Refugees and Early Childhood Human Capital*

Todd Schoellman[†]

March 2013

Abstract

This paper quantifies cross-country differences in early childhood human capital. I first provide new empirical evidence that draws on the experiences of Indochinese refugees who arrived to the U.S. during early childhood. The key moment is the relationship between age at arrival to the U.S. and wages in adulthood. I argue that age at arrival is exogenous for Indochinese refugees and hence that this moment captures only the late-life impact of having spent a year of early childhood in a poor country instead of the U.S. The estimated effect is zero: refugees who arrived at age zero or five have similar outcomes. I use a standard human capital model to help interpret this finding. I show that there are three assumptions that rationalize this finding, only one of which is consistent with the broader literature on early childhood human capital: that parents, and not goods or country environment, are the critical input during early childhood. I conclude that cross-country differences in early childhood human capital are small.

*Thanks to Rody Manuelli, Matt Wiswall and audiences at Ohio State University, UC – Davis, the 2012 Midwest Macroeconomics Meetings, and the 2013 NBER Growth meetings for helpful comments. The usual disclaimer applies.

[†]Address: Department of Economics, Arizona State University, Tempe, AZ 85287. E-mail: todd.schoellman@gmail.com

1 Introduction

One of the central challenges for economists is to explain the large differences in gross domestic product (GDP) per worker across countries. The development accounting literature provides a partial answer: the quantitative role for physical and human capital is limited, implying a large role for total factor productivity (TFP); see Caselli (2005) for a review of this literature. Much of the early work in this literature equates human capital with years of schooling. In this paper I consider whether allowing for a broader notion of human capital would substantially alter this conclusion. In particular, my goal is to quantify the role for early childhood human capital, defined as human capital that is accumulated before school starts (roughly, by age 6). Along the way, I simultaneously establish evidence on the effectiveness of different types of early childhood investments.

The motivation for this paper comes from two facts. First, a large literature has documented the lasting effectiveness of investments and interventions made during early childhood; see Almond and Currie (2011) for an overview. Second, there is substantial cross-country variation in early childhood environment and investments, including disease prevalence, nutrition, or medical spending. The natural question is whether these two facts imply large cross-country differences in human capital even before school starts. Unfortunately, it is not clear how to use the existing evidence to address this question. For example, Cunha et al. (2010) and Del Boca et al. (2012) structurally estimate the human capital production function using U.S. data and find an important role for early childhood human capital investments. It is not obvious how to apply their production function outside the U.S. in the absence of comparable input data from other countries. Likewise, an existing literature documents the effectiveness of specific early childhood interventions (Blau and Currie, 2006; Cunha et al., 2006). Again, it is not clear how to translate the importance of intensive preschool programs into statements about cross-country human capital stocks.

To overcome this challenge I provide new evidence on the late-life outcomes of refugees who immigrated to the U.S. during early childhood. Specifically, I focus on the relationship between age at arrival and socioeconomic outcomes in adulthood. Refugees who arrived at older ages spent more of their early childhood in their country of birth or in refugee camps and less of their early childhood in the U.S. I exploit this variation to estimate the marginal cost of spending a year of early childhood in a poor instead of a rich country. An additional benefit of examining the data in this way is that it is less likely to be biased by selection than the obvious alternative of comparing refugees to American-born workers. My results will be biased only if refugees who arrived to the U.S. at different ages of early childhood were selected differently.

I focus my analysis on the Indochinese refugees, that is, refugees from Vietnam, Cambodia, and Laos in the late 1970s and 1980s. The key advantage of Indochinese refugees is that they could only control when they fled their homes; the transit period and wait times in camps before resettlement were variable and could be quite long. For this reason it is extremely unlikely that refugee families were able to select on the age of their young children, even if there were an underlying motive to do so. An additional benefit of focusing on the Indochinese refugees is that they came from demonstrably poor and disadvantaged backgrounds. GDP per worker in their birth countries was roughly 3 percent of U.S. levels in 1980 (Heston et al., 2012). Refugees also faced an environment that one might expect to be disastrous for human capital accumulation, having survived the Vietnam War, the Khmer Rouge, refugee camps, and other hardship. In the face of these many disadvantages, I find a striking result: there is no relationship between wages or other socioeconomic outcomes and age at arrival for the Indochinese refugees. This result is robust to the details of the estimation or sample selection. I show that similar results extend to other immigrants from poor countries. I also show that the result is special to early childhood: a strong relationship between age at arrival and outcomes does emerge, but only for those who arrive after early childhood.

The next step is to understand the significance of this finding. I embed a standard human capital production function in the spirit of Cunha et al. (2010) and Del Boca et al. (2012) into a model of human capital investment and labor market outcomes across countries. In these models, individuals form human capital in two distinct stages of their lives, early childhood and school. The child's human capital at the end of each stage is a function of their human capital at the beginning of the stage and any investments made during the stage. Investments are allowed to take a variety of forms, including the time of children or parents, goods, or an exogenous environmental input particular to the country they live in. After each individual graduates school they enter the labor force and earn a wage proportional to their final human capital.

I show that this model generically predicts that human capital and wages of refugees decline in age at arrival. This prediction rests on three logical steps that reveal the necessary assumptions on the human capital production function. First, the model predicts that each year spent in a poor country yields less human capital than a year spent in a rich country. There are two channels through which this occurs: poor countries offer an exogenously worse environment for human capital accumulation; and families in low-income countries endogenously purchase fewer goods for their children. The latter effect is precisely the one first emphasized in Manuelli and Seshadri (2010). The implication of this assumption is that refugees who arrive to the U.S. later will have less human capital as of age 5. The second logical step is that early childhood human capital persists as an important input to subsequent human capital accumulation. The implication of this assumption is that refugees who arrive to the U.S. later will also have less human capital as of the end of school if all other investmetns are held equal. Of course, there is no natural reason why other investments should be held equal. Families of children who are disadvantaged at age 5 may choose to invest more in them during their school years to help remediate that disadvantage. Whether they choose to do so, and the effectiveness of such investment, is controlled by the elasticity of substitution between their human capital at age 5 and subsequent investments. As long as that elasticity of substitution is sufficiently low then remediation will be imperfect, and refugees who arrive to the U.S. later will have lower human capital as adults.

Given that this model prediction is counterfactual, it must be the case that one of these three assumptions does not hold: either living in a poor country does not translate to an initial disadvantage in human capital; early childhood human capital does not persist; or early childhood human capital deficits can be remediated easily. I draw some additional evidence from the post migration-investments of refugees. I also discipline the model to be consistent with the evidence of the large existing literature on early childhood human capital. That literature provides an overwhelming body of evidence that early childhood human capital is persistent. It also includes modest although perhaps not conclusive evidence that early childhood human capital deficits are difficult to remediate. Given this evidence, I conclude that it is parents, and not market goods or the country that one lives in, that are critical inputs for human capital accumulation. On this point my paper adds to a growing consensus that positive interactions with parents are the critical input (Del Boca et al., 2012). An additional implication is that early childhood human capital is not likely to differ much across countries.

In addition to the work mentioned above, my paper is related to two other literatures. First, it joins a growing literature that re-considers the role of human capital in accounting for cross-country income differences, mostly by taking a broader view of human capital. Manuelli and Seshadri (2010), Erosa et al. (2010), and Schoellman (2012) study the importance of education quality in addition to years of schooling. Manuelli and Seshadri (2010), Lagakos et al. (2012), and Lagakos et al. (2013) study the impact of experience and on-the-job human capital accumulation. To my knowledge, Manuelli and Seshadri (2010) is the only previous paper to consider early childhood human capital; they use a calibrated model and find larger results than I do here. Second, my paper is related to a literature that studies the effect of age at arrival on immigrant outcomes. Friedberg (1992) first proposed a role for age at arrival and laid out the assumptions necessary for identification. A number of papers that have applied this framework to immigrants who arrived as children found small or no effects of age at arrival on socioeconomic outcomes for those who arrive before age 5, consistent with my findings (Myers et al., 2009; Lee and Edmonston, 2011; Gonzalez, 2003; Bleakley and Chin, 2010).

The rest of the paper proceeds as follows. Section 2 documents the empirical results about the experience of Indochinese refugees in the U.S. Section 3 formulates the model and shows how its predictions can be made consistent with the existing data. Section 4 studies the model's predictions for cross-country human capital differences. Section 5 considers alternative hypotheses for my empirical findings. Section 6 concludes.

2 Empirical Evidence: Experience of Indochinese Refugees

The key empirical idea of this paper is to study the outcomes of immigrants who arrived from poor countries to the U.S. during early childhood. In this section I give some background on the Indochinese refugees and explain why they are ideal for the exercise at hand. I then describe the data and empirical methodology and present the results.

2.1 The Indochinese Refugees

The Indochinese refugees are former citizens of Vietnam, Cambodia, and Laos who fled their home country in response to Communist takeovers in 1975. In all, roughly three million citizens escaped to nearby countries by land or by boat. 1.4 million of those who fled were placed in refugee camps where they waited for an opportunity to settle within the country permanently, repatriate to their former home, or resettle to a third country. Ultimately, 1.3 million refugees were permanently resettled to new homes in countries around the world. The U.S. was the most common destination, accepting over 800,000 Indochinese refugees.[1] This large sample of resettled refugees – in particular, those who entered as young children

[1] No other country took nearly as many refugees. I was able to construct results (available upon request) for Canada that show the same pattern as in the U.S., but with much less precision because of the small sample size. All resettlement figures are taken from United Nations High Commissioner for Refugees (2001).

– are my population of interest.

Indochinese refugees in the U.S. are composed of five major, distinct ethnic groups. The Vietnamese, Lao, and Khmer are the main ethnic groups of the respective countries, but there are also a sizeable number of ethnic Chinese from Vietnam and ethnic Hmong from Laos in the U.S. These ethnic groups differed greatly in their pre-migration characteristics such as education level, occupational background, rural-urban status, and so on; I give some basic indicators of these differences in Appendix A. Given these large between-group differences I will present my results by ethnic group to avoid compositional biases.

Refugees to the U.S. arrived in two large waves, with smaller flows in the remaining years. The first wave consisted of those with close political or military connections to the previous, non-Communist regimes or the U.S., most notably the 130,000 people flown out of Saigon in the final days before the U.S. withdrew from South Vietnam (Hung and Haines, 1996). These immigrants were highly selected on many dimensions, typically spoke English well, and had relatively short transit periods before they were resettled in the U.S. I exclude 1975 arrivals from all subsequent analysis in order to focus on immigrants who were less selected and had more typical refugee and resettlement and experiences.

Remaining Indochinese refugees form an ideal sample for three reasons. The first is the large available sample. The 800,000 Indochinese refugees resettled are second only to Cuban refugees. Further, the Indochinese refugees were unusually young, which makes it possible to construct a large sample of refugees who entered the U.S. as young children.

The second reason I focus on Indochinese refugees is that there is ample evidence that Indochinese refugees were disadvantaged in numerous ways. They immigrated from countries that were much poorer than the U.S.: PPP GDP per worker was roughly 3 percent of U.S. levels in 1980, and did not exceed 5 percent of U.S. levels from the start of the data in 1970 until 2005 (Heston et al., 2012). Additionally, they lived through the Vietnam War and related conflicts between communist and non-communist forces throughout the region. Refugees from Vietnam and Laos fled politically and ethnically motivated discrimination or "re-education" in labor camps at the hands of the Communists (Robinson, 1998). Cambodian refugees fled the Khmer Rouge, who forced the population to switch to an extreme, agriculturally oriented form of Marxism that brought widespread famine, while at the same killing an estimated 20 percent of the Cambodian population directly in the killing fields (Robinson, 1998). Those who decided to flee faced a difficult and unpredictable transit by land or by boat. Violence and piracy en route were common, as were delays or "push-backs" of refugees by the countries of attempted refuge. Those who reached refugee camps faced crowded and unpleasant conditions. Camps for the Khmer in Thailand were particularly notable for shortages of food, poor sanitation, and inadequate medical staff and treatment (United States General Accounting Office, 1980; Robinson, 1998).

These hardships had lasting effects on the refugees that were well-documented by doctors and psychologists after they arrived to the U.S. Groups of refugee children tested around the country consistently averaged between the 5th and 25th percentiles in height-for-age and weight-for-age, with a greatly elevated number scoring below the 5th percentile (Dewey et al., 1986; Barry et al., 1983; Peck et al., 1981). Follow-up studies showed that the gap in height-for-age did not diminish with time in the U.S. (Dewey et al., 1986). Early-arriving Indochinese refugees suffered from tuberculosis, hepatitis B, malaria, and intestinal parasites at rates more than an order of magnitude higher than the general U.S. population, although policy changes around 1982 increased the efficacy of medical screening of refugees (United States General Accounting Office, 1982; Barry et al., 1983; Goldenring et al., 1982). Likewise, psychological studies indicated the presence of post-traumatic stress disorder, depression, and other mental illnesses, with persistence in follow-up studies conducted as much as twenty years after migration (Lustig et al., 2004; Marshall et al., 2005).

The final and most important reason that I focus on the Indochinese refugees is that they mitigate or eliminate concerns about selection. My goal is to study children who experienced dramatic changes in their family income and environment during early childhood. If immigrant families are highly selected, then immigration may not change the family's income or environment that much. I review the available evidence on the selection of Indochinese refugees in Appendix A. That evidence suggests that refugees were selected to a modest extent, but not so strongly as to undo the intent of the experiment. For example, while refugees were generally more educated than their non-migrant peers, their education levels were still far below what would be typical in the U.S., with average years of schooling ranging from 2 to 10 for the various ethnic groups. Few immigrants spoke English before arrival and many came from farming, fishing, or military backgrounds that offered few skills of value in the U.S.

Even more importantly, my identification strategy relies on the assumption that immigrants who arrived at different ages of early childhood are not selected differently. For historical reasons the Indochinese refugees were unlikely to have been selected in this way, even if there were an underlying reason why families might wish to do so. The reason is that Indochinese refugees chose when to flee their home country, not when they would arrive to the U.S. The time elapsed in transit was long and variable. To help illustrate this point I make use of data from the Indochinese Health and Adaptation Research Project (IHARP), which collected a wealth of detailed data from a sample of 599 Indochinese refugees in the San Diego area in the 1980s.[2] The dataset includes the month and year a refugee left home and the month and year they arrived to the U.S., which allows me to provide concrete evidence on the long and variable transit times of refugees. The results are shown in Figure 1; data on the Lao are not available in IHARP. The median transit time ranged from 9 to 30.5 months, with the Khmer and Hmong facing longer transits. Just as importantly, there was large within-ethnic group variation in transit times: the standard deviation of transit times ranged from 6.1 to 22 months for this sample.

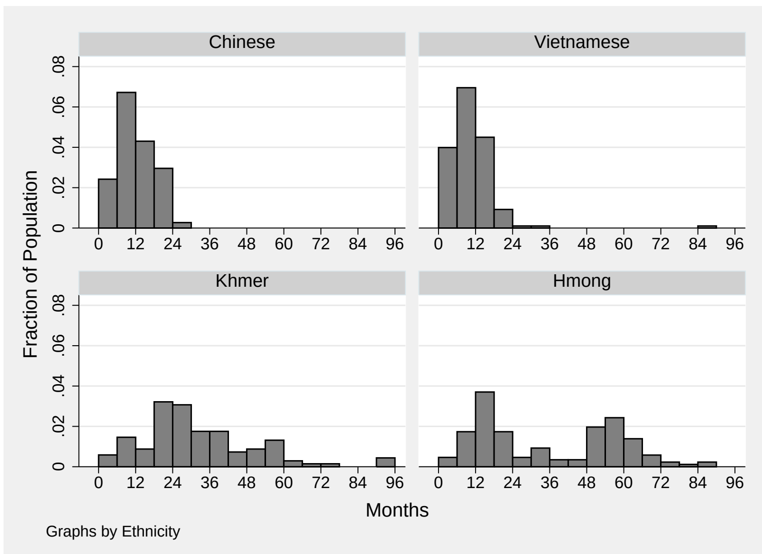


Figure 1: Time Elapsed in Transit by Ethnic Group

The long and variable waits experienced by refugees were for the most part the result of changes in the exogenous, national policies of the involved countries. Origin countries sometimes actively promoted and sometimes hindered attempts to escape. Many refugees needed multiple attempts to escape Vietnam in particular, because arranging transit by boat was difficult once the government worked to prevent refugee departures (Hung and Haines, 1996; Scott, 1989). It was common for refugees to be denied asylum in the first country they reached because countries of asylum worried about the moral hazard of offering easy asylum and worked to redistribute the burden of refugees to neighboring countries. For

[2]This research used the Indochinese Health and Adaptation Research Project, 1982–1984 dataset (made accessible in 2005 in paper and computer data form). These data were collected by Rubén Rumbaut, and are available through the archive of the Henry A. Murray Research Center at the Radcliffe Institute for Advanced Study, Harvard University, Cambridge, Massachusetts (Producer and Distributor).

refugees who reached the camps, the wait was determined mostly by U.S. policy.[3] The total refugee quota for the U.S. was set annually by the President and allocated across countries and camps of refuge in the context of domestic and international politics. Countries of first asylum were willing to accept refugees only if rich countries provided guarantees that they would resettle the refugees abroad (Robinson, 1998). The allocation of quota spots across different nationalities varied primarily due to these concerns, with immigrants from Vietnam arriving steadily throughout the period; immigrants from Laos peaking in the late 1970s; and immigrants from Cambodia peaking later, in the early 1980s (Haines, 1989). Domestic politics also played an important role in determining the number of admissions as concerns about the economic and health impact of refugees began to mount, particularly in light of an influx of Cuban refugees and the severe 1981–82 recession. Domestic politics also lead to changes in the intensity with which immigrants were screened, a tool used primarily to distinguish refugees (those who faced discrimination or violence at home) from economic migrants. More intensive screening increased wait times.

In light of the long, variable, and exogenous delays faced by refugees, it seems implausible that families were able to select when they entered the U.S. In Appendix A I test this hypothesis directly by examining the relationship between age at arrival of the child and family characteristics in the 1990 U.S. Census. I find correlations indistinguishable from zero in most cases, except for four family background characteristics of Vietnamese refugees. A final reason to rule out this form of selection comes from the large size of Indochinese refugee families. Using the 1990 U.S. Population Census, I find that conditional on having any foreign-born children, the average refugee family had three foreign-born children. In this case, it is not even clear what it means for the family to be selected based on the age of "the" child. In light of this evidence I will interpret my findings as measures of acquired human capital and not selection on the innate characteristics of the children or their families. I now turn to the results.

2.2 Benchmark Results

I study the adult outcomes of Indochinese refugees using the 2000 Population Census and the 2005–2011 American Community Surveys (ACSs), available online through Ruggles et

[3]Some refugees from Vietnam were able to bypass waiting in refugee camps entirely by participating in the Orderly Departure Program, which was an agreement for direct resettlement between the U.S. and Vietnamese governments. However, waiting for the entry permit and exit visa necessary for resettlement was an equally long and variable process, taking between two and five years for most refugees (Hung and Haines, 1996).

al. (2010). When combined, these datasets provide a large sample with fairly consistent questions and responses. I identify Indochinese refugees as those born in Vietnam, Cambodia, or Laos and who immigrated to the U.S. during the years of heavy refugee flows.[4] Although the Census does not have a variable that separately identifies refugees, independent sources indicate that essentially all immigrants arriving from these countries during these years were refugees (U.S. Immigration and Naturalization Service, 1980–2000). I construct age at arrival using the year the data were collected, the respondent's age at that time, and the year of immigration. For the baseline analysis I limit the sample to those who immigrated by age 5. I also include the American-born in the sample for comparison.

I use two different outcomes in the main analysis. The first is the wages of refugees. For this outcome I restrict the sample to include only those workers who would typically be used in a wage regression: those aged 23–65, not enrolled in school, who work for wages, usually work at least 30 hours a week and worked at least 30 weeks the previous year, have between 0 and 40 years of potential experience, and report positive wage and salary income the previous year. I construct hourly wage as annual wage and salary income divided by the product of hours worked per week and weeks worked in the previous year.[5] I then regress

$$\log(W) = \beta X + \sum_{a} \alpha_a d_a + \sum_{y} \omega_y d_y + \sum_{aa} \phi_{aa} d_{aa} + \varepsilon, \tag{1}$$

where W is the constructed wage and X is a vector of control variables that consists of state of residence and gender dummies. d_a is a dummy taking the value of 1 if the person's age is a, while d_y and d_{aa} are dummy variables for the year of the dataset and age at arrival. Greek letters denote the corresponding coefficients. The coefficient of interest is ϕ_{aa}, which captures the log-wage of a refugee who arrived at age aa, relative to an American-born worker with the same age, gender, and state of residence. The immigration literature often considers more general forms of equation (1); I show in Appendix B that the coefficients on age at arrival are identified, even in a more general model that allows for cohort or assimilation effects.

Figure 2 shows the results in the format that I will use for the rest of the paper. It plots the estimated coefficients ϕ_{aa} and the corresponding 95 percent confidence interval against age at arrival aa. Refugees with higher age at arrival spent more years of their early childhood in their birth country or in refugee camps and fewer years in the U.S. Given the poverty and

[4] For Vietnam, the years are 1976–1990; for Cambodia, 1976–1994; and for Laos, 1976–1996.

[5] From 2008 onward, weeks worked is reported in categories. I use the 2007 ACS data to estimate the mean value of weeks worked within each category, and apply this to the 2008–2011 ACSs.

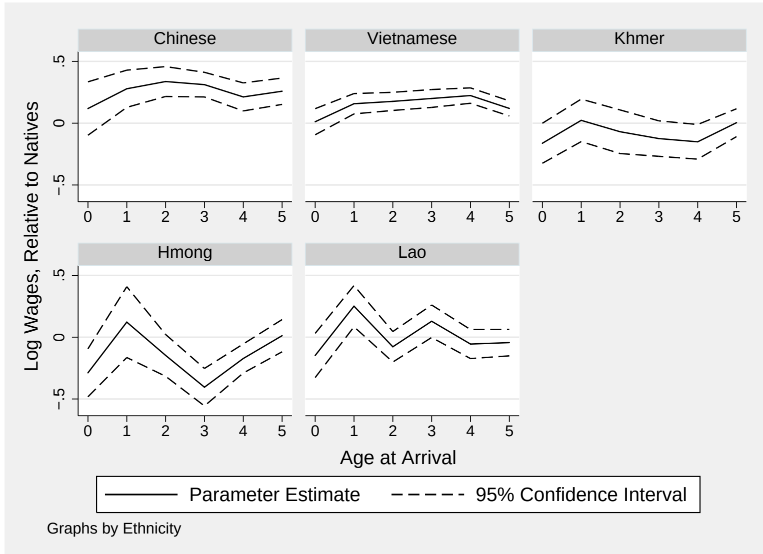


Figure 2: Log-Wages by Age at Arrival

difficult circumstances to which Indochinese refugees were exposed, it is perhaps natural to expect that refugees who were older upon arrival and who had longer exposure to their birth country would have worse outcomes. Graphically, this expectation would imply that ϕ_{aa} declines in aa, or a negative slope to the plotted lines. The striking finding is that there is instead no trend for any of the five ethnic groups. Formally, I cannot reject the null hypothesis that the slopes are zero. The level of the lines reflects the performance of refugees relative to natives. The Chinese and Vietnamese refugees have better outcomes than natives; the Lao, about the same; and the Khmer and Hmong, modestly worse.

As a second outcome I investigate refugees' years of completed schooling. One advantage of using schooling data is that I can impose fewer sample selection criteria; I limit the sample only to those ages 23–65 who are no longer enrolled in school. I transform the provided educational attainment variable into years of schooling in the usual way. I then regress the years of schooling on the same set of control variables as in equation (1). Figure 3 again plots the coefficients ϕ_{aa} against the age at arrival. The same basic finding obtains: there is no trend relationship between age at arrival and completed schooling for Indochinese refugees.

These are the benchmark empirical results of the paper. In Appendix C I show that there is also no trend by age at arrival for a number of other socioeconomic outcomes, including probability of being employed, total earnings, probability of graduating college, and so on. I also show that the result is not driven by breaking out the five ethnic groups; the

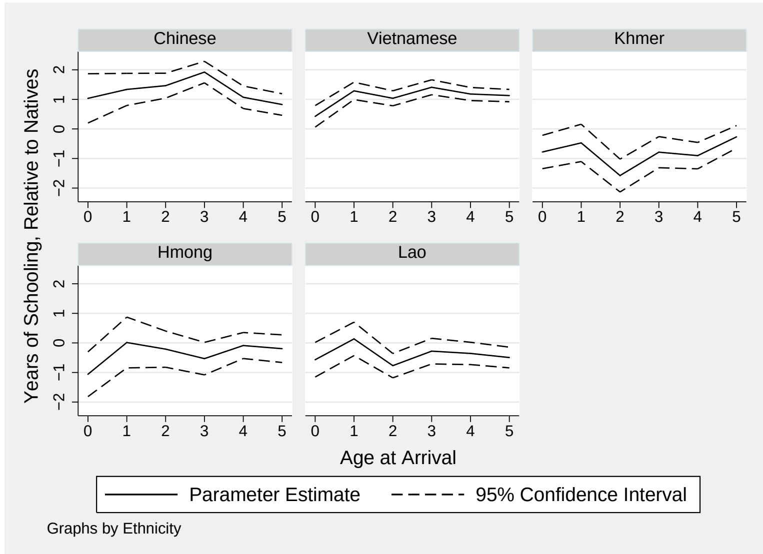
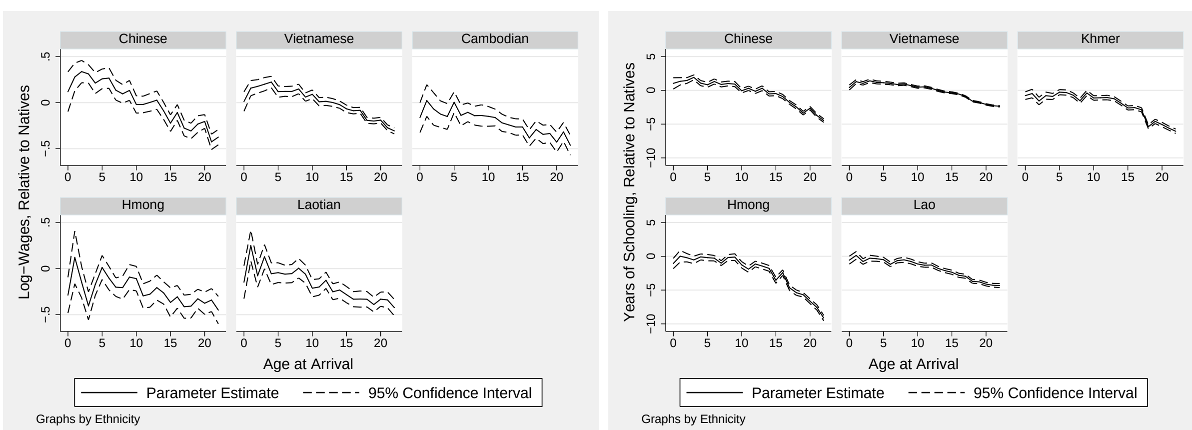


Figure 3: Years of Schooling by Age at Arrival

same results obtain if I instead break results out by country of birth or language spoken at home. Finally, I show that similar results apply for refugees who arrived before or after the increased intensity of medical screenings. In general, there seems to be little difference in the late-life outcomes of Indochinese refugees who arrived to the U.S. at the beginning and at the end of early childhood.



(a) Log-Wages (b) Years of Schooling

Figure 4: Outcomes by Age at Arrival for Older Ages

Finally, I want to emphasize that the results presented here are special to early childhood. To make this point I re-estimate wages and schooling as a function of age at arrival using a sample of refugees who arrived as late as age 22. Figure 4 shows the results. Outcomes begin to decline in age at arrival somewhere around age 10. If I pool the different ethnic groups and use nonlinear least squares to estimate a spline with a single endogenous breakpoint, the regression locates that point to 7.5 years for wages and 11.9 years for schooling. The cumulative effect of worsening outcomes is large for refugees who arrived in early adulthood. Those who arrive at age 22 have 5–10 fewer years of schooling and 30–60 percent lower wages than refugees who arrive in early childhood, with the magnitude varying somewhat by ethnic group.

2.3 Results from Other Immigrants

As I emphasized above, Indochinese refugees are particularly useful because they offer a large sample size, poor backgrounds, and diminished concerns about selection. Most importantly, historical circumstances make it implausible that immigrants who arrived at different ages were selected differently. On the other hand it may be that this form of differential selection (on the age of the young child) is itself implausible. In this case, findings from other groups offer useful additional empirical evidence. In any case, it is useful to know the patterns for immigrants from other countries. Here, I focus on four such groups.

First, I collect a sample of refugees who arrived as children from Afghanistan and Ethiopia; although these countries were quite poor, there were fewer immigrants and their parents are generally considered to have been more selected.[6] Since my strategy is to look at the late-life outcomes of refugees who entered as children, these are the most recent groups of poor-country refugees that I can study; more recent groups such as the Somali or Bosnian refugees are not old enough yet for my analysis. The second group is Cuban immigrants. Cuba is much richer than Vietnam, Cambodia, or Laos, but still only one-fourth as productive as the U.S. on a PPP GDP p.w. basis in 2005; further, a substantial number of Cubans entered the U.S. I make no effort to disentangle Cuban refugees from other immigrants (such as family reunification cases). The third group is all immigrants from countries with 2005 PPP GDP p.w. less than 5 percent of the U.S.[7] These immigrants come from poor countries with worse environments for human capital formation, but are likely to be quite selected. Finally, I study immigrants from Mexico.

[6] I use immigrants who entered between 1980 and 1993 in my sample, which were generally years of high refugee flows (U.S. Immigration and Naturalization Service, 1980–2000).

[7] The countries distinguished in the Census and so included in my sample are: Haiti, Bangladesh, Nepal,

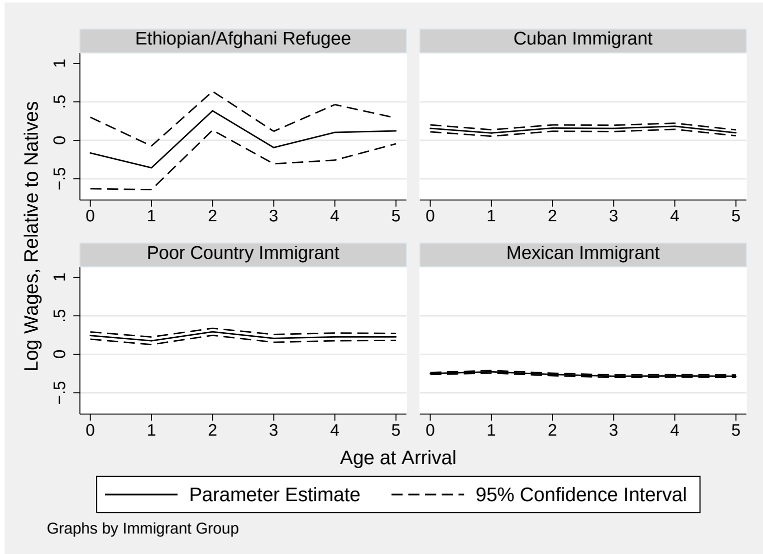


Figure 5: Log-Wages by Age at Arrival for Other Immigrant Groups

Figure 5 shows the results for the log-wage patterns by age at arrival for these four groups. Although each is less ideal than the baseline sample, they confirm the same pattern: there is no trend in log-wages by age at arrival up to age 5. Similar results obtain for other socioeconomic characteristics. At this point it seems that the lack of trend between socioeconomic outcomes and age at arrival is quite general. I now turn to asking how this fact can be reconciled with the large literature on the importance of early childhood human capital.

3 A Model of Early Childhood Human Capital

3.1 Human Capital Accumulation and the Labor Market

The model describes the human capital accumulation and labor market decisions of families in countries $i \in \{1, 2, ..., I\}$. Within each country there is a continuum of heterogeneous families consisting of one working adult and one child that is newly born at time 0; I think of this model as describing the choices of a single birth cohort around the world. There are two forms of ex-ante heterogeneity in the model. First, children are endowed with some human capital at birth h_0. Second, parents are endowed with human capital h_p, which is

Ghana, Guinea, Liberia, Senegal, Sierra Leone, Kenya, Somalia, Tanzania, Uganda, Zimbabwe, and Eritrea.

determined by events that happen before the child is born. The two-dimensional human capital endowment is drawn from a joint distribution $F_i(h_0, h_p)$ defined on $(0,\infty)^2$, which I allow to vary across countries.

Time is continuous and the parent and child are both infinitely lived. Each is endowed with a single unit of time at each instant. The child's life is split into three periods: early childhood, school, and work. Early childhood includes the first five years of the child's life, before schooling. School lasts from age 5 until an endogenously chosen graduation date. After graduation the child joins his or her parent in the labor force and works into the infinite future.

The heart of the model is the human capital production function, which explains how human capital is generated given the family's endowment (h_0, h_p), the country i that they live in, and the investments they make in the child during early childhood and school. Following the literature, I assume that human capital production occurs in two distinct stages. Early childhood human capital h_c is determined by human capital at birth h_0 and a composite of the investments made during early childhood x_c, combined using a CES production function:

$$h_c = \left[\lambda_c h_0^{\frac{\sigma_c-1}{\sigma_c}} + (1-\lambda_c) x_c^{\frac{\sigma_c-1}{\sigma_c}}\right]^{\frac{\sigma_c}{\sigma_c-1}}. \tag{2}$$

λ_c is the weight on human capital at birth and $1-\lambda_c$ the weight on the composite investment, while σ_c is the elasticity of substitution between the two.

The composite investment is a weighted combination of four different underlying inputs. The first input is an exogenous, country-specific environmental effect, denoted by z_i, which includes the prevalence of diseases or the quality of medical infrastructure. The second input is the goods purchased for the child, denoted by g_c, which includes books or vaccines. The final two inputs consist of the time spent by parents with their children p_c and the human capital of the parents, h_p, which I allow to enter the composite investment separately. Given these inputs, x_c is determined using a power function:

$$x_c = z_i^{\omega_{c1}} g_c^{\omega_{c2}} h_p^{\omega_{c3}} p_c^{\omega_{c4}}. \tag{3}$$

The ω_c are weight parameters that determine the relative importance of the different inputs.

At age 5 children start the second phase of their life, which is formal education. They remain in school for an endogenously chosen period of time S, graduating at age $5+S$.

Their human capital at graduation is again a function of their human capital at the start of school h_c and the composite investment made during the school years x_s:

$$h_s = \left[\lambda_s h_c^{\frac{\sigma_s - 1}{\sigma_s}} + (1 - \lambda_s) x_s^{\frac{\sigma_s - 1}{\sigma_s}}\right]^{\frac{\sigma_s}{\sigma_s - 1}}. \tag{4}$$

λ_s is the weight on early childhood human capital and $1 - \lambda_s$ is the weight on composite investment, while σ_s is the elasticity of substitution between the two.

The composite investment during schooling depends on five inputs. The first four are similar to early childhood: country environment z_i; goods g_s; and parental inputs p_s and h_p. Finally, the human capital also depends on how long children stay in school, S. The terms are again combined using a power function,

$$x_s = z_i^{\omega_{s1}} g_s^{\omega_{s2}} h_p^{\omega_{s3}} p_s^{\omega_{s4}} S^{\omega_{s5}}, \tag{5}$$

where the weights on the various factors are allowed to vary over the life cycle.

The human capital production function in equations (2) – (5) borrows heavily from the microeconomic literature on early childhood human capital. Two key insights from that literature will be useful for understanding how the model works. First is the importance of the elasticity parameters σ_c and σ_s (Cunha and Heckman, 2007; Cunha et al., 2010). Their role in this model is to determine the extent to which a disadvantage in the form of low human capital at birth or low early childhood human capital can be remediated by subsequent investments. Cunha et al. (2010) estimate that $\sigma_c > 1$ and $\sigma_s < 1$, indicating that it is relatively easy to remediate low human capital at birth but much more difficult to remediate low early childhood human capital. Second is the importance of allowing for multiple types of investments and allowing their importance to vary over the life cycle. Del Boca et al. (2012) estimate that the relative role for goods inputs and parental inputs changes over the life cycle. My empirical work below offers further support for this finding and shows that it has important macroeconomic implications.

I embed the human capital production function into a simple lifetime income maximization problem. I assume that human capital is fixed after graduation. The Indochinese refugees I studied in the empirical implementation were quite young, which helps to minimize the importance of this assumption. After graduation, the individual enters the labor market and works full-time. They are endowed with a linear production technology that turns the worker's h_s units of human capital into $A_i(t)h_s$ units of the single output good, which can be used for consumption or as the goods inputs in human capital production. $A_i(t)$ is an exogenous TFP term. I assume that $A_i(t)$ grows at a rate γ that is common across countries, but allow for differences in the initial level of productivity $A_i \equiv A_i(0)$.

Finally, I assume that each family behaves altruistically and that children have no preferences over whether they work or study. In this case I can focus on the family's income maximization problem and ignore the (trivial) utility maximization problem given optimal income. Families then choose the duration of schooling for children and the quantity of the two types of inputs at the two stages of the life cycle to maximize the present discounted value of lifetime earnings net of the cost of goods and the foregone earnings of the parents:

$$\max_{g_c, g_s, p_c, p_s, S} \quad \int_{5+S}^{\infty} e^{-rt} A_i(t) h_s dt - g_c - e^{-5r} g_s - A_i h_p p_c - A_i e^{5(\gamma - r)} h_p p_s \tag{6}$$

where h_s is derived from equations (2) – (5) and r is the exogenous interest rate. Here I assume that parents purchase all of the market goods and forego all labor earnings at the start of each of the respective stages of the life cycle, that is, at date 0 for early childhood and at date 5 for school. This assumption simplifies the discounting of investments without foregoing any insights. I now turn to model's predictions for immigrants.

3.2 Model Predictions

I have to take a stand on immigrants' beliefs about the possibility of immigration in order to generate the model's predictions. I start with the simplifying assumption that immigration is entirely unanticipated. In this case, families in poor countries (with low A_i and z_i) make the optimal investments in their children under the assumption that they will live their entire lives in their native country. When the child is aa years old, they are unexpectedly moved to the U.S, at which point the family re-optimizes all future investments, taking past investments as given. I view these beliefs as the natural starting point in the analysis of refugees, but I also consider alternatives in Section 5.

Given these beliefs, the model predicts that human capital and wages of poor country immigrants should decline in age at arrival under fairly generic conditions. This prediction clearly stands at odds with the evidence from Figure 2. The first goal of this section is to clarify the logical steps that lead to this prediction. Once these steps are clear, it is straightforward to see what sets of specific conditions will allow the model to replicate the data.

Table 1: Elasticity of Human Capital with Respect to Exogenous Characteristics

	Adult Human Capital, h_s	Early Childhood Human Capital, h_c
h_0	$\frac{\lambda_c\lambda_s}{1-\Psi_2-\Psi_4}$	$\lambda_c + \frac{\lambda_c\lambda_s(\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}$
z_i	$\frac{\Psi_1}{1-\Psi_2-\Psi_4}$	$\Phi_1 + \frac{\Psi_1(\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}$
A_i	$\frac{\Psi_2}{1-\Psi_2-\Psi_4}$	$\Phi_2 + \frac{\Psi_2(\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}$
h_p	$\frac{\Psi_3-\Psi_4}{1-\Psi_2-\Psi_4}$	$\Phi_2 + \Phi_3 + \frac{(\Psi_2+\Psi_3-1)(\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}$

Note: $\Psi_i \equiv \omega_{ci}\lambda_s(1-\lambda_c) + \omega_{si}(1-\lambda_s)$ and $\Phi_i \equiv \omega_{ci}(1-\lambda_c)$ are the technological weight parameters on factor i in h_s and h_c in the isoelastic case.

The first step is to show that the model generically predicts that there is a disadvantage to growing up in a poor instead of a rich country. The model can generate differences in human capital from each of the four exogenous factors: z_i, h_0, h_p, and A_i. To illustrate these forces, it is useful to specialize momentarily to the isoelastic case, $\sigma_c = \sigma_s = 1$, in which case the model provides simple closed-form solutions for investment and human capital accumulated as a function of these exogenous variables. The appropriate elasticities of h_s and h_c are given in Table 1; derivations are reserved to Appendix D. An additional benefit of the isoelastic case is that it nests a two-stage Ben-Porath human capital production function, which is commonly used in the macroeconomic literature; I exploit this in a moment.

Two of the effects in Table 1 are straightforward. Children who grow up in countries that offer a better environment, or children who have more human capital at birth, will also have higher human capital at the end of early childhood and into adulthood. The quantitative importance of these effects is controlled by two sets of parameters, which are easiest to understand in the expressions for h_s. The numerator of these expressions is simply the technological coefficient on h_0 or z_i; the more weight the technology puts on these factors, the more important they are for accumulated human capital. This effect is amplified because families allocate more of the endogenously chosen inputs (goods and parental time) to children who are more able. This amplification effect is reflected in the markup $(1-\Psi_2-\Psi_4)^{-1}$ for adult human capital, where Ψ_2 and Ψ_4 are the technological weights on goods and parental time and hence $\Psi_2 + \Psi_4$ is the total return to scale of endogenous factors. $\Psi_2 + \Psi_4 < 1$ is thus a necessary condition for an interior solution and is assumed throughout.

The latter two effects are more subtle. Children whose parents have more human capital would seem to enjoy an advantage. However, higher parental human capital also enables parents to earn higher wages in the labor market. For some parameter conditions, parents with higher human capital substitute away from time with their children and towards market work. In extreme cases this substitution is sufficiently strong that the total parental input actually declines in parental human capital. Both of these predictions contradict the existing data, which suggest that parents with more human capital spend more time with their children (Guryan et al., 2008). Finally, A_i provides the TFP multiplier effect that is well-known in the literature (Manuelli and Seshadri, 2010; Erosa et al., 2010). The logic of the multiplier is that higher A_i lowers the cost of goods relative to wages. Hence, families in high-TFP countries endogenously allocate more market goods to their children, which in turn raises average human capital.

I assume that children who arrive at different ages have the same underlying distributions of h_0 and h_p, consistent with the historical evidence surrounding Indochinese refugees. In that case, the difference between refugees who arrived at age 0 and those who arrived at age 5 are due to differences between poor and rich countries in z_i and A_i. The expressions in Table 1 show that h_c is increasing in z_i and A_i under fairly general conditions on the human capital production function. Thus, the model generically predicts that children who arrive at age 5 will find themselves with less human capital than children who arrived at age 0.

The second step is to characterize the conditions under which an initial disadvantage in human capital at age 5 persists until the child enters the labor market. Under the isoelastic case this condition is straightforward; it merely requires $\lambda_s > 0$. This condition ensures that early childhood human capital has a positive weight as an input to subsequent human capital accumulation. Along with the conditions outlined above, this condition is sufficient to ensure that human capital and wages in adulthood are declining in age at arrival to the U.S.

Qualitatively, this prediction stands at odds with the data. It is also useful to have an idea of how large the gap between model and data are quantitatively. To this end, I study a conservative parameterization of a two-stage Ben-Porath human capital production function:

$$h_c = g_c^{0.25},$$
$$h_s = g_s^{0.25}(Sh_c)^{0.4}.$$

This parameterization is conservative in three senses: it abstracts from country environment effects; it puts a low weight on market-purchased goods relative to the literature; and it puts a low weight on early childhood human capital stocks in the second stage relative to the literature.[8]

I simulate a baseline rich economy with $A_R = 1$ and solve for the optimally chosen h_{sR}. I then calibrate a second poor economy by setting A_P so that the endogenously chosen h_{sP} leads to an income gap between rich and poor of $h_{sP}A_P/h_{sR} = 0.03$, in line with the gap between the Indochinese countries and the U.S. in 1980. Finally, I simulate the human capital decisions and wages for natives of both countries, as well as for immigrants who migrate from the poor to the rich country at the beginning of their life (age 0), after early childhood (age 5), and immediately following graduation (which I call age 22).

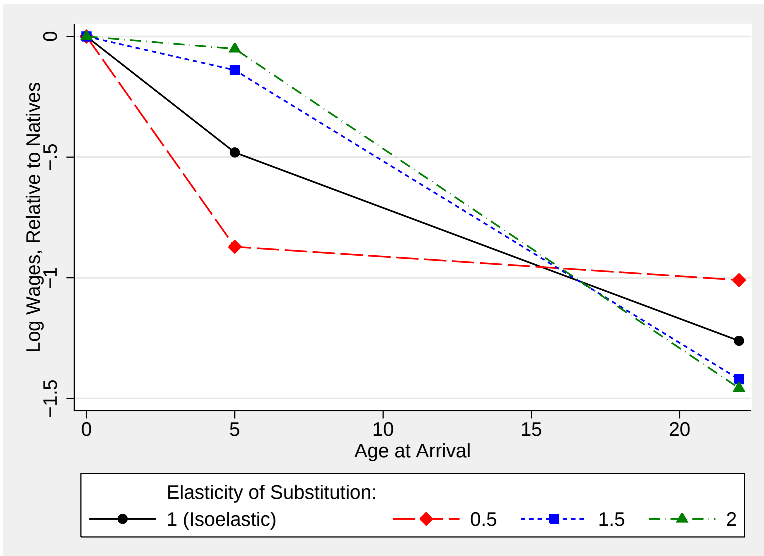


Figure 6: Model-Predicted Log-Wage Patterns for Immigrants

Figure 6 expresses the quantitative predictions of this parameterized experiment in the same format that I used for the empirical section. The solid line shows the results for the isoelastic case. Even a conservative calibration of the model implies a large disadvantage for late-arriving relative to early-arriving refugees; the slope of wages in age at arrival is roughly -0.1, or about ten percent lower wages per year spent abroad. Hence, the gap between the model and data is quantitatively as well as qualitatively important.

The analysis so far has been carried out under the maintained assumption that $\sigma_c = \sigma_s = 1$.

[8]The parameters required are $\lambda_c = 0$, $\lambda_s = 1$, $\omega_{c2} = 0.25$, $\omega_{s2} = 0.417$, and $\omega_{s5} = 0.667$, with the remaining ω set to 0. By comparison, Erosa et al. (2010) use an exponent of 0.40 on market goods in the school period; my value is smaller than naively projecting their value back to the early childhood period. Manuelli and Seshadri (2010) set the exponent on market goods to 0.70 in early childhood and 0.316 thereafter. I also set $r = 0.04$ and $g = 0.02$, although the conclusions are not sensitive to these parameters.

The isoelastic case is convenient for deriving closed form solutions and comparing my results to the existing literature, but it is not innocuous. In particular, σ_s plays a critical role here.[9] $\lambda_s > 0$ ensures that children with higher human capital at age 5 have higher human capital in adulthood if all other investments are held equal. Of course, all other investments need not be held equal; parents can choose to invest more (or less) in their children in response to low human capital at age 5. σ_s controls the extent to which investments made during schooling can be used to remediate deficits in early childhood human capital and the willingness of families to engage in such efforts. To show the importance of σ_s, I re-parameterize the model under different assumptions on the value of σ_s. For each assumed value, I choose a new A_P to generate the poor-rich income ratio of 0.03, then study the model's predictions for the late-life wages of immigrants from the poor country.

The results are plotted in Figure 6. The inelastic case draws on the work of Cunha et al. (2010), who estimate $\sigma_s \approx 0.5$. In this case it is even more difficult for immigrant families to remediate the low early childhood human capital of their children with investments made after arrival, leading them to make fewer such investments. Hence, outcomes are even more strongly declining in age at arrival. On the other hand, allowing for more elasticity than the baseline case generates smaller gaps between refugees who arrive at ages 5 and 0, consistent with the data. For example, $\sigma_s = 1.5$ implies a wage loss of around 3% per year spent abroad, while $\sigma_s = 2$ implies a wage loss of around 1% per year spent abroad. This difference is quantitatively small and would be difficult to detect empirically.

In sum, the generic prediction of the model is that wages of refugees decline in their age at arrival. This prediction rests on three logical steps. First, there is a disadvantage of growing up in poor rather than rich countries: either the poor country environment directly lowers human capital, or low incomes lead parents to spend less on their children. Second, early childhood human capital differences persist to adulthood if all other factors are held equal ($\lambda_s > 0$). Finally, remediation of early childhood human capital gaps is sufficiently difficult that families' endogenously chosen investments do not close such gaps entirely (small σ_s). Given this logic, it is clear that there are three particular parameterizations of the model that are capable of matching the data:

- Case 1: $\omega_{c1} = \omega_{c2} = 0$. This case implies that goods and country environment are not important inputs in early childhood, which also implies that there is no strong

[9] As a reminder, σ_c controls the extent to which differences in human capital at birth h_0 can be remediated through investments made during early childhood. Since I assume that refugees who arrive at different ages have the same distribution of h_0, σ_c plays little role for my findings.

disadvantage to growing up in a poor country.[10]

- Case 2: $\lambda_s = 0$. This case implies that any differences in early childhood human capital are irrelevant for subsequent outcomes. This case corresponds to the idea that early childhood human capital does not matter.

- Case 3: $\sigma_s > 2$. This case implies that there are disadvantages to growing up in poor countries, but that refugees and their families find it easy to remediate deficits in early childhood human capital after arrival to the U.S.

It is important to differentiate between these three cases because they have very different implications for the human capital production function as well as cross-country human capital differences. Cases 1 and 2 both suggest that early childhood human capital is unlikely to be a significant contributor to cross-country income differences. On the other hand, case 3 is largely silent about how large such differences may be, except to note that they can be remediated by refugees using investments made while in school.

The experience of refugees supplies only one additional piece of evidence to discriminate between these cases. Different values of σ_s imply different relationships between early childhood human capital h_c and subsequent investment behavior. The most usxedictions concern the duration of schooling S, since this is an outcome that can be measured in the Census. For $\sigma_s > 1$, the model implies that completed years of schooling should decrease in early childhood human capital. To translate this into the experience of refugees, the implication is that years of schooling should rise with age at arrival. Intuitively, this is because high σ_s induces families of children who were older at arrival and had lower h_c to invest more in their children to remediate their human capital deficit, including sending them to school longer. However, this prediction contradicts the data in Figure 3, which shows that years of schooling are unrelated to age at arrival. The model is consistent with schooling being unrelated to age at arrival if $\sigma_s = 1$, or if refugees who arrive at different ages have roughly the same human capital at the end of early childhood. Hence, the school decisions of refugees point towards cases 1 and 2 rather than case 3.

For further evidence I turn to the findings of the existing literature on early childhood human capital. I show that studies in this literature provide useful evidence that can be

[10] Inspection of Table 1 shows that it is also necessary to set $\omega_{s1} = \omega_{s2} = 0$ for the appropriate elasticities to be exactly zero. The reason is that the model admits a second-order effect whereby parents spend more time with children in early childhood in rich countries in anticipation of the advantages the children will enjoy in their school years stemming from $\omega_{s1} > 0$ or $\omega_{s2} > 0$. This effect is quantitatively small so I abstract from it.

used to discriminate among these three cases.

3.3 Evidence from Existing Literature

The existing literature on early childhood human capital provides substantial evidence on the question at hand. The relevant literature comes in three strands. First, Cunha et al. (2010) and Del Boca et al. (2012) structurally estimate the human capital production function. As was mentioned before, Cunha et al. (2010) find $\sigma_s \approx 0.5$, inconsistent with high values of the elasticity of substitution. They also find a large role for early childhood human capital in the determination of subsequent outcomes, inconsistent with $\lambda_s = 0$. On the other hand, they estimate a human capital production function with only one composite investment good, so they cannot speak to the relative importance of different inputs. Del Boca et al. (2012) provide evidence on exactly this final point. They estimate the relative importance of goods and time spent with children over the life cycle. They find that the importance of time spent with children is relatively high in early years and declines with age, while the opposite pattern prevails for market-purchased goods. This evidence can be thought of as supporting $\omega_{c4}/\omega_{c2} > \omega_{s4}/\omega_{s2}$; my findings are consistent with this but even stronger. Their estimates also support a role for early childhood human capital investments in affecting late-life outcomes, again inconsistent with $\lambda_s = 0$. However, they constrain $\sigma_s = 1$ and hence provide no information on this point.

A second strand of the literature draws on evaluations of early childhood interventions with experimental designs. Many of these studies focus on remedial preschool programs such as the Carolina Abercederian or Perry Preschool programs. Cunha et al. (2006) and Blau and Currie (2006) provide reviews of the structure and effects of these programs. The basic design is to provide a preschool-type program for the children paired with home visits that target the parents. These programs consistently have an impact on late-life outcomes, including test scores, grade retention, high school graduation rates, and college start rates. The longest-lasting survey, the Perry Preschool Experiment, has now run for long enough to verify higher earnings for participants in adulthood. This evidence adds further support to the restriction $\lambda_s > 0$.

Finally, a third strand of the literature studies the effects of extreme deprivation on early childhood development. This strand is the closest in spirit to my paper. Within it, there exists one set of studies nearly identical to my own, following the development of Romanian orphans. A combination of pro-natalist policies and economic stagnation under Ceauşescu led parents to abandon large numbers of children to state-owned orphanages in Communist Romania. Conditions in the orphanages were dire: children were mostly confined to cots, given few toys, spoken to or allowed to play rarely, and fed primarily gruel. When Ceauşescu's regime fell in 1989, some of these orphans were adopted abroad. Thus, Romanian orphans provide a natural experiment of children who leave behind a poor country with a damaging environment that is very similar to my own. The critical difference is that Romanian orphans received dramatic upgrades in their parental input (h_p and p_c), trading essentially no parental input in the orphanages for the inputs of their adoptive parents in Britain. On the other hand, 95 percent of Indochinese refugees in my sample migrated with at least one biological parent and so had similar parental inputs before and after migration.[11] Thus, the critical distinction between the refugees and Romanian orphans is that the refugees *only* changed their country environment and labor market productivity, z_i and A_i, while the Romanian orphans also changed their parental input h_p and p_c.

A team of researchers has intensively studied the ongoing progress of a sample of Romanian orphans adopted into Britain (Rutter and The English and Romanian Adoptees (ERA) study team, 1998; O'Connor et al., 2000; Beckett et al., 2006, 2010). To date they have completed surveys of the orphans at ages 4, 6, 11, and 15. They compare outcomes for British-born versus Romanian-born orphans, but also for Romanian-born orphans adopted in different age groups (<6 months; 6–24 months; and 24–42 months). Hence, they analyze the same relationship between outcomes and age at arrival as I do. Their most striking result is a consistent, negative relationship between age at arrival and outcomes, exactly the opposite of my finding. Romanian orphans who arrive before they are six months old do as well as British-born adoptees and better than Romanian orphans who are adopted at older ages. They also find smaller and less persistent differences between those who were adopted at 6–24 months and those who were adopted at 24–42 months, with the former sometimes doing better and sometimes roughly the same.

The difference between the experiences of Romanian orphans and Indochinese refugees suggests that parental inputs, rather than goods or country environment, are the critical input in early childhood. The experience of Romanian orphans also suggests that early childhood human capital matters for late-life outcomes ($\lambda_s > 0$) since the differences found at age 4 are persistent until at least age 15. Finally, it suggests that it is difficult to remediate deficits in early childhood human capital that have developed as little as six months after birth (low σ_s).

[11] I compute this statistic using the 1990 U.S. Census, which captures the household structure of Indochinese refugees when they were young enough to be living in a household headed by someone else.

In summary, the evidence from a large number of studies supports long-lasting effects of early childhood that are inconsistent with the hypothesis that $\lambda_s = 0$. A smaller number of papers have addressed the potential for remediation of early childhood human capital deficits and found that it is difficult, which implies a low value of σ_s. A low value of σ_s is also consistent with the post-migration schooling investments of refugees documented in this paper. On the other hand, a small role for goods or country environment effects in early childhood ($\omega_{c1} = \omega_{c2} = 0$) provides the unique explanation that is consistent with the experience of Indochinese refugees as well as the existing literature on early childhood human capital. In the next section I address what a model parameterized along these lines implies for cross-country early childhood human capital differences.

4 Implications for Cross-Country Human Capital Differences

The findings of the previous section have strong implications for cross-country differences in early childhood human capital. As has been emphasized throughout, there are two key advantages to growing up in a rich rather than a poor country. The first is that high TFP makes goods inputs to human capital production cheap, providing rich country families incentives to purchase more books or vaccines for their children. The second is that the environment may itself be better, perhaps by offering better infrastructure or a lower disease burden. The findings of the previous section suggest that neither of these channels is quantitatively large in early childhood. This severely limits the scope for large cross-country differences in early childhood human capital.

The results are consistent with a critical role for parents during early childhood. Indeed, the differences in the level of outcomes between ethnic groups are strongly correlated with the differences in the backgrounds of their parents. I review several measures of the parental characteristics for refugees in Appendix A. Chinese and Vietnamese children had the the most educated parents who were the most likely to be literate pre-migration, had the most usable occupational skills, and were most likely to have paid a bribe to facilitate escape (suggesting higher family income). These children had the highest level of outcomes. Hmong children had the least educated parents, most of whom were illiterate pre-migration and worked as farmers, fishers, or soldiers; they also had the lowest levels of outcomes. In this sense the experiences of refugees support an important role for parental inputs. A perhaps more surprising result is that the level gaps between ethnic groups or between refugees and natives are small, despite the striking difference in parental backgrounds. For example, Hmong children who arrived in early childhood had wages just 18 percent lower than U.S.-born children, and 54 percent lower than Vietnamese children. These facts suggest that cross-country differences in acquired parental characteristics are also not likely to generate sizable cross-country differences in early childhood human capital.

The paper's findings for human capital formed during the schooling years are more mixed. Figure 4 shows that both wages and school completion decline in age at arrival past age 10 or so. There are a number of possible explanations for this finding. In the context of the model it could be the case that ω_{s1} or ω_{s2} is positive, so that goods or environment are important for accumulating human capital during the schooling years, consistent with the findings of Manuelli and Seshadri (2010) and Erosa et al. (2010). However, a number of alternative explanations seem equally promising. For example, age 10 is approximately when cognitive plasticity in children seems to decline, implying a lower ability to adapt (Lenneberg, 1967). This explanation is particularly promising in light of the finding of Bleakley and Chin (2010) that English language ability of immigrants declines in age at arrival past age 10 or so. Alternatively, these patterns could reflect low exposure to formal education or low education quality in these countries (Schoellman, 2012). The evidence at hand is not well-suited to discriminate between these theories.

5 Alternative Explanations

The main empirical finding of the paper is that there is no trend relationship between late-life outcomes and age at arrival for immigrants who arrive to the U.S. by age 5. The baseline interpretation of this fact is that goods inputs and country environment matter little for early childhood human capital formation, which implies small cross-country differences in early childhood human capital. In this section I consider two alternative explanations. First, I allow for immigration to be an anticipated event, contrary to the simplifying assumption in the baseline model. Second, I consider whether measurement error provides a plausible alternative explanation.

5.1 Anticipated Immigration

In the baseline model I treat immigration as an unanticipated, one-off event. For the early waves of the Indochinese refugees this is probably the natural assumption, because the possibility of immigrating from these countries before the U.S. started accepting refugees was trivial. For example, over the entire 1960s the U.S. accepted 1200 immigrants from Cambodia, 100 from Laos, and 4600 from Vietnam. These figures were minuscule compared either to the 3.3 million immigrants the U.S. accepted that decade or the nearly 50 million person population of the three countries in 1965 (Heston et al., 2012). However, the Indochinese refugee flows were protracted, and the consensus is that by the mid-1980s Indochinese citizens became aware of a significant probability of third-country resettlement for refugees. The relevant question is: to what extent does that change the model's predictions about the relationship between age at arrival and late-life outcomes? The main concern is that forward-looking parents may have anticipated resettlement and invested heavily in their young children, which could help explain the lack of a strong relationship between late-life outcomes and age at arrival.

This hypothesis is logically consistent in the context of the model. If parents view immigration as a lottery with some probability of resettlement then they will alter their investments in their children accordingly. The strength of this effect depends on how well parents can predict resettlement and the parameterization of the model. As an extreme example, I consider the parameterization from section 3.2 and figure 6, but allow future refugees to anticipate their resettlement perfectly. Under this change in beliefs, the model predicts that future migrant families invest as much in their children as families in the U.S. After resettlement, refugees would have the same late-life outcomes as children born in the U.S., with no relationship between age at arrival and late-life outcomes. In other words, allowing future refugees to anticipate their resettlement generates predictions consistent with the data.

There are at least three reasons to be skeptical of this extreme case. First, the probability of resettlement was much lower than this calculation suggests. As was mentioned earlier, less than half of those who fled their country of birth during this period were resettled abroad; many others were settled locally or repatriated to their home country. If I impute this probability as a lottery to all refugees, then the model's predicted negative relationship between late-life outcomes and age at arrival is qualitatively restored. Generating a lack of relationship between age at arrival and outcomes relies on all refugees anticipating perfectly their immigration from birth.

A second reason to be skeptical of this extreme case is that it requires implausible expenditures for the necessary early childhood investments. For example, the U.S. Department of Agriculture produces an annual report that estimates the expenditures of U.S. families on their children (Lino, 2010). They also report a number of estimates from other sources. The lowest estimate from any source in the 2010 report is that families spend 21 percent of their budget on a child. To make this calculation conservative, I focus my attention on expenditures on food, health care, child care, and education, which comprise about 40 percent of the total spending on children, or a little more than 8 percent of the family's expenditures. Given the 33:1 income difference between the U.S. and Indochinese countries, the implication is that future refugee families would have to spend about 250 percent of their annual income per child to match U.S. spending levels. It is difficult to see how families could have borrowed or spent this much, particularly in the context of refugee camps.

Finally, a third reason is that evidence presented in Appendix D.1 shows that the patterns for those who arrived before or after 1981 are roughly similar, despite the fact that earlier refugees must have known much less about how and where they could be resettled. In sum, allowing immigration to be anticipated changes the quantitative predictions of the model by flattening the predicted relationship between late-life outcomes and age at arrival. However, for anticipated immigration to explain my findings entirely would require both an implausible amount of foresight on the part of refugee families and extraordinary borrowing and spending power while in their country of birth or refugee camps. Further, it would seem to contradict the experiences of early and late-arriving refugees.

5.2 Measurement Error in Age at Arrival

A second alternative explanation for my empirical findings is that refugees may not accurately report their arrival year, which I use to construct age at arrival. It is well-known that measurement error in the right-hand side variable tends to attenuate the estimated coefficient, so if arrival year is measured with sufficient noise then this could explain my findings. This is potentially important given the existing evidence that year of arrival may not be well-measured (Lubotzky, 2007). However, this evidence is unlikely to play a significant role for my findings. First, the major problem documented in Lubotzky (2007) is that the standard question on year of arrival is not well-designed to capture the experience of immigrants who enter and exit the U.S. multiple times for extended spells. This issue is unlikely to apply to refugees. Second, while refugees might have some recall bias for their year of immigration, the required magnitude of this bias is extraordinarily large for such an important event in their lives.

To quantify this point, I return to the parameterized model of section 3.2, which makes predictions about the log-wages of refugees who arrive at age 0 and age 5. If I regressed log-wage on age at arrival in that parameterized model, I would find a coefficient of -0.096, or a wage decline of roughly 10 percent per year. I then simulate measurement error in that model and run the same regression. I assume that measurement error takes a simple form: a fraction π of each arrival group (age 0 or age 5) misreports that they are in the other group. In this case, $\pi = 0$ means no measurement error, and $\pi = 0.5$ implies that reported age at arrival is pure noise. I find that to cut the estimated coefficient in half - to -0.048 - would require $\pi \approx 0.25$, so that half of the observed signals are noise. It would take $\pi = 0.43$ to generate a coefficient greater than -0.01, which would be difficult to distinguish from zero in the data. I conclude that measurement error is unlikely to explain my findings.

6 Conclusion

This paper used the adult outcomes of refugees who arrived to the U.S. in early childhood to quantify cross-country differences in early childhood human capital. The key fact is that outcomes for such immigrants are unrelated to the age at which they arrived to the U.S. To help understand this finding, I embedded a standard human capital production function into a cross-country model of skill acquisition. I showed that there are three parameterizations of that model that are consistent with the data. The most promising in light of the work here and existing findings in the literature is to assume that there is little role for goods or country environment as inputs to early childhood human capital. Given this, I find that it is unlikely that early childhood human capital explains much of cross-country income differences. It is hard to reconcile large cross-country differences in early childhood human capital with the fact that refugees who immigrated from Vietnam, Cambodia, and Laos to the U.S. had similar outcomes whether they immigrated at age 0 or age 5.

The paper is less conclusive on the interpretation of two other empirical findings. First, there is a strong negative relationship between age at arrival and late-life outcomes for refugees who arrive after roughly age 10. The problem here is that there are many alternative theories that are all consistent with this empirical finding, and that there is no additional evidence to help discriminate between them. Second, the model shows that there are modest differences in average outcomes between ethnic groups that line up well with the average observed characteristics of parents by ethnic group. This evidence suggests that parents are an important input to early childhood human capital formation. However, it is difficult to isolate what particular characteristics of parents or investments by parents matter most using this data. These would seem to be fruitful areas for further research.

References

Almond, Douglas and Janet Currie, "Human Capital Development before Age Five," in David Card and Orley Ashenfelter, eds., *Handbook of Labor Economics*, Vol. 4b, Elsevier, 2011, chapter 15, pp. 1315–1486.

Barry, Michele, Joseph Craft, David Coleman, Holy O. Coulter, and Ralph Horwitz, "Clinical Findings in Southeast Asian Refugees," *Journal of the American Medical Association*, 1983, *249* (23), 3200–3203.

Beckett, Celia, Barbara Maughan, Michael Rutter, Jenny Castle, Emma Colvert, Christine Groothues, Jana Kreppner, Suzanne Stevens, Thomas G. O'Connor, and Edmund J.S. Sonuga-Barke, "Do the Effects of Early Severe Deprivation on Cognition Persist into Early Adolescence? Findings from the English and Romanian Adoptees Study," *Child Development*, 2006, *77* (3), 696–711.

_ , Jennifer Castle, Michael Rutter, and Edmund J. Sonuga-Barke, "Institutional Deprivation, Specific Cognitive Functions, and Scholastic Achievement: English and Romanian Adoptee (ERA) Study Findings," *Monographs of the Society for Research in Child Development*, 2010, *75* (1), 125–142.

Blau, David and Janet Currie, "Preschool, Daycare, and Afterschool Care: Who's Minding the Kids?," in Eric Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Vol. 2, Amsterdam: North Holland, 2006, chapter 20, pp. 1163–1278.

Bleakley, Hoyt and Aimee Chin, "Age at Arrival, English Proficiency, and Social Assimilation Among U.S. Immigrants," *American Economic Journal: Applied Economics*, 2010, *2* (1), 165–192.

Borjas, George J., "The Economic Analysis of Immigration," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Elsevier Science, North-Holland Publishers, 1999, pp. 1697–1760.

Caselli, Francesco, "Accounting for Cross-Country Income Differences," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1A, Elsevier, 2005, chapter 9, pp. 679–741.

Cunha, Flavio and James Heckman, "The Technology of Skill Formation," *American Economic Review*, 2007, *97* (2), 31–47.

— , **James J. Heckman, and Susanne M. Schennach**, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

— , — , **Lance Lochner, and Dimitriy V. Masterov**, "Interpreting the Evidence on Life Cycle Skill Formation," in Eric Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Vol. 2, Amsterdam: North Holland, 2006, chapter 12, pp. 697–812.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall, "Household Choices and Child Development," June 2012. mimeo, Arizona State University.

Dewey, Kathryn G., Johanne Daniels, Kiok S. Teo, Elissa Hassel, and June Otow, "Height and Weight of Southeast Asian Preschool Children in Northern California," *American Journal of Public Health*, 1986, *76* (7), 806–808.

Ebihara, May M., "Khmer," in David W. Haines, ed., *Refugees in the United States. A Reference Handbook*, Westport: Greenwood Press, 1985, pp. 127–147.

Erosa, Andrès, Tatyana Koreshkova, and Diego Restuccia, "How Important is Human Capital? A Quantitative Theory Assessment of World Income Inequality," *Review of Economic Studies*, October 2010, *77* (4), 1421–1449.

Friedberg, Rachel M., "The Labor Market Assimilation of Immigrants in the United States: The Role of Age at Arrival," December 1992. mimeo, Brown University.

Goldenring, John M., Jeffrey Davis, and Margaret McChesney, "Pediatric Screening of Southeast Asian Immigrants," *Clinical Pediatrics*, 1982, *21* (10), 613–616.

Gonzalez, Arturo, "The Education and Wages of Immigrant Children: The Impact of Age at Arrival," *Economics of Education Review*, 2003, *22*, 203–212.

Guryan, Jon, Erik Hurst, and Melissa Kearney, "Parental Education and Parental Time with Children," *Journal of Economic Perspectives*, 2008, *22* (3), 23–46.

Haines, David W., "Introduction," in David W. Haines, ed., *Refugees as Immigrants: Cambodians, Laotians, and Vietnamese in America*, Rowman & Littlefield Publishers, Inc., 1989.

Heston, Alan, Robert Summers, and Bettina Aten, *Penn World Table Version 7.1*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, November 2012.

Hung, Nguyen Manh and David W. Haines, "Vietnamese," in David W. Haines, ed., *Refugees in America in the 1990s: A Reference Handbook*, Greenwood Press, 1996, pp. 305–327.

Lagakos, David, Ben Moll, Tommaso Porzio, and Nancy Qian, "Experience Matters: Human Capital and Development Accounting," 2012. mimeo, Arizona State University.

_ , Benjamin Moll, Tommaso Porzio, Nancy Qian, and Todd Schoellman, "People vs. Places: Understanding the Determinants of Age-Earnings Profiles," March 2013. mimeo, Arizona State University.

Lee, Sharon M. and Barry Edmonston, "Age-at-Arrival's Effects on Asian Immigrants' Socioeconomic Outcomes in Canada and the U.S.," *International Migration Review*, 2011, *45* (3), 527–561.

Lenneberg, Eric H., *Biological Foundations of Language*, New York: Wiley & Sons, 1967.

Lino, Mark, "Expenditures on Children By Families, 2010," 2010. U.S. Department of Agriculture, Center for Nutrition Policy and Promotion. Miscellaneous Publication No. 1528-2010.

Lubotzky, Darren, "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings," *Journal of Political Economy*, 2007, *115* (5), 820–867.

Lustig, Stuart L., Maryam Kia-Keating, Wanda Grant Knight, Paul Geltman, Heidi Ellis, J. David Kinzie, Terence Keane, and Glenn N. Saxe, "Review of Child and Adolescent Refugee Mental Health," *Journal of the American Academy of Child and Adolescent Psychiatry*, 2004, *43* (1), 24–36.

Manuelli, Rodolfo E. and Ananth Seshadri, "Human Capital and the Wealth of Nations," 2010. Mimeo, University of Wisconsin - Madison.

Marshall, Grant N., Terry L. Schell, Marc N. Elliott, S. Megan Berthold, and Chi-Ah Chun, "Mental Health of Cambodian Refugees 2 Decades After Resettlement in the United States," *Journal of the American Medical Association*, 2005, *294* (5), 571–579.

Minnesota Population Center, "Integrated Public Use Microdata Series — International: Version 6. [Machine-readable database].," Minneapolis: University of Minnesota. 2010. http://www.ipums.org.

Mortland, Carol A., "Khmer," in Donald W. Haines, ed., *Refugees in American in the 1990s: A Reference Handbook*, Greenwood Press, 1996, chapter 11, pp. 232–258.

Myers, Dowell, Xin Gao, and Amon Emeka, "The Gradient of Immigrant Age-at-Arrival Effects on Socioeconomic Outcomes in the U.S.," *International Migration Review*, 2009, *43* (1), 205–229.

O'Connor, Thomas G., Michael Rutter, Celia Beckett, Lisa Keaveney, Jana M. Kreppner, and The English and Romanian Adoptees Study Team, "The Effects of Global Severe Privation on Cognitive Competence: Extension and Longitudinal Follow-up," *Child Development*, 2000, *71* (2), 376–390.

Office of Refugee Resettlement, "Report to the Congress: Refugee Resettlement Program," 1980–1995. Washington DC.

Peck, Richard E., Margaret Chuang, Gordon E. Robbins, and Milton Z. Nichaman, "Nutritional Status of Southeast Asian Refugee Children," *American Journal of Public Health*, 1981, *71* (10), 1144–1148.

Robinson, W. Courtland, *Terms of Refuge: The Indochinese Exodus and the International Response*, New York, NY: Zed Books Limited, 1998.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek, "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]," Minneapolis: University of Minnesota 2010.

Rumbaut, Rubén G., "Portraits, Patterns, and Predictors of the Refugee Adaptation Process: Results and Reflections from the IHARP Panel Study," in David W. Haines, ed., *Refugees as Immigrants: Cambodians, Laotians, and Vietnamese in America*, Rowman & Littlefield Publishers, Inc., 1989.

Rumbaut, Ruben G. and John R. Weeks, "Fertility and Adaptation: Indochinese Refugees in the United States," *International Migration Review*, 1986, *20* (2), 428–466.

Rutter, Michael and The English and Romanian Adoptees (ERA) study team, "Developmental Catch-up, and Deficit, Following Adoption after Severe Global Early Privation," *Journal of Child Psychology and Psychiatry*, 1998, *39* (4), 465–476.

Schoellman, Todd, "Education Quality and Development Accounting," *Review of Economic Studies*, 2012, *79* (1), 388–417.

Scott, Joanna C., *Indochina's Refugees: Oral Histories from Laos, Cambodia and Vietnam*, Jefferson, North Carolina: McFarland & Company, Inc., 1989.

Strand, Paul J. and Woodrow Jones Jr., *Indochinese Refugees in America: Problems of Adaptation and Assimilation*, Durham, North Carolina: Duke University Press, 1985.

United Nations High Commissioner for Refugees, "The State of the World's Refugees: Fifty Years of Humanitarian Action," Oxford University Press 2001. New York.

United States General Accounting Office, "Indochinese Refugees: Protection, Care, and Processing Can Be Improved," 1980. Washington DC.

_ , "Improved Overseas Medical Examinations and Treatment Can Reduce Serious Diseases in Indochinese Refugees Entering the United States," 1982. Washington DC.

U.S. Immigration and Naturalization Service, *Statistical Yearbook of the Immigration and Naturalization Service*, Washington, D.C.: U.S. Government Printing Office, 1980–2000.

A Appendix: Selection of Refugees

Refugees are typically considered to be less selected than other forms of migrants, for two reasons. First, in refugee situations there is typically a large "push" factor that leads refugees to leave, mitigating the effects of self-selection. In the Indochinese case, the main elements were political persecution, ethnic discrimination, and in some cases ongoing conflict (such as between the Vietnamese government and the Khmer Rouge). Second, once refugees are formally labeled as such, they face a different set of immigration standards from potential host countries. In essence, they are accepted on humanitarian grounds even if they lack the usual marketable skills that countries desire in their immigrants.

Anecdotal evidence generally tends to point towards modest selection for Indochinese refugees after 1975. For example, Robinson (1998) gives a common characterization of the post-1975 refugees: "Beyond that, this second wave of refugees bore scant resemblance to the relatively homogeneous, well-educated Vietnamese of the first wave. These were peasant Khmer fresh from the 'killing fields' of Cambodia; they were pre-literate Hmong from the highlands of Laos; they were ethnic Chinese and Vietnamese traumatized by perilous boat journeys, push-backs, and pirate attacks." The characterization of the Khmer as rural farmers fleeing the horror of the Khmer Rouge is nearly universal (Ebihara, 1985; Mortland, 1996). On the other hand, two margins of selection are well-known for Indochinese refugees. First, most refugees from Vietnam fled by boat. Doing so required paying a fare and, in some cases, a bribe to Vietnamese officials (Robinson, 1998). For this reason, boat refugees were probably selected based on family income. Second, the Indochinese refugee flows were protracted. While the initial refugees fled persecution, there was widespread agreement by the mid-1980s that many migrants were making a conscious decision to seek resettlement for the sake of improved economic opportunity. This process again likely implies a degree of self-selection.

In this section I review two forms of available evidence to quantify the extent of selection. First, I draw on a number of studies that were conducted in the 1980s by or on behalf of the Office of Refugee Resettlement. Their primary goal was to quantify the challenges faced by refugees and their rate of assimilation, particularly with respect to finding work and leaving public assistance. The Office of Refugee Resettlement conducted an annual, nationally representative survey of newly-arrived refugees. They asked a few basic questions about the refugees' backgrounds before arriving, without distinguishing between subcategories of Indochinese refugees. As a whole, refugees averaged between 4 and 6 years of schooling over this time period, with roughly half of the new arrivals speaking no English and only a few percent reporting speaking English well or fluently. They also asked about the refugees' occupational backgrounds in their home country; roughly one-third of refugees from the years of interest report having been farmers or fishermen, with about one-quarter reporting sales or clerical jobs and the rest distributed among managerial, technical, and blue collar jobs (Office of Refugee Resettlement, 1980–1995).

The studies sponsored by the Office of Refugee Resettlement were typically more detailed in the questions they asked, but focused on a narrow geographic region at the expense of national representativeness. They are particularly likely to over-represent areas where refugees clustered since these were areas where it was cost-effective to sample refugees. Table 2 gives some of the basic descriptive statistics of refugees taken a few years after their arrival. I focus on refugees' education, their ability to write their own language and English, their occupational background, and the characteristics of how they arrived to the U.S. There are two key messages from this table. First, there are dramatic differences in the pre-arrival background of refugees of different ethnic groups. Vietnamese and Chinese refugees in particular were well-educated, wrote their own language well and sometimes wrote English, and had professional occupational backgrounds. The Hmong and Khmer, on the other hand, had very little education, were unlikely to be literate even in their own language, and were almost all farmers, fishermen, or soldiers in their home country. The Lao performed somewhere in the middle. Second, the Hmong in the U.S. in particular come from disadvantaged backgrounds that show little evidence of families with a large degree of acquired human capital. Their experiences are consistent with their background as isolated, rural farmers with no written language until the 1950s.

I add to this evidence by comparing the characteristics of refugee migrants before and after they immigrated to the characteristics of non-migrants from the same country. I focus on education because it is the most useful variable that is easily compared across countries. I utilize data from three different sources. First, I use data from the IHARP study introduced in the text in Section 2.1. That study asked a small sample of refugees about their pre-migration characteristics, including their education level as of 1975. Second, I use data from the 1990 U.S. Census. In this dataset I can observe the post-migration educational outcomes of a large and representative sample of refugees. Finally, I compare these outcomes to the educational attainment of non-migrants using data from population censuses from Cambodia and Vietnam. Unlike Laos, these countries have conducted censuses that collect information on age and education, with the earliest census taking place in 1989 in Vietnam

Table 2: Characteristics of Refugees

	Vietnamese	Chinese	Lao	Hmong	Khmer
Schooling					
Average Years[a]	9.8	6.7	4.9	1.6	5.0
Percent with None[b]	14.0		13.8	41.0	11.8
Percent with Primary – Some High School[b]	45.8		70.1	58.0	84.5
Percent with High School Degree or More[b]	39.8		16.3	1.0	3.6
Literacy					
Native Language (Pre-Migration)[c]	98.8	81.6		26.6	65.9
Native Language (Post-Migration)[b]	93.0		91.3	54.0	77.3
English (Pre-Migration)[c]	33.8	8.8		0.9	4.2
English (Post-Migration)[b]	52.6		42.8	31.0	24.5
Occupation					
Percent Military, Farming, or Fishing[c]	36.1	20.4		90.9	59.9
Migration					
Average Months in Camps[a]	7.8	10.3	23.0	34.3	25.5
Percent Paying Bribes[c]	32.7	71.7		21.3	19.3

[a] Source: Rumbaut and Weeks (1986).
[b] Source: Strand and Jones Jr. (1985). Education responses are for household heads. Ethnic Chinese are included with Vietnamese. Literacy figures are as of the study time and presumably include some learning of English since arrival.
[c] Source: Rumbaut (1989). Refugees were asked about their literacy and occupation as of 1975. Figures for ethnic Lao not reported.

Table 3: **Schooling Comparison: Refugees and Non-Migrants**

	0 Years	1–11 Years	12+ Years
Panel A: Vietnam			
Non-Migrants	22%	69%	9%
Refugees, 1975	6%	63%	31%
Refugees, 1990	12%	25%	63%
Panel B: Cambodia			
Non-Migrants	48%	50%	2%
Refugees, 1975	24%	66%	10%
Refugees, 1990	32%	28%	40%

Table notes: Results represent author's calculation using data from censuses of Vietnam and Cambodia (for non-migrants); from the 1982–83 wave of the Indochinese Health and Adaptation Research Project (for refugees in 1975); and from the 1990 U.S. population census (refugees in 1990). See text for details.

and 1998 in Cambodia.[12]

In each data source I focus on adults who were born before 1957, with the idea that they would ordinarily have completed any schooling by 1975. I impose this cutoff because the retrospective pre-migration educational outcomes are measured as of 1975 in the IHARP. For immigrant groups I also confine my attention to those who arrived to the U.S. after 1975. I code educational attainment into three categories: no schooling; 1–11 years (e.g., less than a high school degree) and 12+ years (e.g., at least a high school degree). Table 3 gives the figures. There are two main results. First, refugees were modestly selected already as of 1975. Refugees from both countries were somewhat less likely to be entirely uneducated and more likely to have at least a high school degree. Second, refugees increased their educational attainment substantially between the end of the Vietnam War in 1975 and the 1990 Census.

I also use the 1990 U.S. Census to provide empirical evidence on the plausibility of the hypothesis that refugees who arrived at different ages were endowed with different family backgrounds in terms of education, ability to find work, willingness to invest, and so on. I test whether there is a correlation between age at arrival for the child and observable family attributes such as family income or parental English-language ability.[13] Differences in outcomes could arise if there was differential selection in age at arrival (despite the historical

[12] Available online at Minnesota Population Center (2010).

[13] I can connect 87 percent to their biological mother and 81 percent to their biological father.

evidence that it was unlikely); or if parents of late-arriving children make systematically different choices in how to allocate their time between investing in their children, investing in themselves, and working in the labor market. It is not necessarily clear which direction of correlation would be more worrying. A positive correlation indicates that parents of late-arriving children have more education, speak English better, or earn more, and can provide more financial resources to their children. On the other hand, a negative correlation could indicate that parents are foregoing the labor market and investments in their own human capital in favor of investing more in their children.

To implement the test I use a host of family attributes: family income; hourly wage of the mother or father; education of the mother or father, measured several ways; and English language ability of the mother or father. For four of the five ethnic groups I find that there is no strong relationship between family characteristics and age at arrival. By this I mean that the correlation is not statistically significant and is as likely to be negative as positive. For ethnic Vietnamese immigrants I find that the estimated coefficients are more likely to be positive than negative, and four of the estimated coefficients (family income, father's hourly wage and years of schooling, and mother's English language ability) are statistically significant at conventional levels. Hence, for four of the five ethnic groups there is not much room for a story of differential selection or differences in investment patterns by age at arrival of the children. For the fifth group I cannot rule such hypotheses out definitively.

B Online Appendix: Identification of Age at Arrival Effects

In the empirical section I propose to estimate the effects of age at arrival by regressing outcomes such as years of schooling or wages on full sets of dummies for age, census year, and age at arrival using a pooled sample of natives and immigrants. Identification of the effect on age at arrival requires some assumptions, which I formulate explicitly and justify here.

To simplify the discussion, I specialize to the case where all time variables enter the regression equations in linear fashion; the same insights apply to the dummy variable specifications I use. With linear time effects, my estimation model for the determination of some outcome of interest y is

$$y = \beta X + \alpha A + \omega Y + \phi AA + \varepsilon,$$

where the right hand side includes a vector of controls X, the age A, the year of the Census Y, and (for immigrants) the age at arrival, AA. Greek letters denote the corresponding coefficients.

Research in the literature often proposes a far more general model of the determination of outcomes (particularly log-wages) for immigrants and natives. Adapting from Friedberg (1992) and Borjas (1999), a flexible model for the determination of native outcomes y^N and immigrant outcomes y^I is:

$$\begin{aligned} y^N &= \beta^N X^N + \alpha^N A^N + \omega^N Y^N + \varepsilon^N, \\ y^I &= \beta^I X^I + \alpha^I A^I + \omega^I Y^I + \phi^I AA^I + \gamma^I C^I + \delta^I YUS^I + \varepsilon^I. \end{aligned}$$

This specification is more general in two ways. First, it allows the effect of the controls, age, and year to be different for immigrants and natives. Second, immigrant outcomes are affected by year-of-immigration cohort effects C^I, which are intended to capture changes in the composition of immigrants by year of entry, and the assimilation term YUS^I which measures the number of years an immigrant has spent in the U.S.

It is well-known that some of the coefficients in this general model are not identified without further assumptions. The problem arises from two linear dependencies in the immigrant equation, namely $YUS^I + C^I = Y^I$ and $AA^I + YUS^I = A^I$. For my purposes the latter dependency is the problem, since it means that the coefficient on age at arrival is not identified without further assumptions. Friedberg (1992) proposes imposing the restriction $\alpha^I = \alpha^N$ to resolve this dependency. In words, the assumption is that immigrants and natives share the same age effects, which can be identified for the natives. The effect of age at arrival on immigrant outcomes is thus identified as the differential effect of a year spent abroad for immigrants as opposed to a year spent in the U.S. for natives, which is exactly consistent with the logic of my model. To implement this strategy I pool natives and immigrants and impose the further restriction $\beta^N = \beta^I$. In this case, a general model for the outcome y is

$$y = \beta X + \alpha A + \omega^N Y^N + \omega^I Y^I + \phi^I AA^I + \gamma^I C^I + \delta^I YUS^I + \varepsilon.$$

There is still a linear dependency in this model, but it is irrelevant for the coefficient of interest, ϕ^I; this can be seen by plugging in for the year effects for immigrants:

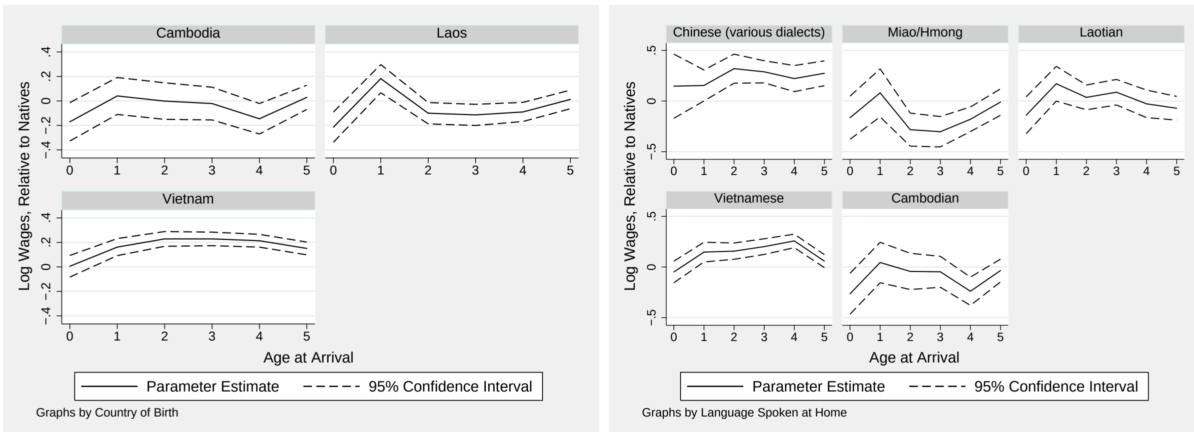
$$y = \beta X + \alpha A + \omega^N Y^N + \phi^I AA^I + (\gamma^I + \omega^I)C^I + (\delta^I + \omega^I)YUS^I + \varepsilon. \quad (7)$$

The effect of age at arrival is identified, although cohort effects and assimilation effects are not. This estimation model is more general than the one used in the text, because it also includes cohort effects as a regressor (even though the estimated coefficient does not measure "true" cohort effects). Nonetheless, implementing this equation produces essentially the same results, which are available upon request.

C Online Appendix: Robustness

C.1 Alternative Decompositions

In this subsection I explore alternative decompositions of the Indochinese refugees into subgroups. The main idea is that self-reported ethnicity may not appropriately capture the different groups of refugees. Also, some Indochinese refugees report ethnicities that do not fall neatly into the five major categories, and so are excluded from earlier figures. As a check on the baseline results, I also decompose refugees by their country of birth, which captures all Indochinese refugees; and by their reported language spoken at home, in case language rather than ethnicity is a better way of grouping immigrants. Figures 7a and 7b show that the patterns for wages are similar to those for the decomposition based on ethnicity, with no trend in outcomes by age at arrival. Figures for schooling are similar and available upon request.



(a) Birth Country

(b) Language Spoken at Home

Figure 7: Log-Wages by Age at Arrival for Alternative Cuts of Indochinese Refugees

As mentioned in the text, U.S. immigration policy towards Indochinese refugees shifted in 1982. A report by the General Accounting Office documented that as of 1981 the required health screenings were cursory (lasting roughly 20 seconds per person); that children under 15 were not routinely screened; and that the results of examinations did not play a part in admissions decisions. The report led to much stricter screening after it was issued. Hence, one might suspect that pre-1982 refugees are less selected on health status, and post-1982 refugees more so. I cut refugees based on whether they arrive before or after the policy change. Figure 8 shows that the lack of a trend is consistent for the less selected, pre-1982 refugees, although the more selected, post-1982 refugees do display a more mixed pattern. For them, although outcomes for those who arrive at age 5 are better than outcomes for those who arrive at age 0, there is a downward trend in outcomes between ages 2 and 5.

Finally, I consider two limits on the sample of interest. First, I exclude from the sample refugees who live in “ethnic enclaves”, areas with high concentrations of other residents of the same ethnicity. I define a person as living in an ethnic enclave if they live in a Public Use Microdata Area (PUMA) where more than 5 percent of the population shares their ethnicity or if they live in a metropolitan statistical area (MSA) where more than 2.5 percent of the population shares their ethnicity. The PUMA is the smallest geographic region publicly available in the Census and includes between 100,000 and 200,000 people,

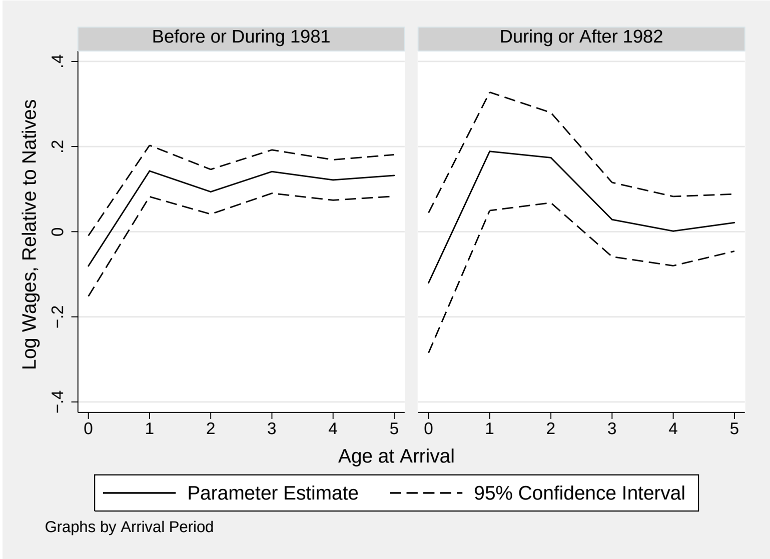


Figure 8: Log-Wage by Age at Arrival for Early and Late Arrivals

corresponding typically to a portion of a city; MSAs are cities and the surrounding areas. My definition of ethnic enclaves excludes roughly 30 percent of refugees from the sample. Figure 9 shows that the wage patterns are similar for those who live outside of enclaves. These findings suggest that my results are not driven by the ability of refugees to live and work in areas with others who share a similar cultural background or language.

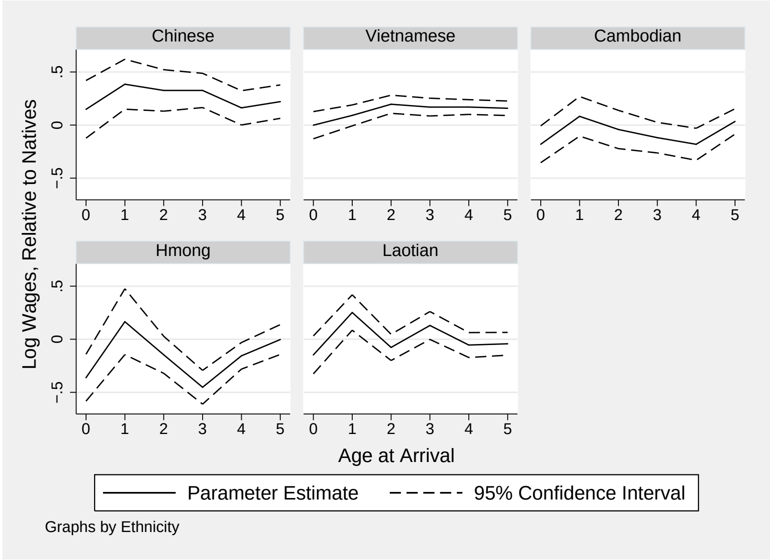


Figure 9: Log-Wage by Age at Arrival for Refugees Living Outside Enclaves

As a second sample restriction, I re-estimate my key regressions using only workers who are 23–26 years of age. My model abstracts from post-graduation human capital accumulation. Although most refugees in my sample are young some are older, and hence may have invested significantly in their human capital since graduation. Figure 10 shows that I obtain similar results for wage patterns if I look exclusively at the very young workers who are unlikely to have made significant post-graduation investments.

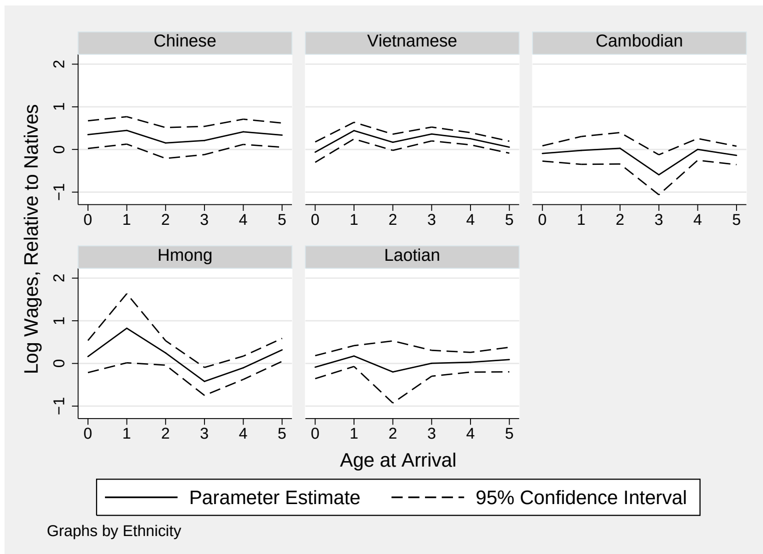


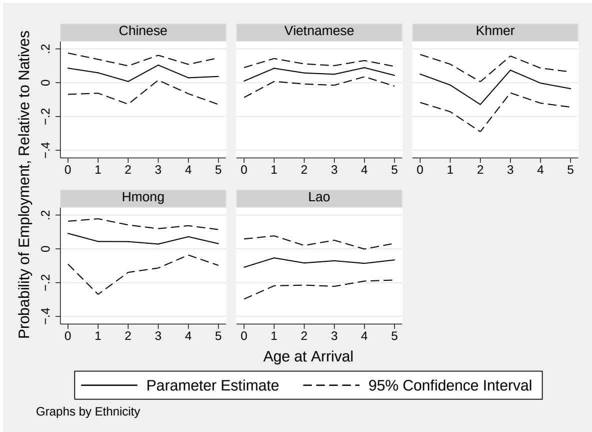
Figure 10: Log-Wage by Age at Arrival for Refugees 23–26 Years Old

C.2 Alternative Outcomes

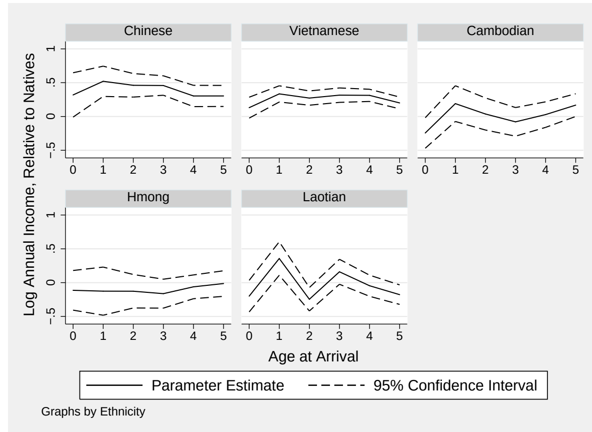
Finally, I consider alternative socioeoconomic outcomes. If I instead estimate the probability of employment, I find a similar pattern: no relationship between age at arrival and probability of employment as an adult, with small level differences between refugees and natives.[14] The same conclusion also applies if I estimate a regression with log-income instead of log-wages (to include differences in hours worked per year) or if I estimate the probability of having graduated college rather than years of schooling; see Figure 11.

I also extend the analysis to look at the outcomes of children born in the U.S. to refugee parents. This test is useful if there are important effects of exposure to adverse conditions

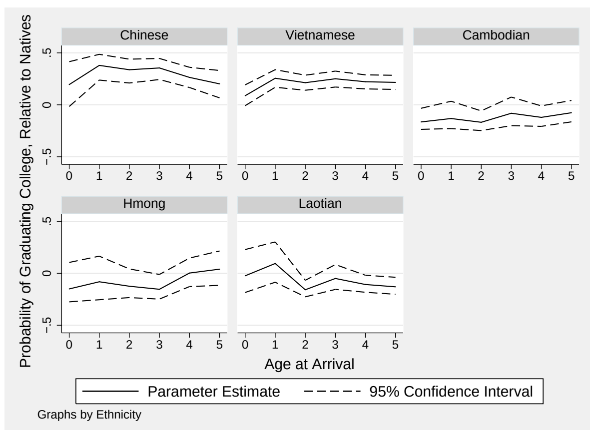
[14]For this and subsequent binary outcomes, estimation is performed via a probit model. The reported coefficient is the model-predicted change in average enrollment for each age-at-arrival group if they had instead been native-born children of non-refugee parents. Standard errors are simulated via Monte Carlo.



(a) Probability of Employment



(b) Log-Income



(c) Probability of Graduating College

Figure 11: Alternative Outcomes by Age at Arrival

while in utero. The test is somewhat more difficult to conduct because I can only link children to their parents while they still live in the same household, so I need an outcome more relevant to the experience of those living at home than completed schooling or wages. I focus my attention on children age 16–18 and use the outcome variable of still being enrolled in school. I include natives, as a control group; child refugees; and children born in the U.S. to Indochinese refugee parents. I identify children as having refugee parents if both parents immigrated from Vietnam, Cambodia, or Laos during the refugee period as defined above and, additionally, the parents were born in the same country and immigrated in the same year. I then regress a school attendance dummy on the same set of controls as in equation (1).

Figure 12 shows the results. Since the sample of 16–18 year olds is smaller, I pool all

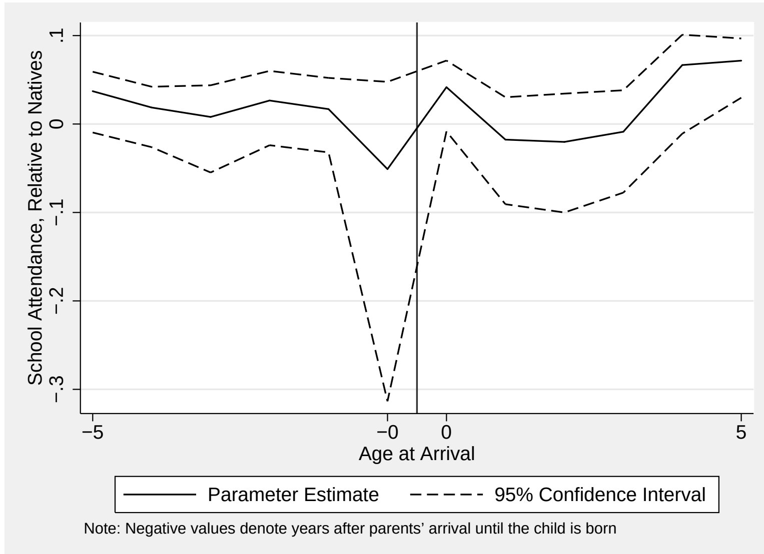


Figure 12: Probability of Attending School at Ages 16–18, by Age at Arrival

Indochinese refugees. I use negative age at arrival for the native-born children of refugee parents; the value denotes how many years after their parents' arrival the children were born. For example, while ϕ_0 is the estimated coefficient for children who are born abroad and immigrate before their first birthday, ϕ_{-0} is the estimated coefficient for children who are born in the U.S. in the first year after their parents immigrated and so on. If in utero development were key, then one would expect to see important differences between children born before their parents' immigration and children born at least one year after their parents' immigration. Instead, there are no significant differences between groups.

D Online Appendix: Derivations for the Isoelastic Model

From the text, the family's problem is:

$$\begin{aligned}
\max_{g_c, g_s, p_c, p_s, S} \quad & \int_{5+S}^{\infty} e^{-rt} A_i(t) h_s dt - g_c - e^{-5r} g_s - A_i h_p p_c - A_i e^{5(\gamma - r)} h_p p_s \\
\text{s.t.} \quad & h_s = \left[\lambda_s h_c^{\frac{\sigma_s - 1}{\sigma_s}} + (1 - \lambda_s) \left(z_i^{\omega_{s1}} g_s^{\omega_{s2}} h_p^{\omega_{s3}} p_s^{\omega_{s4}} S^{\omega_{s5}} \right)^{\frac{\sigma_s - 1}{\sigma_s}} \right]^{\frac{\sigma_s}{\sigma_s - 1}} \\
& h_c = \left[\lambda_c h_0^{\frac{\sigma_c - 1}{\sigma_c}} + (1 - \lambda_c) \left(z_i^{\omega_{c1}} g_c^{\omega_{c2}} h_p^{\omega_{c3}} p_c^{\omega_{c4}} \right)^{\frac{\sigma_c - 1}{\sigma_c}} \right]^{\frac{\sigma_c}{\sigma_c - 1}} .
\end{aligned}$$

Integrating out and substituting in the isoelastic case yields:

$$\max_{g_c,g_s,p_c,p_s,S} \quad \frac{A_i e^{(\gamma-r)(S+5)}}{r-\gamma} h_s - g_c - e^{-5r} g_s - A_i h_p p_c - A_i e^{5(\gamma-r)} h_p p_s \tag{D1}$$

$$\text{s.t.} \quad h_s = h_0^{\lambda_c \lambda_s} \left(z_i^{\omega_{c1}} g_c^{\omega_{c2}} h_p^{\omega_{c3}} p_c^{\omega_{c4}} \right)^{\lambda_s(1-\lambda_c)} \left(z_i^{\omega_{s1}} g_s^{\omega_{s2}} h_p^{\omega_{s3}} p_s^{\omega_{s4}} S^{\omega_{s5}} \right)^{1-\lambda_s} \tag{D2}$$

with $h_c = h_0^{\lambda_c} \left(z_i^{\omega_{c1}} g_c^{\omega_{c2}} h_p^{\omega_{c3}} p_c^{\omega_{c4}} \right)^{1-\lambda_c}$ already substituted out.

The first-order conditions for the problem are:

$$S: \quad A_i e^{(\gamma-r)(S+5)} h_s = \frac{A_i e^{(\gamma-r)(S+5)}}{r-\gamma} \omega_{s5}(1-\lambda_s) \frac{h_s}{S} \tag{D3}$$

$$g_c: \quad \frac{A_i e^{(\gamma-r)(S+5)}}{r-\gamma} \omega_{c2} \lambda_s (1-\lambda_c) \frac{h_s}{g_c} = 1 \tag{D4}$$

$$g_s: \quad \frac{A_i e^{(\gamma-r)(S+5)}}{r-\gamma} \omega_{s2} (1-\lambda_s) \frac{h_s}{g_s} = e^{-5r} \tag{D5}$$

$$p_c: \quad \frac{A_i e^{(\gamma-r)(S+5)}}{r-\gamma} \omega_{c4} \lambda_s (1-\lambda_c) \frac{h_s}{p_c} = A_i h_p \tag{D6}$$

$$p_s: \quad \frac{A_i e^{(\gamma-r)(S+5)}}{r-\gamma} \omega_{s4} (1-\lambda_s) \frac{h_s}{p_s} = A_i e^{5(\gamma-r)} h_p. \tag{D7}$$

Inspection of (D3) reveals that it pins down $S = \frac{\omega_{s5}(1-\lambda_s)}{r-\gamma}$, which implies that S does not vary within or across countries. Equations (D4)–(D7) link together the optimal market goods and parental investments in the two periods. Inspection shows that the model predicts $g_c \propto g_s \propto A_i h_p p_s \propto A_i h_p p_c$, where the proportionality factors are functions of the share parameters (ω and λ) as well as discount and growth rates (e^{-5r} and $e^{5(\gamma-r)}$), and so do not vary within or across countries.

Using proportionality, it is possible to rewrite (D6) as:

$$\kappa_1 h_0^{\lambda_c \lambda_s} z_i^{\omega_{c1}\lambda_s(1-\lambda_c)+\omega_{s1}(1-\lambda_s)} (A_i h_p p_c)^{\omega_{c2}\lambda_s(1-\lambda_c)+\omega_{s2}(1-\lambda_s)} h_p^{\omega_{c3}\lambda_s(1-\lambda_c)+\omega_{s3}(1-\lambda_s)} p_c^{\omega_{c4}\lambda_s(1-\lambda_c)+\omega_{s3}(1-\lambda_s)} = h_p p_c$$

where κ_1 captures functions of parameters and discount and growth rates that do not vary within or across countries. Solve for the time parents spend with their children p_c in terms of exogenous parameters to find:

$$p_c = \kappa_2 h_0^{\frac{\lambda_c \lambda_s}{1-\Psi_2-\Psi_4}} z_i^{\frac{\Psi_1}{1-\Psi_2-\Psi_4}} A_i^{\frac{\Psi_2}{1-\Psi_2-\Psi_4}} h_p^{\frac{\Psi_2+\Psi_3-1}{1-\Psi_2-\Psi_4}} \tag{D8}$$

where κ_2 is again a constant and $\Psi_i \equiv \omega_{ci}\lambda_s(1-\lambda_c) + \omega_{si}(1-\lambda_s)$. Finally, use the proportionality relationship again as well as (D8) to substitute in for h_s to find:

$$h_s = \kappa_3 h_0^{\frac{\lambda_c \lambda_s}{1-\Psi_2-\Psi_4}} z_i^{\frac{\Psi_1}{1-\Psi_2-\Psi_4}} A_i^{\frac{\Psi_2}{1-\Psi_2-\Psi_4}} h_p^{\frac{\Psi_3-\Psi_4}{1-\Psi_2-\Psi_4}} \tag{D9}$$

Likewise, taking equation (D8) and the proportionality relationship and plugging in for h_c yields:

$$h_c = \kappa_4 h_0^{\lambda_c + \frac{\lambda_c \lambda_s (\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}} z_i^{\Phi_1 + \frac{\Psi_1(\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}} A_i^{\Phi_2 + \frac{\Psi_2(\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}} h_p^{\Phi_2+\Phi_3+\frac{(\Psi_2+\Psi_3-1)(\Phi_2+\Phi_4)}{1-\Psi_2-\Psi_4}} \tag{D10}$$

with $\Phi_i \equiv \omega_{ci}(1-\lambda_c)$. The elasticity properties in Table 1 follow directly.

Last, I characterize the problem of the refugee who moves after early childhood. They take their level of h_c as given and choose subsequent investments g_s, p_s, and S. Their problem then is:

$$\max_{g_s, p_s, S} \quad \frac{A_i e^{(\gamma-r)S+5\gamma}}{r-\gamma} h_s - g_s - A_i e^{5\gamma} h_p p_s \tag{D11}$$

$$\text{s.t.} \quad h_s = h_c^{\lambda_s} \left(z_i^{\omega_{s1}} g_s^{\omega_{s2}} h_p^{\omega_{s3}} p_s^{\omega_{s4}} S^{\omega_{s5}} \right)^{1-\lambda_s} \tag{D12}$$

The first-order conditions for the problem are:

$$S: \quad A_i e^{(\gamma-r)S+5\gamma} h_s = \frac{A_i e^{(\gamma-r)S+5g}}{r-\gamma} \omega_{s5}(1-\lambda_s)\frac{h_s}{S} \tag{D13}$$

$$g_s: \quad \frac{A_i e^{(\gamma-r)S+5\gamma}}{r-\gamma} \omega_{s2}(1-\lambda_s)\frac{h_s}{g_s} = 1 \tag{D14}$$

$$p_s: \quad \frac{A_i e^{(\gamma-r)S+5\gamma}}{r-\gamma} \omega_{s4}(1-\lambda_s)\frac{h_s}{p_s} = A_i e^{5\gamma} h_p. \tag{D15}$$

It is still the case that $S = \frac{\omega_{s5}(1-\lambda_s)}{r-\gamma}$. Likewise, it is still the case that there is a proportionality relationship between the remaining two inputs, with $g_s \propto A_i h_p p_s$. Plugging this information into (D15) yields:

$$p_s = \kappa_5 h_c^{\frac{\lambda_s}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}} z_i^{\frac{\omega_{s1}(1-\lambda_s)}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}} A_i^{\frac{\omega_{s2}(1-\lambda_s)}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}} h_p^{\frac{(\omega_{s2}+\omega_{s3})(1-\lambda_s)-1}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}}$$

where κ_5 is a function of share parameters and other constants. Substitution yields an expression for h_s:

$$h_s = \kappa_6 h_c^{\frac{\lambda_s}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}} z_i^{\frac{\omega_{s1}(1-\lambda_s)}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}} A_i^{\frac{\omega_{s2}(1-\lambda_s)}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}} h_p^{\frac{(\omega_{s3}-\omega_{s4})(1-\lambda_s)}{1-(\omega_{s2}+\omega_{s4})(1-\lambda_s)}}$$

where κ_6 is a final constant. It follows that human capital in the labor force is increasing in early childhood human capital for $\lambda_s > 0$.