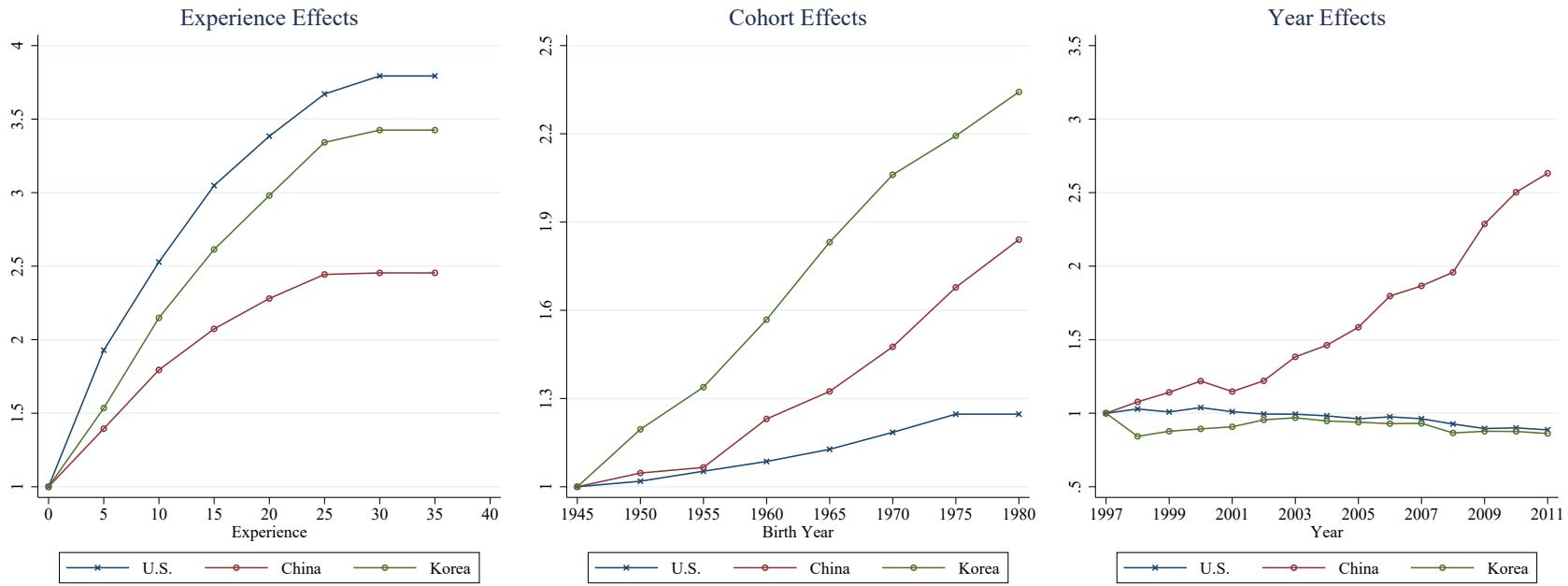
Source: Calculated from the authors' estimation

Figure 16: Cross-Sectional Age-(Monthly)Earnings Profiles of Korean Male Workers



Source: Calculated from the Korean Labor & Income Panel Study (KLIPS)

Figure 17: Decomposition



5 Conclusion

This paper documents three stark differences in the age-earnings profiles between the U.S. and China. First, the cross-sectional "golden age" stays stable at around 45-50 years old in the U.S. but continuously decreases from 50 to 35 years old in China during the past 30 years. Second, age-specific labor earnings grow dramatically in China but do not witness much growth in the U.S. Third, the cross-sectional and life-cycle age-earnings profiles look remarkably similar in the U.S. but differ substantially in China.

As the first step to explain these puzzles, we exploit a decomposition framework to address these facts. The decomposition suggests that China has experienced a much larger inter-cohort improvement in human capital and increase in the rental price to human capital, compared to the U.S. But the return to experience is higher in the U.S. We apply the decomposition result to adjust for the changes in human capital and obtain a series of estimated TFP growth that is lower than previous estimates.

In the baseline analysis, we assume workers are perfect substitutes. We discuss the possibility that different education groups are not perfect substitutes, but different cohorts are still restricted to be perfect substitutes. There might be some technical changes that makes older cohorts obsolete, and therefore, cohorts are not perfect substitutes in the production function. In that case, cohort sizes will matter. But it is hard to distinguish technical changes that favors younger generations and inter-cohort human capital growth without better-suited data or richer structures. We leave this question for future research.

We believe another fruitful direction is to link the decomposition results to specific institutions and evaluate how much does each policy contribute the three effects. We list some of our hypotheses here. The college expansion is the most related to the inter-cohort improvement in human capital. Joining the WTO and SOE reforms may well increase the overall efficiency of the economy and hence also the rental price to human capital. Throughout the current paper, we focused on urban males. A natural question is how would structural change from agriculture to industry, increasing female labor participation, and internal migration (especially in China) be reflected in the decomposition.

References

Acemoglu, Daron. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature*, 40(1): 7–72.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review*, 108(6): 1488–1542.

Aguiar, Mark, and Erik Hurst. 2013. "Deconstructing Life Cycle Expenditure." *Journal of Political Economy*, 121(3): 437–492.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75(4, Part 1): 352–365.

Berger, David W, Kyle F Herkenhoff, and Simon Mongey. 2019. "Labor Market Power." National Bureau of Economic Research.

Bils, Mark, and Peter J Klenow. 2000. "Does Schooling Cause Growth?" *American Economic Review*, 90(5): 1160–1183.

Bowlus, Audra J, and Chris Robinson. 2012. "Human Capital Prices, Productivity, and Growth." *American Economic Review*, 102(7): 3483–3515.

Brandt, Loren, Johannes Van Biesebroeck, Luhang Wang, and Yifan Zhang. 2017. "WTO Accession and Performance of Chinese Manufacturing Firms." *American Economic Review*, 107(9): 2784–2820.

Browning, Martin, Lars Peter Hansen, and James J Heckman. 1999. "Micro Data and General Equilibrium Models." *Handbook of Macroeconomics*, 1: 543–633.

Burdett, Kenneth, and Dale T Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review*, 257–273.

Cai, Yong, Feng Wang, Ding Li, Xiwei Wu, and Ke Shen. 2014. "China's Age of Abundance: When Might It Run Out?" *The Journal of the Economics of Ageing*, 4: 90–97.

Casanova, Maria. 2013. "Revisiting the Hump-Shaped Wage Profile." *Unpublished mimeo, UCLA*.

Chen, Yuyu, Mitsuru Igami, Masayuki Sawada, and Mo Xiao. 2019. "Privatization and Productivity in China." *Available at SSRN 2695933*.

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* The World Bank.

Deaton, Angus, and Christina Paxson. 1994. "Intertemporal Choice and Inequality." *Journal of Political Economy*, 102(3): 437–467.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *The Quarterly Journal of Economics*.

Elsby, Michael WL, Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity*, 2013(2): 1–63.

Fernald, John G. 2015. "Productivity and Potential Output before, during, and after the Great Recession." *NBER Macroeconomics Annual*, 29(1): 1–51.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, and J. Robert Warren. 2018. "Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]." *Minneapolis, MN: IPUMS*.

Gourinchas, Pierre-Olivier, and Jonathan A Parker. 2002. "Consumption over the Life Cycle." *Econometrica*, 70(1): 47–89.

Hall, Robert E. 1968. "Technical Change and Capital from the Point of View of the Dual." *The Review of Economic Studies*, 35(1): 35–46.

Hall, Robert E, and Charles I Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 83–116.

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

Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2011. "Sources of Lifetime Inequality." *American Economic Review*, 101(7): 2923–54.

Islam, Asif, Remi Jedwab, Paul Romer, and Daniel Pereira. 2018. "The Sectoral and Spatial Allocation of Labor and Aggregate Returns to Experience."

Kambourov, Gueorgui, and Iourii Manovskii. 2009. "Accounting for the Changing Life-Cycle Profile of Earnings."

Katz, L. F., and K. M. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, 107(1): 35–78.

Kleiner, Morris M., and Alan B. Krueger. 2013. "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." *Journal of Labor Economics*, 31(S1): S173–S202.

Kuruscu, Burhanettin. 2006. "Training and Lifetime Income." *American Economic Review*, 96(3): 832–846.

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman. 2018. "Life Cycle Wage Growth across Countries." *Journal of Political Economy*, 126(2): 797–849.

Lipsius, Ben. 2018. "Labor Market Concentration Does Not Explain the Falling Labor Share." *Available at SSRN 3279007*.

Lise, Jeremy, and Fabien Postel-Vinay. 2019. "Multidimensional Skills, Sorting, and Human Capital Accumulation."

Macaluso, Claudia, Brad Hershbein, Chen Yeh, et al. 2019. "Concentration in US Local Labor Markets: Evidence from Vacancy and Employment Data." Society for Economic Dynamics.

Manuelli, Rodolfo E., and Ananth Seshadri. 2014. "Human Capital and the Wealth of Nations." *American Economic Review*, 104(9): 2736–2762.

McKenzie, David J. 2006. "Disentangling Age, Cohort and Time Effects in the Additive Model." *Oxford Bulletin of Economics and Statistics*, 68(4): 473–495.

Mincer, Jacob A. 1974. "Schooling, Experience, and Earnings."

Rupert, Peter, and Giulio Zanella. 2015. "Revisiting Wage, Earnings, and Hours Profiles." *Journal of Monetary Economics*, 72: 114–130.

Schulhofer-Wohl, Sam. 2018. "The Age-Time-Cohort Problem and the Identification of Structural Parameters in Life-Cycle Models." *Quantitative Economics*, 9(2): 643–658.

Violante, Giovanni L. 2008. "Skill-Biased Technical Change." *The New Palgrave Dictionary of Economics*, 2.

Weiss, Yoram. 1986. "The Determination of Life Cycle Earnings: A Survey." *Handbook of Labor Economics*, 1: 603–640.

Zhu, Xiaodong. 2012. "Understanding China's Growth: Past, Present, and Future." *Journal of Economic Perspectives*, 26(4): 103–24.