

Refugees and Early Childhood Human Capital*

Todd Schoellman[†]

October 2012

Abstract

This paper quantifies cross-country differences in early childhood human capital. I embed a standard human capital production function into a cross-country model of human capital investment and labor market outcomes. The model predicts that only some human capital investment channels generate cross-country differences in early childhood human capital. I derive an empirical test of the importance of these channels. The test compares the late-life outcomes of otherwise identical immigrants who entered the U.S. at age 0 or age 5. I implement this test using the Indochinese refugees, who immigrated from poor countries during trying times, and for whom selection is unlikely to bias my results. The empirical results document a striking fact: there is no difference in late-life outcomes between Indochinese refugees who arrived at age 0 or age 5. I conclude that cross-country differences in early childhood human capital are small.

*Thanks to Matt Wiswall and audiences at Ohio State University and the 2012 Midwest Macroeconomics 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).¹ 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.

A key challenge in evaluating the role of early childhood human capital is a lack of comparable cross-country data. 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 data from other countries. Likewise, an existing literature documents the effectiveness of specific types of interventions, but again it is difficult to translate these results into cross-country differences in early childhood human capital (Blau and Currie, 2006; Cunha et al., 2006).

To overcome this challenge I provide new evidence on the late-life outcomes of refugees who immigrated to the U.S. during early childhood. Natives who grow up and work in the U.S. are productive in the labor market, but it is difficult to disentangle whether this productivity is due to human capital formed in early childhood, human capital formed in school, or other factors such as TFP. On the other hand, the refugees I study have spent part of their early childhood in a much poorer country before sharing the education system and labor market conditions of the U.S. with natives. By studying the investments that refugee families make in their children and the late-life outcomes of the children, I can isolate and quantify the effect of spending early childhood in a poor instead of a rich country. Since both natives and refugees are now in the U.S., I can circumvent the difficulty of finding comparable cross-country data.

To motivate the work, I embed a 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. Individuals form human capital in two distinct

¹See Caselli (2005) for a review of the literature that reaches this conclusion.

phases of their lives, early childhood and school. The human capital formed in each phase is affected by the child's initial level of human capital, the exogenous environment of the country, and investments made by the child's family. There are three different types of investment: the child's own time; the time of the parents; and goods, which are purchased in the market. After each individual graduates school they enter the labor force and earn a wage proportional to their final human capital.

The model predicts that two mechanisms can generate cross-country differences in early childhood human capital. First, countries that have exogenously better environments will have higher early childhood human capital. Second, countries with exogenously higher TFP will have higher early childhood human capital. This latter effect is exactly the one stressed in previous work by Manuelli and Seshadri (2010) and Erosa et al. (2010): high TFP lowers the relative price of goods (in terms of wages), which leads families in high-TFP countries to invest more goods in human capital formation. I then show that cross-country differences in early childhood human capital are affected only by the magnitude of these two channels. In particular, I show that cross-country differences are independent of the importance of own or parental time in the human capital production function. An implication of this finding is that the effectiveness of early childhood interventions in the U.S. is insufficient to guarantee that there will be cross-country differences in early childhood human capital. What is needed is evidence on the importance of two particular channels.

I use the model to derive a test that speaks exactly to this point. The empirical moment of interest is the difference in late-life outcomes between otherwise similar immigrants who arrive at the beginning or the end of early childhood (e.g., before their first birthday or at age 5). The model's predictions are clear. If either or both of the above mechanisms operates, then older-aged arrivals will have less human capital than otherwise similar immigrants who arrived at younger ages, because the older-aged arrivals had more exposure to the poor country. Although parents have the chance to remediate these gaps with investments made in the U.S., they will not find it optimal to remediate the gaps entirely under fairly general conditions on the human capital production function. In this case the adult human capital and wages of immigrants will be declining in their age at arrival. This is the prediction that I propose to test. The prediction offers two useful features. First, it has a natural economic interpretation: it captures the marginal cost of spending an additional year in a poor country instead of the U.S. Second, it is less likely to be biased by selection, because I am comparing immigrants who arrived to the U.S. at different ages, rather than comparing natives to immigrants. Hence, my key identifying assumption is that immigrants who arrive at different ages in early childhood are not selected differently.

I provide results from this empirical test for a number of immigrant groups, but I focus mostly on the Indochinese refugees, that is, immigrants 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 for a refugee camp; the transit period and wait times in camps before resettlement were variable and could be quite long. For this reason they are unlikely to be differentially selected on the age at arrival of their child, even if there were an underlying motive to do so. Additionally, Indochinese refugees 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., 2011). Refugees also faced an environment that should have been 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 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 hold for other outcomes (such as completed schooling) and other groups of immigrants. Finally, I 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 model can be consistent with this finding under three different special cases. I then consider additional evidence from immigrants to the U.S. as well as from the large existing literature on early childhood human capital formation to discriminate among the three. Only one special case is consistent with the findings from refugees as well as the structural estimates of the human capital production function, the evidence from early childhood interventions, and evidence from children who are exposed to extreme deprivation. The case of interest requires that market goods and country environment are quantitatively unimportant in early childhood. This finding is consistent with and supports the recent results of Del Boca et al. (2012), who find that the importance of goods relative to parental time rises over the life of the child. A key implication is that cross-country differences in early childhood human capital are likely to be small, implying that abstracting from this dimension is innocuous for development accounting exercises.

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), Erosa et al. (2010), and Lagakos et al. (2012) 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 describes the model and formulates its predictions for natives and for immigrants to the U.S. Section 3 presents the empirical results. Section 4 combines the insights from this paper’s empirical work and the existing literature to discipline the model and its predictions for cross-country early childhood human capital differences. Section 5 considers alternative hypotheses for my empirical findings. Section 6 concludes.

2 A Model of Early Childhood Human Capital

2.1 Human Capital Accumulation and the Labor Market

The model describes the human capital accumulation and labor market decisions of families in countries $i \in \{1, 2, \dots, I\}$. 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, the parent has exogenously given human capital h_p . The two are drawn from a joint distribution $F_i(h_0, h_p)$ defined on $(0, \infty)^2$. I assume that the unconditional distribution of h_0 is the same across countries. I allow the unconditional distribution of parental human capital to differ across countries, reflecting prior investment decisions not modeled here.

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:

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

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 during early childhood depends on three factors. First, it depends on the exogenous environment of the country z_i , which captures for example the prevalence of diseases or the quality of medical infrastructure. Second, it depends on investments using market-purchased goods m_c , which includes books and vaccines. Finally, it depends on the inputs provided by the parent, for example by caring for or reading to the child. The total parental input depends on the time parents spend with their children p_c but also on the quality of that time, which I assume is measured by the parent's human capital h_p . Given these three inputs, x_c is determined using a power function:

$$x_c = z_i^{\omega_{c1}} m_c^{\omega_{c2}} (h_p p_c)^{\omega_{c3}}. \quad (2)$$

The ω_c are share 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 :

$$h_s = \left[\lambda_s h_c^{\frac{\sigma_s}{\sigma_s-1}} + (1 - \lambda_s) x_s^{\frac{\sigma_s}{\sigma_s-1}} \right]^{\frac{\sigma_s}{\sigma_s-1}}. \quad (3)$$

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

The composite investment during schooling depends on four factors. The first three are similar to early childhood: country environment z_i ; market-purchased goods m_s ; and

parental inputs $h_p p_s$. Finally, the human capital also depends on how long children stay in school, S . The terms are again combined using a power function,

$$x_s = z_i^{\omega_{s1}} m_s^{\omega_{s2}} (h_p p_s)^{\omega_{s3}} S^{\omega_{s4}}, \quad (4)$$

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

The human capital production function in equations (1) – (4) borrows heavily from the microeconomic literature on early childhood human capital. Two key insights from that literature will be useful for my empirical approach. 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.² 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 market goods 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; in the empirical implementation I study young workers. 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 g 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 individuals 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

²On the other hand, Cunha et al. (2010) have only one form of investment good and so do not estimate an intratemporal elasticity of substitution between different types of investment. I have followed Del Boca et al. (2012) and implicitly restricted the intratemporal elasticity of substitution to be one. While this elasticity is important for a number of questions of interest, it plays only a minor role here.

value of lifetime earnings net of the cost of market goods and the foregone earnings of the parents:

$$\max_{m_c, m_s, p_c, p_s, S} \int_{5+S}^{\infty} e^{-rt} A_i(t) h_s dt - m_c - e^{-5r} m_s - A_i h_p p_c - A_i e^{5(g-r)} h_p p_s \quad (5)$$

where h_s is derived from equations (1) – (4) 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.

2.2 Cross-Country Human Capital Differences

In this section I explore the qualitative properties of the model under the special case $\sigma_c = \sigma_s = 1$. In this case the model provides simple closed-form solutions for the investment and human capital accumulated as a function of the exogenous variables h_0 , h_p , z_i , and A_i . These solutions provide intuition for the key mechanisms of the model. Further, in this special case the model nests a two-stage Ben-Porath (1967) human capital production function. This feature is useful because much of the existing cross-country human capital literature has utilized the Ben-Porath setup, so I can easily compare my findings to theirs (Manuelli and Seshadri, 2010; Erosa et al., 2010). I return to the more general case below.

I focus here on the properties of the model; the derivations are reserved to Appendix B. This special case is much easier to solve, but does have one significant drawback, formalized in Proposition 1:

Proposition 1 *The optimally chosen duration of schooling S is independent of the exogenous variables.*

Hence the model does not generate a distribution of schooling within or across countries. I explore this result and how to rid the model of it in Section 2.4.

The parental and market-purchased goods inputs at each stage do scale with exogenous factors. An advantage of focusing on the isoelastic case is that there are strong equilibrium relationships among the different inputs that make it straightforward to derive the elasticity of investments and human capital with respect to the exogenous driving forces, given in Proposition 2.

Proposition 2 *The elasticity of h_s with respect to h_0 is $\frac{\lambda_c \lambda_s}{1-\Psi}$; with respect to z_i is $\frac{\omega_{c1} \lambda_s (1-\lambda_c) + \omega_{s1} (1-\lambda_s)}{1-\Psi}$; and with respect to A_i is $\frac{\omega_{c2} \lambda_s (1-\lambda_c) + \omega_{s2} (1-\lambda_s)}{1-\Psi}$, where $\Psi \equiv (\omega_{c2} + \omega_{c3}) \lambda_s (1-\lambda_c) + (\omega_{s2} + \omega_{s3}) (1-\lambda_s)$. The elasticity with respect to h_p is 0.*

Ψ is the total returns to scale in market-purchased goods and parental inputs. $\Psi < 1$ (diminishing returns to accumulable factors) is necessary for an interior solution to exist, and is assumed for the remainder of the paper. Under this condition, h_s is generally increasing in h_0 , z_i , and A_i , unless some of the weights (ω or λ) are set to 0. On the other hand, h_s does not depend on parental human capital h_p . This result follows from the assumption that the parents' human capital is equally productive when working in the market or investing in their children. Under this neutrality assumption, parents choose an optimal level of $h_p p$; parents with higher human capital simply spend less time with their children. To my knowledge there is no existing research on the relative productivity of parental human capital at home and in the market. My empirical findings support the neutrality assumption. Finally, I note that qualitatively similar but algebraically more complex expressions characterize the elasticity of h_c with respect to exogenous factors; I reserve the exact expressions to the Appendix.

The model allows for differences in human capital within and across countries. The distribution of human capital within a country is generated by the heterogeneity in human capital at birth h_0 . Children who have more human capital at birth enjoy an innate advantage, which is amplified by the endogenous decisions of their families to increase investment. However, h_0 does not contribute to cross-country human capital differences because I have assumed that the unconditional distribution of h_0 is the same across countries.

The remaining two exogenous factors can generate cross-country differences in human capital. The effect of z_i is straightforward: countries that provide better environments for children have higher average human capital, with the effect again amplified by families' endogenous decisions to invest more in children in countries with better environments. 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.

The same logic does not apply to investments of parents' time. The reason is that investments of parental time trade foregone labor earnings for the parent today against higher future labor earnings for the child. Higher TFP raises the value of current and future time proportionately, and so has offsetting changes in the marginal costs and marginal benefits of parental time inputs. This logic explains why the strength of the TFP multiplier

depends critically on ω_{c2} and ω_{s2} : these are the weights on goods in the human capital production function.

An existing literature has established that investment in early childhood human capital can affect late-life outcomes. The preceding analysis shows that this fact alone is insufficient to imply cross-country differences in early childhood human capital. The critical question is what form effective investments take: are they goods purchased in markets or time spent by parents with their children? To my knowledge only the work of Del Boca et al. (2012) is designed to distinguish between these hypotheses. One contribution of this paper is to provide additional empirical evidence to this debate about which channel matters. Further, new evidence is still needed on the importance of country environment effects before age 5; this question is unlikely to be settled by results drawn from a single country. I now show how empirical evidence from immigrants to the U.S. can be used to address these questions.

2.3 Immigrants and Identification

One goal of this paper is to isolate and quantify the contribution of early childhood human capital to cross-country income differences. The experience of natives is generally difficult to use for this purpose. For example, U.S. natives are born into a high-TFP country. The model predicts that this will lead them to accumulate high average early childhood human capital because the market goods input will be relatively cheap. But the same effect also applies while they are in school; and further, TFP directly raises income once they enter the labor force. This collinearity problem makes it challenging to use the experience of natives to isolate the importance of early childhood human capital. Similarly, country environment affects both early childhood and school human capital formation.

In this paper I introduce novel empirical evidence that draws on the experience of immigrants who arrive to the U.S. from poor countries during early childhood. Intuitively, these immigrants solve the collinearity problem presented by natives. They spend part or all of their early childhood in a foreign country with a lower TFP and a worse environment. Afterwards, they immigrate and then share the environment, TFP, and labor market with natives. The goal of this section is to show that their post-migration investment decisions and labor market outcomes can be used to infer the quantitative magnitude of their early childhood human capital deficit.

To derive the model's predictions about the behavior and outcomes of immigrants, I have to take a stand on immigrants' beliefs about the possibility of immigration and also on who immigrates. It is useful to start with two strong simplifying assumptions and then consider the effects of relaxing them. First, I assume 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.; aa denotes age at arrival. At this point I allow the family to re-optimize all future investments, taking past investments as given. Second, I assume that immigrants are randomly sampled from their country of origin with respect to the initial heterogeneity h_0 .

From the perspective of the model, there are three ages at arrival of interest: $aa = 0, 5,$ and $5 + S$. If immigrants are randomly sampled and arrive to the U.S. at age 0, the model predicts that they will have the same late-life outcomes as natives, because they share the same environment as natives for their entire life and are still able to choose all of their investments. On the other hand, immigrants who arrive at ages 5 and $5 + S$ have already made investments before immigrating. Proposition 2 tells us that they will have invested less and have lower human capital at arrival than natives under fairly general conditions. Those who arrive at age 5 will re-optimize and choose higher investments while in school than they would have had they remained in their country of birth. However, Proposition 3 says that their lower early childhood human capital at arrival leads them to have lower human capital at graduation as well.

Proposition 3 h_s is increasing in h_c if $\lambda_s > 0$.

Immigrants who arrive at age $5 + S$ have fixed all their human capital investments at lower levels. Hence, they have even less human capital as adults than immigrants who arrive at age 5.

The model predicts that an immigrant's human capital as an adult is declining in their age at arrival to the U.S. under fairly general conditions (the relevant weight terms need to be positive; see Propositions 2 and 3). Under the assumption that labor markets in the U.S. are competitive, this generates the testable prediction that the adult wages of immigrants are declining in their age at arrival. The strength of the relationship depends on the role of country environment and market-purchased goods in production ($\omega_{c1}, \omega_{s1}, \omega_{c2},$ and ω_{s2}) and on the role of prior human capital at each stage (λ_c and λ_s). It is useful to have an idea of the quantitative magnitude of the relationship that is implied by typical calibrations in the literature. To this end, I study a conservative parameterization of a two-stage Ben-Porath

human capital production function:

$$h_c = m_c^{0.25},$$

$$h_s = m_s^{0.25}(Sh_c)^{0.4}.$$

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.³

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 1 expresses the quantitative predictions of this parameterization in the standard format that I will utilize for the remainder of the paper. The y-axis gives the log-wage that an immigrant will earn relative to a native, so that -0.1 indicates a log-wage 0.1 lower than a native (or a wage roughly 10 percent lower). I plot this wage gap against the age at arrival along the x-axis. Under the assumption that immigrants are randomly selected, a fairly conservative parameterization of the model predicts that children who enter after early childhood will earn 40 percent less than natives, and those who immigrate after graduation will earn 70 percent less.

However, immigrant-native wage gaps are not the ideal statistic to take to the data. The reason is that immigrants to the U.S. are likely selected. In this case the empirically observed gap in wages between immigrants and natives represents the human capital gap that would obtain given randomly selected immigrants, plus a difficult to quantify bias that arises from the non-random selection. As an alternative, I propose to use the slope of the relationship between outcomes and age at arrival in Figure 1. The slope has a natural economic interpretation: it measures the wage loss that arises from spending an additional

³The parameters required are $\lambda_c = 0$, $\lambda_s = 1$, $\omega_{c2} = 0.25$, $\omega_{s2} = 0.417$, and $\omega_{s4} = 0.667$, 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.

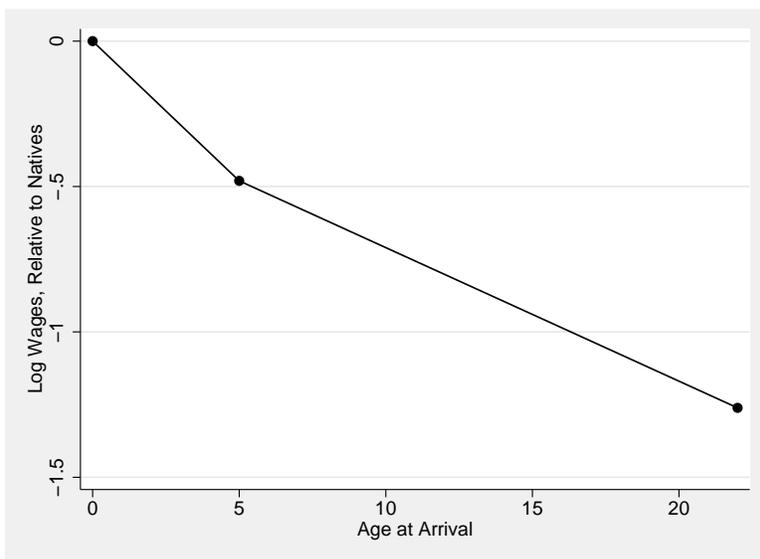


Figure 1: Model-Predicted Log-Wage Patterns for Immigrants

year in a poor country instead of the U.S. In the current calibration that slope is -0.096 during early childhood, implying a loss of 9.6 percent of wages per year of early childhood spent in the poor country. The slope is also less likely to be affected by selection. By construction the slope compares otherwise identical immigrants who differ only in their age at arrival to the U.S. For selection to bias my results, it would have to be the case that immigrants who arrive at different ages are systematically selected differently. The absence of differential selection thus represents my key identifying assumption.

My empirical approach is sufficiently simple that I produce estimates of the relationship between log-wage and age at arrival for a number of different groups of immigrants. The slopes can be interpreted as the log-wage cost of an additional year in poor countries as long as my identifying assumption holds. However, this identifying assumption is difficult to test. I focus my attention on Indochinese refugees because historical circumstances surrounding their migration make it extremely unlikely that they were differentially selected.

2.4 Properties of the Generalized Model

The analysis so far has been carried out under the maintained assumption that $\sigma_c = \sigma_s = 1$. Now I consider the properties of the more general model with unconstrained elasticities of substitution. As noted before, the primary function of these elasticities is to govern the extent to which low levels of h_0 and h_c can be remediated through subsequent investments x_c and x_s . The value of σ_c is relatively unimportant for my results because I assume that

the unconditional distribution of h_0 is the same in all countries. On the other hand, σ_s plays an important role.

Empirically, I propose to estimate the relationship between age at arrival and wages for immigrants from poor countries. This slope is determined by a combination of two factors. First, it is affected by the human capital gap when the immigrant arrives to the U.S., e.g., how far they are behind immigrants who arrived at younger ages. This gap is in turn a function of cross-country differences in TFP and country environment and the relevant parameters. Second, it is affected by the extent to which the human capital gap at arrival can be remediated through post-migration investments, which is exactly the role of σ_s . To show the importance of σ_s , I return to the quantitative predictions of the previous section. I parameterize the model as before, but deviate by allowing σ_s to vary. For each value of σ_s 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.

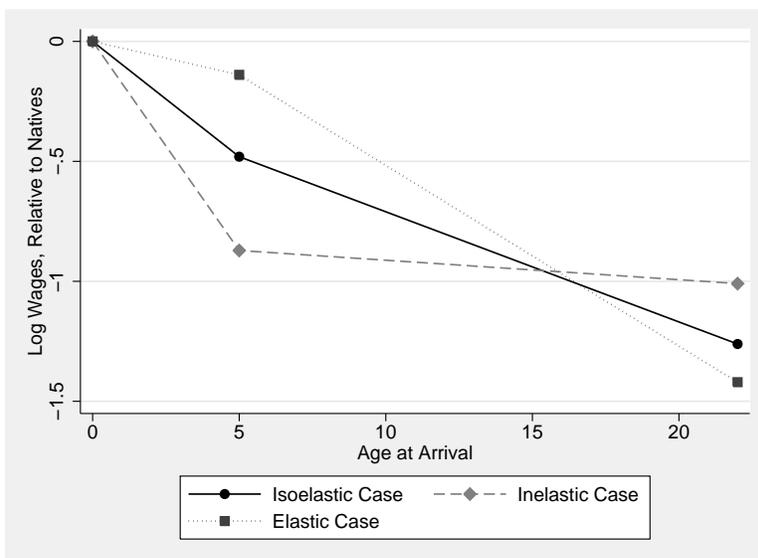


Figure 2: Model-Predicted Log-Wage Patterns for Immigrants

Figure 2 gives the results in the same format as before. The isoelastic case fixes $\sigma_s = 1$ as in the baseline. 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 the elastic case sets $\sigma_s = 1.5$, in which case outcomes are only weakly declining in age at arrival for those who arrive by age 5. For values of $\sigma_s > 2$,

the quantitative difference between age 5 and age 0 arrivals is economically small (less than 5 percent of wages) and would be difficult to detect empirically. I note again that such high substitutability and easy remediation of low early childhood human capital is inconsistent with the evidence from the literature; I return to this point in Section 4.

An additional reason to consider $\sigma_s < 1$ is that in this case the model produces a non-degenerate distribution of S within and across countries, with S weakly increasing in z_i , z_j , and A_i . Further, if h_c is taken as given then S is weakly increasing in h_c . Thus, under the inelastic case, the model also predicts that school attainment is decreasing in age at arrival for immigrants. This prediction is useful because unlike other human capital investments, the years of completed schooling are easily observable.

In summary, the model suggests two empirical moments that are useful for measuring cross-country differences in early childhood human capital. The first is the correlation between adult wages and age at arrival; the second is the correlation between completed years of schooling and age at arrival. These two slope moments allow me to gauge the disadvantage of a year spent abroad instead of in the U.S. and to parameterize the model. In the next section I estimate these relationships and show that the slope is essentially zero for each.

3 Empirical Implementation

The model of the previous section allows for cross-country differences in early childhood human capital driven by underlying differences in TFP or country environment. The key idea of this paper is to isolate and quantify these forces by using the experience of young children who migrate from poor countries to the U.S. I focus on the U.S. so that I can exploit the large quantity of data on immigrants in the U.S. population censuses, both in terms of sample size and also in terms of available information. While I will present results for immigrants from many poor countries I focus my attention on results for Indochinese refugees. The Indochinese refugees offer three key benefits: they are a large subsample of immigrants; they immigrated from poor countries during trying times, and were demonstrably disadvantaged at arrival; and the historical circumstances of their immigration makes it unlikely that they were differentially selected, which is key to my identification strategy. I now give a brief overview of the Indochinese refugees and explain these benefits in more detail.

3.1 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 all, roughly three million citizens escaped to neighboring countries by land or by boat (United Nations High Commissioner for Refugees, 2001). 1.4 million of those who fled spent time in refugee camps, and 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; it is these refugees who I use for my empirical analysis.⁴

It is useful to distinguish two waves of refugees. 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 and had relatively short transit periods before resettlement in the U.S. I exclude these immigrants from my analysis in order to focus on those who were less selected and who had more typical refugee experiences.

Subsequent refugees generally faced significant hardship before resettlement. In their home country they were exposed to politically and ethnically motivated persecution, violence, and economic hardship (Robinson, 1998). For example, the rise of the Khmer Rouge in Cambodia meant a switch to an extreme, agriculturally oriented form of Marxism that brought widespread famine, while at the same time the Khmer Rouge killed an estimated 20 percent of the Cambodian population directly (Robinson, 1998). Those who decided to flee faced a difficult and unpredictable transit period 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). Because of unpredictable and changing policies, refugees spent anywhere from a few months to years in camps; indeed, some were ultimately repatriated to their home country rather than resettled abroad.

I focus my work on the Indochinese refugees for three primary reasons. First, the Indochinese refugees were one of the largest refugee flows to enter the U.S., second only to the Cuban refugees. Further, the Indochinese refugees were young, with a median age less

⁴Other countries took many fewer refugees; only France, Australia, or Canada resettled even 25,000. 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).

than 20 (Haines, 2010). These two facts together imply that I can construct a large sample of Indochinese refugees who entered the U.S. as young children.

The second reason I focus on the Indochinese refugees is that there exists a general perception that refugees are less selected than other immigrants and that the Indochinese refugees are less selected than other refugees. Minimizing selection is useful for two reasons. First, my goal is to study children who experienced dramatic changes in their family income and environment. 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, roughly one-quarter of refugee parents reported no education at all, and roughly one-half of refugees spoke no English at the time of arrival. The second reason that minimizing selection is useful is that my identification strategy relies on the assumption that immigrants are not differentially selected in age at arrival. Two historical features of Indochinese refugees make this form of selection particularly unlikely, even if there were an underlying reason why families might want to select in this manner. First, Indochinese refugees had only loose control over when they would be resettled. They could choose when to flee their home country, but they faced substantial variation in both transit time to the refugee camp and how long it took them to be resettled out of the refugee camp. These delays were largely determined by the policies of the countries of refuge and resettlement and were outside the control of the refugee families. Second, the typical Indochinese family was large. Using the 1990 U.S. Population Census, I document that the average Indochinese family with any children had three children who were born abroad. In this case, it is not even clear what it means for the family to be selected based on the age of "the" child.

The final reason I focus on the Indochinese refugees is that there is ample evidence that they 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., 2011). Given their background, one might expect them also to have low z_i from exposure to war, violence, food shortages, and other forms of hardship. Doctors and psychologists in the U.S. who cared for refugee children conducted assessments that quantified their disadvantaged status. 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). Indochinese children suffered from tuberculosis, hepatitis B, malaria, and intestinal parasites at rates more than an order of magnitude higher than the general U.S. population (United States General Accounting Office, 1982; Barry et al., 1983; Goldenring et al., 1982).⁵ Likewise, psychological studies indicated the presence of post-traumatic stress disorder, depression, and other mental illnesses, with some persistence in follow-up studies six years later (Lustig et al., 2004; Sack et al., 1993).

In short, the ideal model experiment would be to move unexpectedly a large number of randomly chosen young children from a poor country to the U.S., allow them to re-optimize their behavior, and study their choices and outcomes. The Indochinese refugees fit this description well.

3.2 Benchmark Results

To estimate outcomes for the child refugees, I combine data from the 2000 Population Census and the 2005–2010 American Community Surveys (ACSs), available online through Ruggles et al. (2010). When combined, these datasets provide a large sample with fairly consistent variables 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.⁶ 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 estimate and present results separately for each of the five main ethnic groups of Indochinese refugees (Chinese, Vietnamese, Cambodian, Hmong, and Laotian) so that composition effects do not drive my results. Results from alternative decompositions are similar and are reserved to Appendix D.1. I also include natives in the sample.

⁵All refugees were supposed to be screened for the presence of contagious diseases and certain other mental or physical conditions before admission to the U.S. However, a report by the General Accounting Office documented that as of 1981 the required examinations 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 stricter screening after it was issued. I explore this change in policy further in Appendix D.1.

⁶For Vietnam, the years are 1976–1990; for Cambodia, 1976–1994; and for Laos, 1976–1996. These cutoff dates were chosen to include the years of highest refugee flows, and to exclude later years when refugees began to make up less than 50 percent of immigrants.

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.⁷ I then regress

$$\log(W) = \beta X + \sum_a \alpha_a d_a + \sum_y \omega_y d_y + \sum_{aa} \phi_{aa} d_{aa} + \varepsilon, \quad (6)$$

where W is the constructed wage and X is a vector of control variables that consists of state of residence and gender. 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. I show in Appendix C that the coefficients on age at arrival are identified, even in a more general model that allows for cohort or assimilation effects.

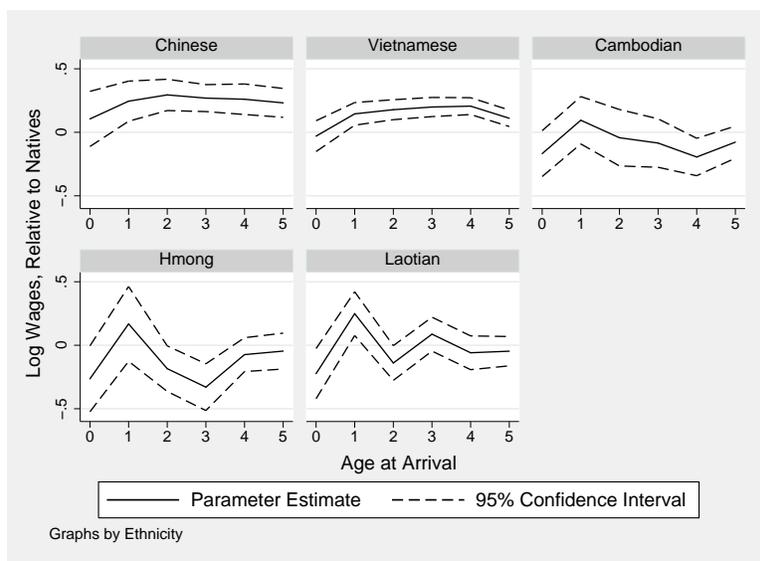


Figure 3: Log-Wages by Age at Arrival

Figure 3 plots the estimated coefficients on age at arrival ϕ_{aa} as a function of age at arrival, as well as the 95 percent confidence interval. Given the regression equation

⁷From 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 2008–2010 ACSs.

(6), these coefficients can be interpreted as the marginal difference in log-wages between an immigrant who arrived at age aa and a native with the same age, gender, and state of residence. In other words, it is exactly the empirical analogue to the model-generated figure 1. However, the empirical findings contrast sharply with the model-generated results. The critical empirical finding is that the slope of the line is consistently estimated to be zero: there is no trend relationship between age at arrival and wages as an adult for Indochinese refugees.

As a second outcome I investigate the years of completed schooling. For this outcome, I limit the sample 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 (6). Figure 4 again plots the coefficients ϕ_{aa} against the age at arrival aa . The same basic finding obtains: there is no trend relationship between age at arrival and completed schooling for Indochinese refugees.

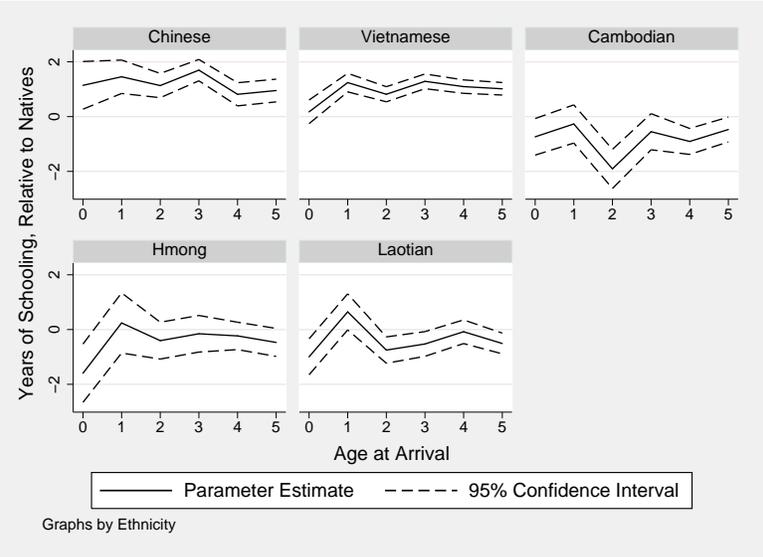


Figure 4: Years of Schooling by Age at Arrival

These are the benchmark empirical results of the paper. In Appendix D.2 I show that similar results hold for a number of other socioeconomic outcomes. A similar picture emerges if we look at probability of being employed or incarcerated: small to no level differences as compared to natives, and no trend in age at arrival. The same patterns apply if I use earnings instead of wages or if I use probability of graduating college instead of years of schooling. I also study the outcomes for refugees who live outside areas of high

ethnic concentration defined in several ways and find similar patterns.

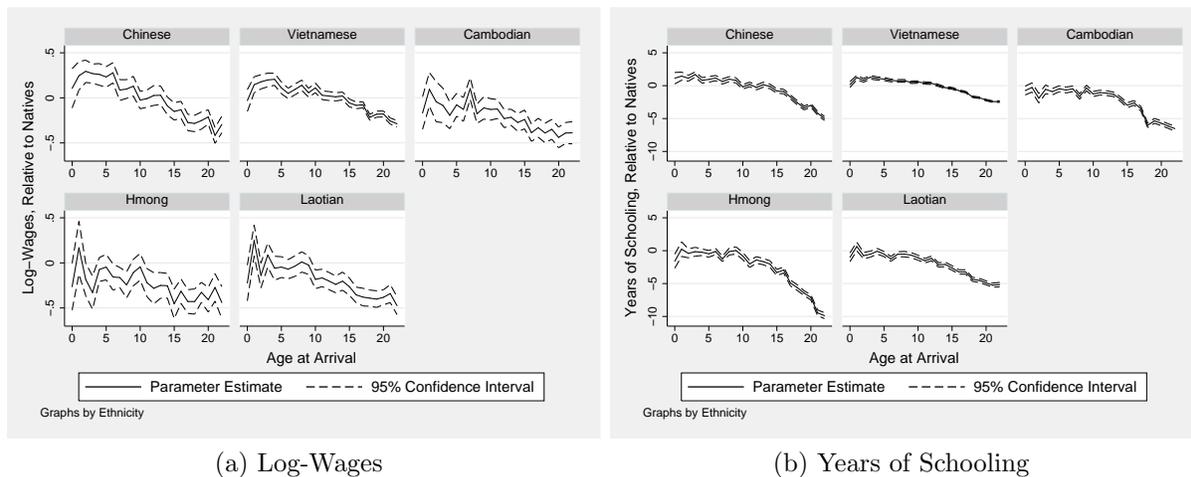


Figure 5: Outcomes by Age at Arrival for Older Ages

Finally, I want to emphasize that the results from early childhood are special. Figure 5 shows the patterns of log-wages and years of schooling by age at arrival for those arriving between ages 0 and 22. It shows that a pronounced negative trend in outcomes begins somewhere between age 5 and 10, with the exact turning point varying somewhat by ethnicity and outcome. For those who arrive at age 22 the effects are quantitatively large: they have 5–10 fewer years of schooling and 30–60 percent lower wages than refugees who arrive in early childhood. Hence, the lack of a pattern for refugees who arrive during early childhood is different from the pattern shown for those who arrive afterwards.

3.3 Other Immigrant Groups

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 they were differentially selected. 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.⁸ 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.⁹ 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.

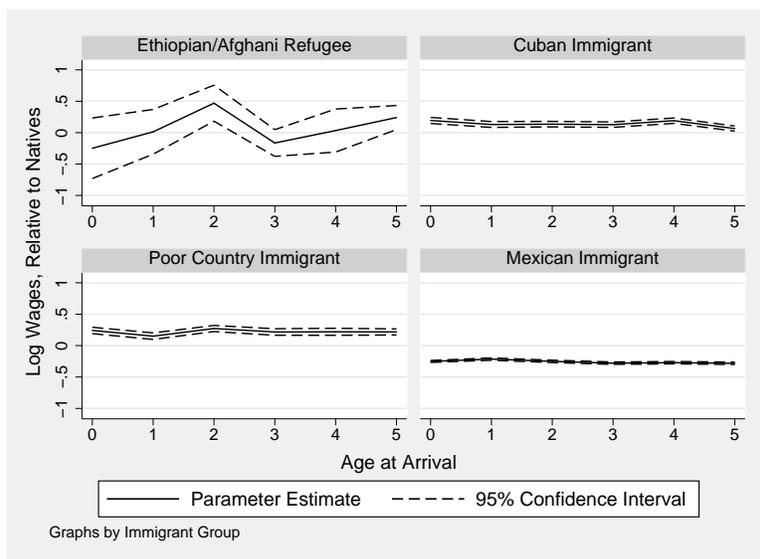


Figure 6: Log-Wages by Age at Arrival for Other Immigrant Groups

Figure 6 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 schooling and are available upon request.

⁸I 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).

⁹The countries distinguished in the Census and so included in my sample are: Haiti, Bangladesh, Nepal, Ghana, Guinea, Liberia, Senegal, Sierra Leone, Kenya, Somalia, Tanzania, Uganda, Zimbabwe, and Eritrea.

4 Implications for Model and Cross-Country Income Differences

The previous section established a robust empirical fact: late-life outcomes of child refugees are uncorrelated with age at arrival for those who arrive by age 5. This fact contradicts the generic prediction of the model from Section 2. It may also appear to contradict the conventional wisdom that “early childhood human capital matters”. In this section I show how to reconcile the model with both my empirical findings and those of the literature. I also show that the model which is capable of fitting both my findings and those of the literature makes a sharp prediction: early childhood human capital differs little across countries.

4.1 Reconciling the Model to the Data

Generically, the model predicts that wages are declining in age at arrival (Propositions 2 and 3). The next natural question is whether the model can generate the flat relationship between wages and age at arrival. It can for three different special cases of parameters:

- Case 1: $\omega_{c1} = \omega_{c2} = 0$. In this case country environment and market goods are unimportant for the production of early childhood human capital. Since these are the channels that generate cross-country differences, the implication is that there are no cross-country differences in early childhood human capital.
- Case 2: $\lambda_s = 0$. In this case there may be cross-country differences in early childhood human capital but they are irrelevant for late-life outcomes.
- Case 3: $\sigma_s = \infty$. In this case poor countries have a deficit in early childhood human capital. However, these deficits can easily be remediated by young immigrants through investments made while they are in schooling. In practice, $\sigma_s > 2$ generates levels of remediation large enough to generate nearly flat profiles of wages in age at arrival.

It is important to differentiate between these three cases because they have very different implications for human capital production functions and also for cross-country differences in early childhood human capital. The experience of refugees provides one additional piece of information, which is that years of completed schooling are also uncorrelated with age at arrival. This is inconsistent with case 3 because if $\sigma_s > 1$ then years of completed schooling should be increasing in age at arrival as immigrants work to remediate their

presumed deficit in human capital at arrival. This is the only additional evidence provided by comparing the predictions of the model to the empirical results derived from immigrants. To dig deeper into this matter, I turn to the available evidence from the existing literature on early childhood human capital.

4.2 Other Evidence on Early Childhood Human Capital

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 market 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 a large ratio ω_{c3}/ω_{c2} and a lower ratio ω_{s3}/ω_{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 Abercderian 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 Ceaușescu 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 Ceaușescu’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.

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. Critically, 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 experience of the Romanian orphans suggests both that early childhood human capital matters for late-life outcomes ($\lambda_s > 0$) and that it is difficult to intermediate deficits in early childhood human capital that have developed as little as six months after birth ($\sigma_s < 1$). The most revealing finding is that Romanian orphans showed a strong relationship between late-life outcomes and age at arrival, unlike refugees. The difference is that Romanian orphans received dramatic upgrades in their parental input (h_{ppc}), 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 their parents and so had similar parental inputs before and after migration.¹⁰ 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

¹⁰To verify this point I use the 1990 U.S. Census, which captures the household structure of the Indochinese refugees when they were young enough to be living in a household headed by someone else. I construct a sample of those who immigrated during the appropriate years and were five years old or younger at arrival, equivalent to the sample that I used for wage regressions in 2000–2010. I find that 95 percent of the sample was living with at least one biological parent; the majority of the remaining five percent were living with someone else of the same ethnic group.

Romanian orphans also changed their parental input $h_p p$.

4.3 Interpretation and Implications

In Section 4.1, I proposed three special cases of the model that were capable of replicating the data. Evidence drawn from the experience of refugees in the U.S. as well as the existing literature on early childhood human capital suggests that two of these three cases have counterfactual implications. There is substantial evidence that early childhood human capital affects late-life outcomes, including the results of structural estimation of the human capital production function, the results of early childhood interventions such as Perry Preschool, and the experiences of Romanian orphans. In this case I can rule out Case 2 ($\lambda_s = 0$). Likewise, there is substantial evidence suggesting that early childhood human capital deficits are difficult to remediate, including again structural estimates and particularly the experiences of Romanian orphans who were adopted before or after six months of age. In this case I can rule out Case 3 ($\sigma_s = \infty$). On the other hand, all of the evidence considered so far is consistent with Case 1 ($\omega_{c1} = \omega_{c2} = 0$).

It is important to re-iterate that Case 1 does not imply that early childhood human capital does not matter. Rather, it constrains the channels through which it can matter. In particular, it says that the effect of country environment and market goods are small in early childhood. There can still be a rich distribution of early childhood human capital within each country, driven for example by differences in human capital at birth h_0 and amplified by investments in the form of parental inputs $h_p p_c$. Further, it can still be the case that intervention-based changes of parental inputs have large effects (as long as $\omega_{c3} > 0$). For example, many of the existing interventions involve sending children to preschool programs; this can be thought of as implicitly substituting the time of qualified educational professionals for the time of parents. However, the average early childhood human capital will be the same across countries, leading to the conclusion that early childhood human capital is not important for understanding cross-country income differences.

The paper has less to say about human capital formation during the school years. It does document that there is a strong negative relationship between age at arrival and late-life outcomes for refugees who arrive after roughly age 8. This empirical finding is consistent with a number of alternative theories, but does not provide a convincing way to test among competing theories. Within the framework of the model, this negative relationship is consistent with a role for market-purchased goods after early childhood, as was previously found by Manuelli and Seshadri (2010) and Erosa et al. (2010). Likewise, it is consistent with the finding of Del Boca et al. (2012) that the importance of market-

purchased goods in human capital production rises over the life cycle. On the other hand, it is also consistent with important country environment effects after age 5. In a broader sense, a negative relationship after age 5 is also consistent with the view that refugees immigrated from countries with low education quality and hence will have less human capital after age 5 even if they remain enrolled in school (Schoellman, 2012). The fact that the negative slope often seems to start around age 8 or so also matches with the well-known decline in cognitive plasticity and particularly in language formation skills (Lenneberg, 1967). Bleakley and Chin (2010) previously showed that there is a similar pattern between age at arrival and English-language ability: a weak relationship up until around age 8 or so, then a negative relationship for those who arrive at older ages. Again I reiterate that the empirical findings are consistent with all of these theories but cannot distinguish between them.

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 three 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 “overinvestment” by immigrant parents in their children can explain my findings. Finally, 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 (Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, 2011). However, the Indochinese refugee flows were protracted, and the consensus is that during the 1980s In-

dochinese 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 2.3 and figure 1, but allow refugees to perfectly predict their future resettlement. Under this change in beliefs, the model predicts that future migrant families invest as much in their children as native families. After resettlement, the model predicts that refugees would have the same late-life outcomes as natives, with no relationship between age at arrival and late-life outcomes. In other words, allowing refugees to perfectly anticipate their resettlement generates predictions consistent with the data.

There are at least four 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. 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.

A third reason to be skeptical of this extreme case is that there is substantial evidence documented in the previous section that refugee children were disadvantaged at the time of entry relative to native-born children. Finally, a fourth 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 health and well-being measures of refugees taken shortly after their arrival and the experiences of early and late-arriving refugees.

5.2 Immigrant Overinvestment

My empirical findings relate to a growing literature on the outcomes for young first-generation and second-generation immigrants in the U.S. It is commonly held that immigrants invest heavily in their young children and that their children fare well in the U.S. The evidence is actually somewhat mixed; for example, Chiswick and DebBurman (2004) find that second-generation immigrants have higher levels of schooling than any other group, but Borjas (2006) finds that wage convergence for second-generation immigrants is incomplete. Nonetheless, given that this hypothesis is so common, I explore whether “overinvestment” by immigrant parents in their young children can explain my findings.

First, I note that high levels of investment by immigrants do not necessarily affect my empirical findings because I compare immigrants who arrive at different ages, not immigrants and natives. For parental investment patterns to explain away my findings it would need to be the case that immigrants who arrive at older ages have less human capital at arrival, but that their parents invest more in them, thus closing the human capital gap with immigrants who arrive at early ages. This pattern is inconsistent with the model because low human capital at arrival lowers the marginal productivity of all future investments. In the context of the model, parents of such children actually choose to invest less, not more, in their children.

I also provide empirical evidence on this hypothesis. To do so, I return to the 1990

U.S. Census, where I could link the Indochinese child refugees to their families.¹¹ Using this sample, 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. Differences in outcomes could arise if there was differential selection in age at arrival (despite the historical 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 and 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.

5.3 Measurement Error in Age at Arrival

A final 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

¹¹I can connect 87 percent to their biological mother and 81 percent to their biological father.

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 2.3, 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. It embedded a human capital production function in line with the literature into a cross-country model of human capital investment and labor market outcomes. Some human capital investment channels generated cross-country differences in early childhood human capital, but others did not. I derived a measure of the size of such differences that compares the late-life outcomes of otherwise identical immigrants who entered the U.S. at different ages. I implemented this measure using the Indochinese refugees as well as immigrants from other countries and consistently found no difference in late-life outcomes between those who immigrated at the beginning and the end of early childhood. These findings appear to be robust and are unlikely to be explained by selection.

I drew two conclusions from these findings. First, it seems 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. I also concluded that country environment

and market-purchased goods do not have a large role in early childhood human capital production. I showed that this interpretation is consistent with the existing evidence on early childhood human capital formation.

The empirical evidence presented here is inconclusive on two other important questions. First, the evidence does not pin down the importance of parental investment in children. I drew on the existing literature to argue that this channel probably operates but provided no independent evidence that it does. Second, the evidence is much less conclusive on the nature of investments made or the human capital acquired during schooling. The empirical finding of a negative relationship between late-life outcomes and age at arrival after roughly age 8 is consistent with a number of existing theories but does not discriminate between them. I view these questions as fruitful areas for future research.

References

- 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.
- Ben-Porath, Yoram**, “The Production of Human Capital and the Life Cycle of Earnings,” *Journal of Political Economy*, August 1967, *75* (4), 352–265.
- 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.
- , “Making It In America: Social Mobility in the Immigrant Population,” *The Future of Children*, 2006, *16* (2), 55–71.
- 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.

- Chiswick, Barry R. and Noyna DebBurman**, “Educational attainment: analysis by immigration generation,” *Economics of Education Review*, 2004, 23 (4), 361–379.
- 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.
- Haines, David W.**, *Safe Haven? A History of Refugees in America*, Sterling, VA: Kumarian Press, 2010.

- Heston, Alan, Robert Summers, and Bettina Aten**, *Penn World Table Version 7.0*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, May 2011.
- 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.
- 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.
- 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,” 1985. 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.
- Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat**, “World Population Prospects: The 2010 Revision,” 2011. New York, NY.
- 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.
- 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.
- Sack, William H., Greg Clarke, Chanrithy Him, Dan Dickason, Brian Goff, Kathleen Lanham, and J. David Kinzie**, “A 6-Year Follow-up Study of Cambodian Refugee Adolescents Traumatized as Children,” *Journal of the American Academy of Child and Adult Psychiatry*, 1993, 32 (2), 431–437.
- Schoellman, Todd**, “Education Quality and Development Accounting,” *Review of Economic Studies*, 2012, 79 (1), 388–417.

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 warfare. 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.

The available quantitative evidence also points towards a modest degree of selection. Contemporary studies noted that roughly two-thirds of Indochinese refugees during the peak years, and 40 percent of refugees during the first ten years, spoke no English (Office of Refugee Resettlement, 1985). Likewise, many newly-arrived refugees were illiterate, with 22 percent of Cambodians and 28 percent of Laotians reporting no schooling, a stark contrast to other large groups such as Cuban refugees (Haines, 2010).

Most studies report statistics for a broad set of immigrants. I add to this information by constructing a narrower comparison of refugee parents to non-migrant parents who remained in Indochina. I do so in two steps. First, I identify parents of Cambodian and

Table 1: Schooling Comparison: Refugees and Non-Migrants

Schooling	Men			Women		
	None	Primary	High School Grad	None	Primary	High School Grad
<i>Panel A: Vietnam</i>						
Refugees	10%	28%	62%	15%	37%	48%
Non-Migrants	5%	77%	18%	11%	79%	10%
<i>Panel B: Cambodia</i>						
Refugees	28%	30%	42%	40%	24%	26%
Non-Migrants	20%	76%	4%	50%	49%	1%

Table notes: Comparison of completed schooling of parents of refugee children in the 1990 U.S. Census to parents born 1940–1960 in the 1989 Vietnamese and 1998 Cambodian census.

Vietnamese-born refugees in the 1990 U.S. Census. I measure the education of their parents in three broad categories: none; some, but less than twelve years; and twelve or more years. I focus on education because it is the easiest variable to compare across countries.

In the second step I collect data on the educational attainment of non-migrants in Cambodia and Vietnam. Unlike Laos, these countries have conducted censuses that collect information on age, education, and family structure, with the earliest census taking place in 1989 in Vietnam and 1998 in Cambodia.¹² I measure the education of parents born between 1940 and 1960 in these countries in the same three categories; those birth years match up with the birth years of the parents of refugee children in the U.S.

Table 1 compares the education of mothers and fathers who migrated to the education of those who did not. The main result is that refugees from both countries come disproportionately from the extremes: both those with no education and those with at least a high school degree are overrepresented, while those with intermediate levels (such as some primary) are underrepresented. This evidence is consistent with selection, although it is also possible that refugees pursued education after arriving in the U.S. On the other hand, it is striking that 10–40 percent of refugees report no education at all, an outcome that is essentially unheard of for U.S.-born parents in the same cohorts.

¹²Available online at Minnesota Population Center (2010).

B Online Appendix: Derivations for the Isoelastic Model

From the text, the family's problem is:

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

Integrating out and substituting in the isoelastic case yields:

$$\max_{m_c, m_s, p_c, p_s, S} \quad \frac{A_i e^{(g-r)(S+5)}}{r-g} h_s - m_c - e^{-5r} m_s - A_i h_p p_c - A_i e^{5(g-r)} h_p p_s \quad (\text{B1})$$

$$\text{s.t.} \quad h_s = h_0^{\lambda_c \lambda_s} (z_i^{\omega_{c1}} m_c^{\omega_{c2}} (h_p p_c)^{\omega_{c3}})^{\lambda_s (1-\lambda_c)} (z_i^{\omega_{s1}} m_s^{\omega_{s2}} (h_p p_s)^{\omega_{s3}} S^{\omega_{s4}})^{1-\lambda_s} \quad (\text{B2})$$

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

The first-order conditions for the problem are:

$$S : \quad A_i e^{(g-r)(S+5)} h_s = \frac{A_i e^{(g-r)(S+5)}}{r-g} \omega_{s4} (1 - \lambda_s) \frac{h_s}{S} \quad (\text{B3})$$

$$m_c : \quad \frac{A_i e^{(g-r)(S+5)}}{r-g} \omega_{c2} \lambda_s (1 - \lambda_c) \frac{h_s}{m_c} = 1 \quad (\text{B4})$$

$$m_s : \quad \frac{A_i e^{(g-r)(S+5)}}{r-g} \omega_{s2} (1 - \lambda_s) \frac{h_s}{m_s} = e^{-5r} \quad (\text{B5})$$

$$p_c : \quad \frac{A_i e^{(g-r)(S+5)}}{r-g} \omega_{c3} \lambda_s (1 - \lambda_c) \frac{h_s}{p_c} = A_i h_p \quad (\text{B6})$$

$$p_s : \quad \frac{A_i e^{(g-r)(S+5)}}{r-g} \omega_{s3} (1 - \lambda_s) \frac{h_s}{p_s} = A_i e^{5(g-r)} h_p. \quad (\text{B7})$$

Inspection of (B3) reveals that it pins down $S = \frac{\omega_{s4}(1-\lambda_s)}{r-g}$, which implies that S does not vary within or across countries, verifying the first proposition. Equations (B4)–(B7) link together the optimal market goods and parental investments in the two periods. Inspection shows that the model predicts $m_c \propto m_s \propto A_i h_p p_s \propto A_i h_p p_c$, where the proportionality factors are functions of the share parameters (ω and λ) and discount rates (e^{-5r} and $e^{5(g-r)}$), and so do not vary within or across countries.

Using proportionality, it is possible to rewrite (B6) as:

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

where κ_1 captures functions of parameters and discount rates that do not vary within or across countries. Solve for the total parental input $h_p p_c$ in terms of exogenous parameters to find:

$$h_p p_c = \kappa_2 h_0^{\frac{\lambda_c \lambda_s}{1-\Psi}} z_i^{\frac{\omega_{c1} \lambda_s (1-\lambda_c) + \omega_{s1} (1-\lambda_s)}{1-\Psi}} A_i^{\frac{\omega_{c2} \lambda_s (1-\lambda_c) + \omega_{s2} (1-\lambda_s)}{1-\Psi}} \quad (\text{B8})$$

where κ_2 is again a constant and $\Psi \equiv (\omega_{c2} + \omega_{c3})\lambda_s(1 - \lambda_c) + (\omega_{s2} + \omega_{s3})(1 - \lambda_s)$ is the total returns to scale. Finally, use the proportionality relationship again as well as (B8) to substitute in for h_s to find:

$$h_s = \kappa_3 h_0^{\frac{\lambda_c \lambda_s}{1-\Psi}} z_i^{\frac{\omega_{c1} \lambda_s (1-\lambda_c) + \omega_{s1} (1-\lambda_s)}{1-\Psi}} A_i^{\frac{\omega_{c2} \lambda_s (1-\lambda_c) + \omega_{s2} (1-\lambda_s)}{1-\Psi}}. \quad (\text{B9})$$

Likewise, taking equation (B8) and the proportionality relationship and plugging in for h_c yields:

$$h_c = \kappa_4 h_0^{\lambda_c + \frac{\lambda_c \lambda_s}{1-\Psi} (\omega_{c2} + \omega_{c3})(1-\lambda_c)} z_i^{\omega_{c1} (1-\lambda_c) + \frac{\omega_{c1} \lambda_s (1-\lambda_c) + \omega_{s1} (1-\lambda_s)}{1-\Psi} (\omega_{c2} + \omega_{c3})(1-\lambda_c)} \times \quad (\text{B10})$$

$$A_i^{\omega_{c2} (1-\lambda_c) + \frac{\omega_{c2} \lambda_s (1-\lambda_c) + \omega_{s2} (1-\lambda_s)}{1-\Psi} (\omega_{c2} + \omega_{c3})(1-\lambda_c)}$$

The elasticity properties in Proposition 2 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 m_s , p_s , and S . Their problem then is:

$$\max_{m_s, p_s, S} \frac{A_i e^{(g-r)S + 5g}}{r - g} h_s - m_s - A_i e^{5g} h_p p_s \quad (\text{B11})$$

$$\text{s.t.} \quad h_s = h_c^{\lambda_s} (z_i^{\omega_{s1}} m_s^{\omega_{s2}} (h_p p_s)^{\omega_{s3}} S^{\omega_{s4}})^{1-\lambda_s} \quad (\text{B12})$$

The first-order conditions for the problem are:

$$S : \quad A_i e^{(g-r)S+5g} h_s = \frac{A_i e^{(g-r)S+5g}}{r-g} \omega_{s4} (1-\lambda_s) \frac{h_s}{S} \quad (\text{B13})$$

$$m_s : \quad \frac{A_i e^{(g-r)S+5g}}{r-g} \omega_{s2} (1-\lambda_s) \frac{h_s}{m_s} = 1 \quad (\text{B14})$$

$$p_s : \quad \frac{A_i e^{(g-r)S+5g}}{r-g} \omega_{s3} (1-\lambda_s) \frac{h_s}{p_s} = A_i e^{5g} h_p. \quad (\text{B15})$$

It is still the case that $S = \frac{\omega_{s4}(1-\lambda_s)}{r-g}$. Likewise, it is still the case that there is a proportionality relationship between the remaining two inputs, with $m_s \propto A_i h_p p_s$. Plugging this information into (B15) yields:

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

where κ_5 is a function of share parameters and other constants, and $\Phi \equiv (\omega_{s2} + \omega_{s3})(1-\lambda_s)$. Finally, substitution yields an expression for h_s :

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

where κ_6 is a final constant. $\Psi < 1$ is sufficient to guarantee $\Phi < 1$. In this case, it follows that human capital in the labor force is increasing in early childhood human capital for $\lambda_s > 0$, which verifies Proposition 3. This result arises through the direct effect of lower early childhood human capital but is also amplified by an endogenous decision to allocate less market-purchased goods and parental human capital to children with less early childhood human capital.

C 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

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

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 cohort effects:

$$y = \beta X + \alpha A + \omega^N Y^N + (\omega^I + \gamma^I) Y^I + \phi^I A A^I + (\delta^I - \gamma^I) Y U S^I + \varepsilon. \quad (\text{B16})$$

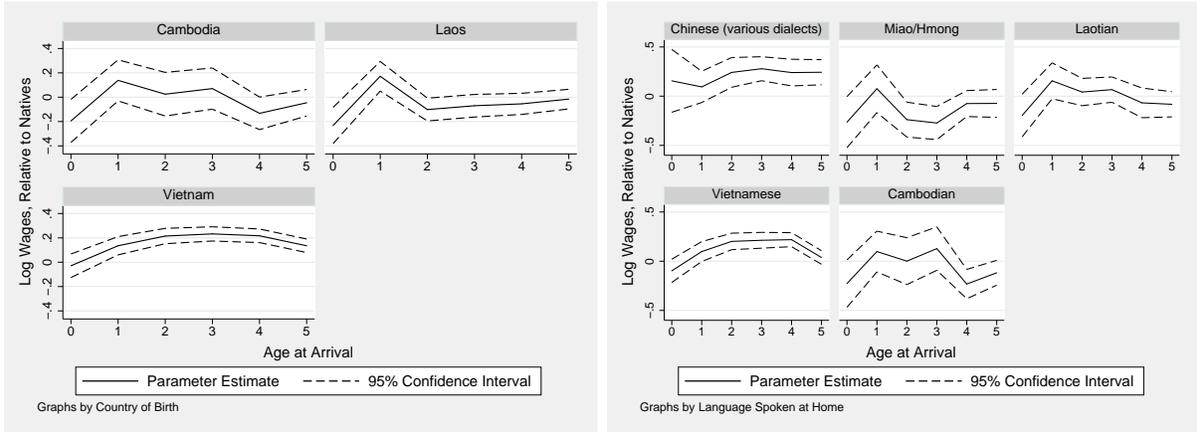
The effect of age at arrival is identified, although cohort effects and assimilation effects are not. This estimation model is more general than the one used in the text, because it also includes cohort effects as a regressor (even though the estimated coefficient does not measure “true” cohort effects).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

D Online Appendix: Robustness

D.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.

As mentioned in the text, U.S. immigration policy towards Indochinese refugees shifted in 1982. Prior to that time, health screening of refugees was cursory and only loosely related to admissions decisions; afterwards, the screening was improved and integrated into the admissions process. 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



(a) Birth Country

(b) Language Spoken at Home

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

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.

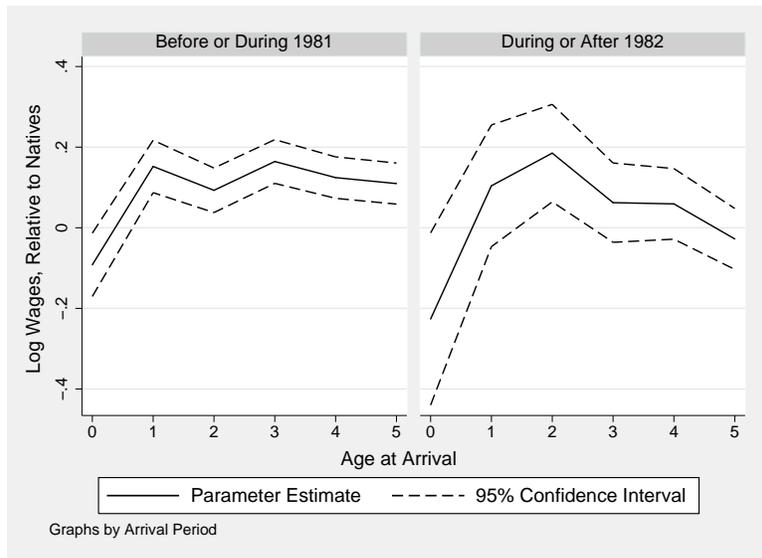


Figure 8: Log-Wage by Age at Arrival for Early and Late Arrivals

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, 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.

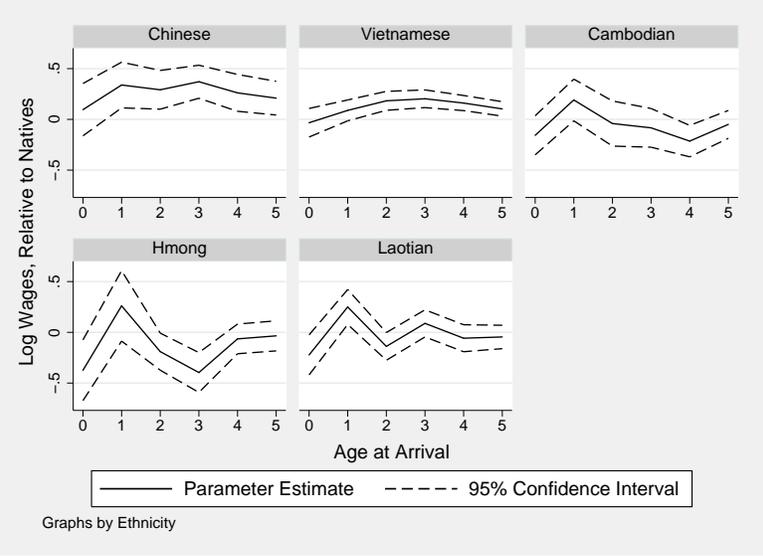


Figure 9: Log-Wage by Age at Arrival for Refugees Living Outside Enclaves

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.

D.2 Alternative Outcomes

Finally, I consider alternative outcomes besides those in the paper (log-wages and completed years of schooling). 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,

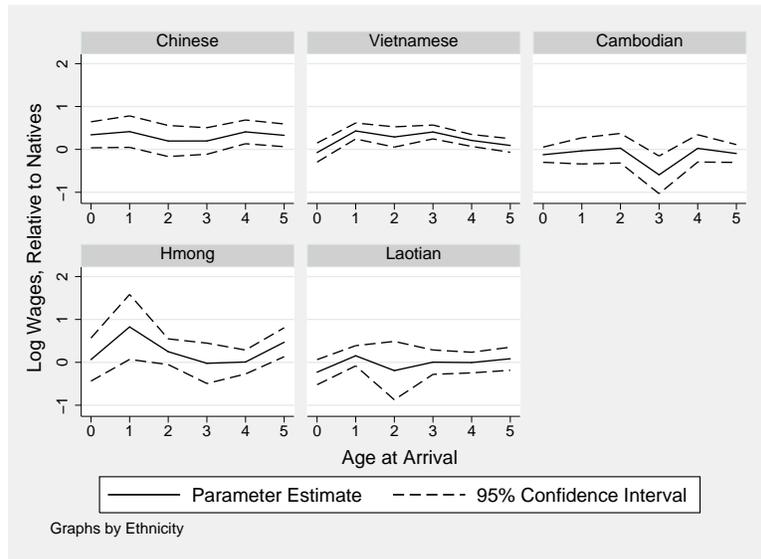


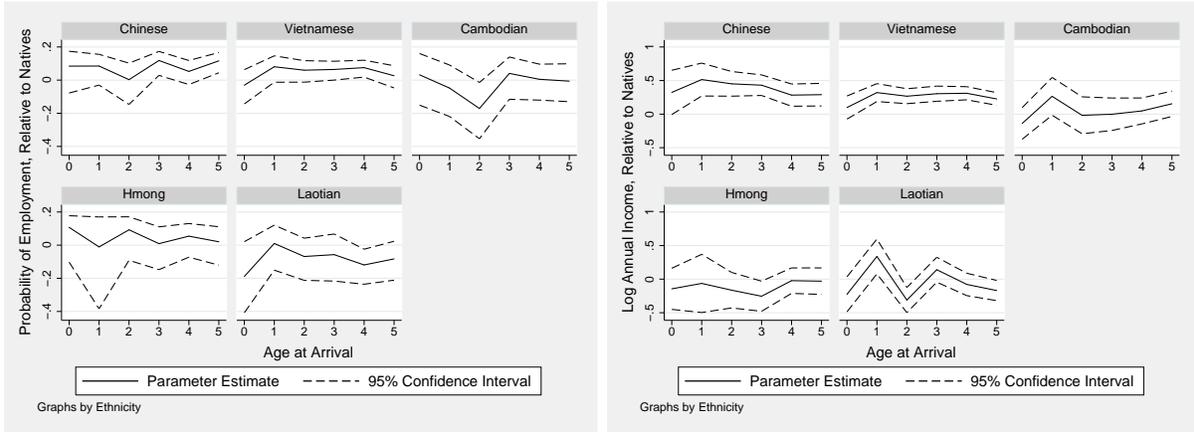
Figure 10: Log-Wage by Age at Arrival for Refugees 23–26 Years Old

with small level differences between refugees and natives.¹³ The same conclusion also implies 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 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 (6).

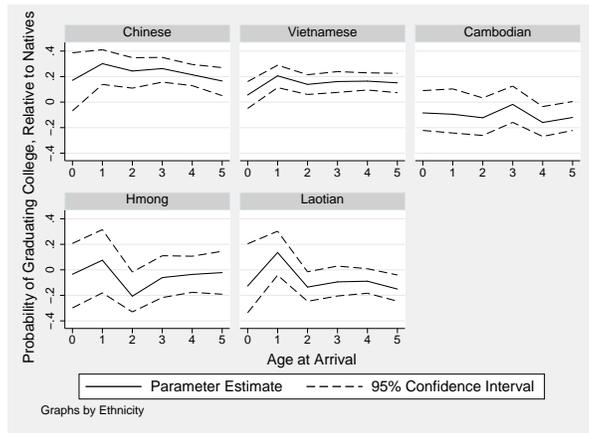
Figure 12 shows the results. Since the sample of 16–18 year olds is smaller, I pool all Indochinese refugees. I use negative age at arrival for the native-born children of refugee

¹³For 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

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.

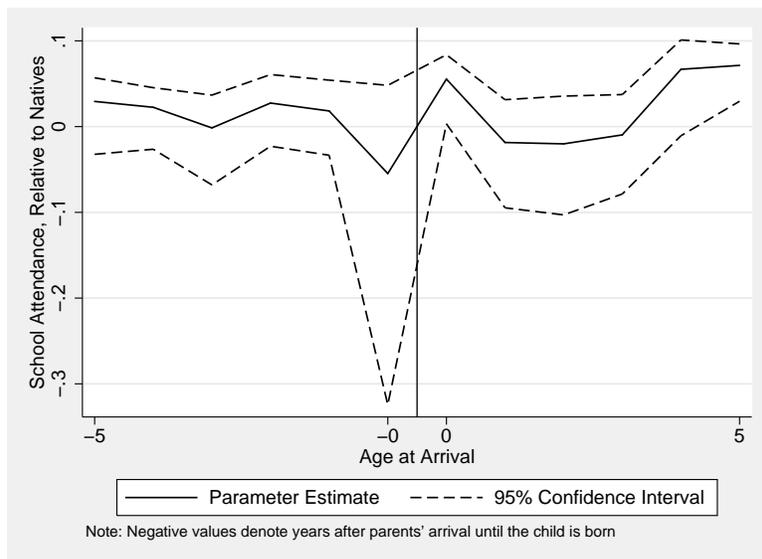


Figure 12: Probability of Attending School at Ages 16–18, by Age at Arrival