

Replication: Schoellman (2016) AEJ: Macro

Eric Hsienchen Chu*

University of Wisconsin-Madison

February 21, 2024

1 Paper Overview

Schoellman (2016) provided new evidence in the cross-country early childhood human capital of Indochinese refugees and their future earnings and education in the U.S. To carry out empirical analysis, Schoellman (2016) defined Indochinese refugees to be the group born in Vietnam, Cambodia, or Laos and migrated to the U.S. during the corresponding heavy refugee flows.¹ The author used 2005-2012 American Community Surveys (ACSs) and 2000 Population Census to construct the defined Indochinese refugees' adult outcomes, and the baseline model and figures suggested that, comparing to the U.S. natives, there is no obvious relationship between early arrival age of Indochinese refugees and their future earnings.

The rest of the report proceeds as follows: In Section 2, I briefly describe the Replication process and important notes to carry out this replication. In Section 3, I present replication results of Table 1 and Figure 2-5, which are the main findings and evidence in Schoellman (2016). In Section 4, I touch on replication difficulties, discuss the disparities between my estimations and the original paper, and briefly justify for my alternative IV identification approach as extension.

2 Replication Description

The replication resources for Schoellman (2016) is available here: <https://www.openicpsr.org/openicpsr/project/114117/version/V2/view>. This replication report is executed based on Version 2 package. Most of the data for this paper comes from IPUMS-USA, which was already included in the replication package above. The only necessary raw data to replicate Table 1 and Figure 2-5 is `'/census/usa_00131.dat'`. Additionally, one needs to download my modified Stata code, namely `'master-modified.do'` and `'ec_adult-modified.do'`, to carry out this replication exercise. Please see my Github repo for modified Stata programs and more detailed instructions.²

*Department of Economics, University of Wisconsin-Madison. hchu38@wisc.edu. This is a replication exercise for Spring 24 ECON580.

¹I consider this as relatively loose definitions to identify refugee groups so that the estimates are some overestimations of the true parameters. But even if under overestimation, the coefficients are statistical insignificant. Also see Section 4.

²The corresponding repo is accessible here: https://github.com/hsienc/econ_580_assignment_2/

3 Replication Results

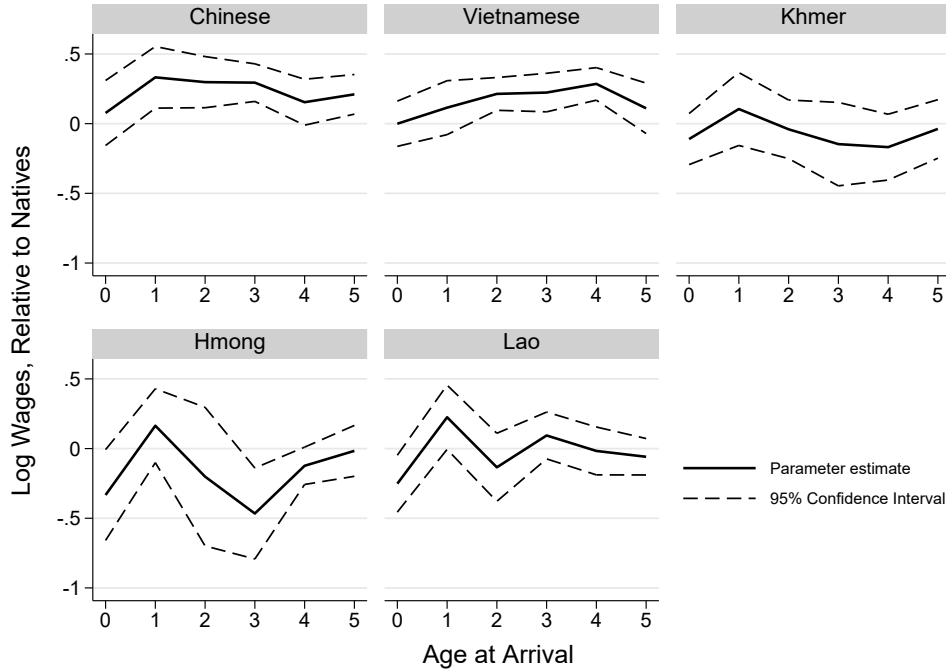
To keep track of the income differences and human capital, the outcomes of interest are the wages of refugees and their years of schooling. Table 1 replicates the baseline estimation in Schoellman (2016) by running the following stacked fixed effect OLS:

$$\mathbf{Y} = \beta \mathbf{X} + \sum_a \alpha_a d_a + \sum_y \omega_y d_y + \sum_{aa} \sum_e \phi_{aa,e} d_{aa,e} + \varepsilon, \quad (1)$$

where $\phi_{aa,e}$ is the main coefficient of interest that captures the **relative** wage levels of refugees with specific ethnicity to the wage levels of natives when we throw in bunch of controls and dummies.

Table 1 reported such replication estimates for the two outcome variables, looking at the interaction terms between ethnicity (e) and age at arrival (aa) to the U.S., and the benchmark results following the layout of Schoellman (2016). The estimates are fairly "consistent" in terms of signs and magnitudes, except for the Khmer ethnicity group (also see Section 4). None of the estimates in Table 1 are statistically significant, and we also do not see negative trend/slope for $\phi_{aa,e}$ in Figure 2. These indicates that refugees who arrived at the U.S. during early ages earn roughly equivalent wages to natives.

Figure 4A focuses on a longer range of age at arrival. In particular, it indeed exhibits negative correlation between age at arrival and refugee future earnings. This does not contradict to findings in Table 1 and Figure 2. Instead, it shows that the critical early childhood turns out to be the irrelevant period for explaining future outcome, which is one of the striking contribution of this paper. Therefore, Schoellman (2016) concluded that this human capital story can help predict future earnings above age six (i.e., lower bound) and that **arrivals at early childhood from age 0 to 5** have no relationship to refugee future earnings.



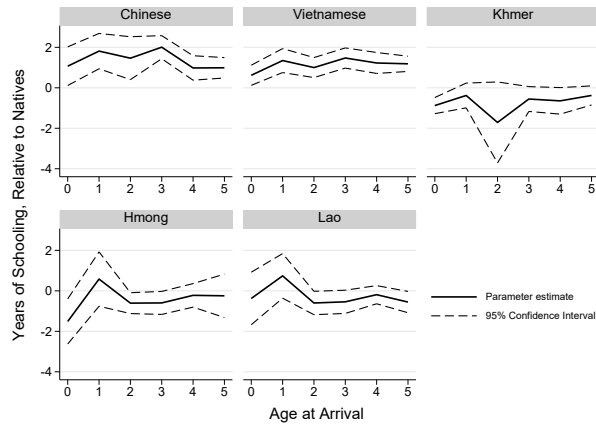
REPLICATION OF FIGURE 2: LOG-WAGES BY AGE AT ARRIVAL

REPLICATION OF TABLE 1: EFFECT OF AGE AT ARRIVAL ON WAGES AND SCHOOLING

	Vietnamese	Chinese	Lao	Hmong	Khmer	Pooled
<i>Dependent var: Wage (log)</i>						
Repl. Estimate	0.0179 (0.0224)	-0.0117 (0.0236)	-0.0029 (0.0258)	0.0425 (0.0351)	-0.0118 (0.0268)	0.0090 (0.0127)
Observations	800 (45.66%)	355 (20.26%)	265 (15.13%)	160 (9.13%)	172 (9.82%)	1,752 -
<i>Benchmark: Schoellman (2016)</i>						
Estimate	0.0064 (0.0156)	-0.0043 (0.0204)	-0.0090 (0.0224)	0.0454 (0.0268)	0.0116 (0.0228)	0.0070 (0.0097)
Observations	1,202 (46.72%)	500 (19.43%)	384 (14.92%)	267 (10.38%)	220 (8.55%)	2,573 -
<i>Dependent var: yrs of schooling</i>						
Repl. Estimate	0.0545 (0.0592)	-0.144 (0.0984)	-0.0923 (0.1043)	0.1422 (0.1321)	0.0894 (0.0677)	0.017 (0.038)
Observations	1,285 (44.96%)	540 (18.89%)	471 (16.48%)	258 (9.03%)	304 (10.64%)	2,858 -
<i>Benchmark: Schoellman (2016)</i>						
Estimate	0.058 (0.050)	-0.113 (0.083)	0.000 (0.089)	0.137 (0.103)	0.125 (0.063)	0.043 (0.033)
Observations	1,984 (46.83%)	757 (17.87%)	673 (15.88%)	362 (10.38%)	461 (10.88%)	4,237 -

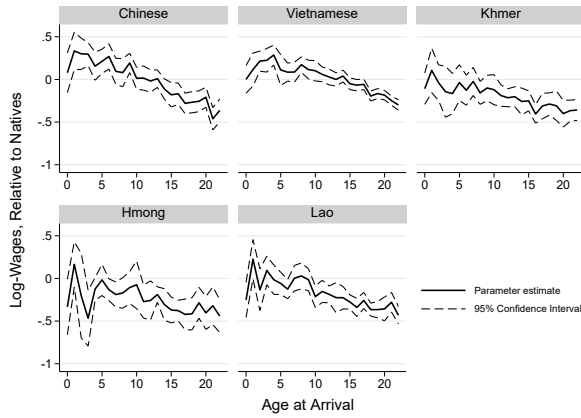
Notes: Reported effects control for age, state of residence, gender, and ethnic group fixed effect. Orders follow Schoellman (2016).

REPLICATION OF FIGURE 3: YEARS OF SCHOOLING BY AGE AT ARRIVAL

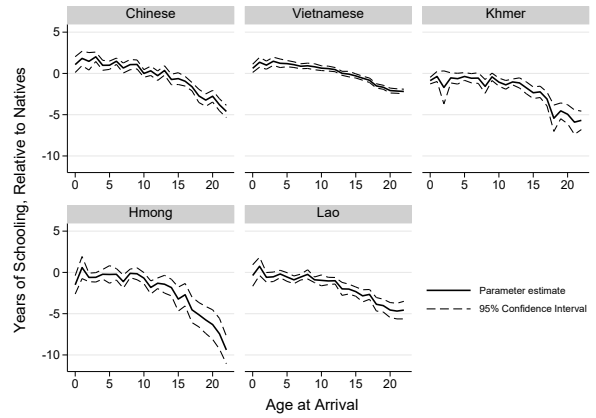


REPLICATION OF FIGURE 4: OUTCOMES BY AGE AT ARRIVAL FOR OLDER AGES

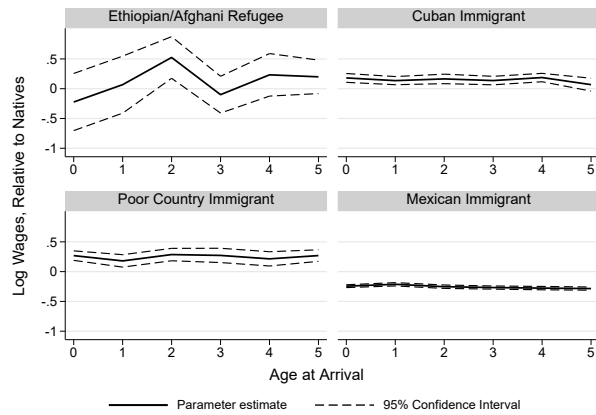
Panel A. log-wages



Panel B. Years of schooling



REPLICATION OF FIGURE 5: LOG-WAGES BY AGE AT ARRIVAL FOR OTHER IMMIGRANT GROUPS



4 Discussion & Extension

Schoellman (2016) provided interesting evidences stating that it exhibits no correlation between early human capital of refugees (arrival age) to their future wages. My notes are threefold. First, the earning disparity still emerge but only happening for older arrival ages (i.e., after childhood), and this paper contributes to the finding that the age-0-to-5 window is independent to adult outcome prediction. Second, the definition of refugees is simply by immigration heavy flows, so I consider Schoellman (2016) as providing upperbounds of estimates. Lastly, the baseline estimation only captures the relative effect comparing to natives, not suggesting inferences on aggregate effects.

Upon my replication, I noticed that the numbers of observations for replicated baseline estimations did not equate the numbers reported in the original paper. This issue arise possibly due to **runtime** or **computational constraints** under the basic Stata editions (SE). When loading in the (huge) ‘.dat’ data, I recalled my Stata program stopped responding on the SSCC WinStat server with some "I/O error" or "crashed warning" message. It may ultimately lead to a smaller sample size used for replications and generally larger standard errors in my Table 1 report.

Additionally, recall the the sign disparity of the Khmer group in Table 1 mentioned in Section 3. Even though the other estimates show fair robustness in sign and magnitude, the Khmer group behaves oppositely. I laid out three plausible explanations for this:

- (i) Among my "missing" samples, they may contain Khmer observations with extensively positive correlations between arrival age and earnings. This leads to a potential downward bias in my Table 1 replication estimate for Khmer. Yet, the share of observations by ethnicity in my report generally fit the shares in the original paper. I am inclined to believe the observation with (Khmer = 1) locates in the dataset uniformly.
- (ii) The author accidentally deleted the negative sign when formatting regression tables. While ‘eyeballing’ my replication Figure 2 or the original paper Figure 2, it seems like $\text{sgn}(\phi_{aa,e})$ for Khmer should be a small negative number. Considering the similarity between my reported estimate and the author’s estimate in absolute values, this may serve as one of the possibilities.
- (iii) Maybe the arrival age is somehow endogeneous so that OLS estimators from baseline models are biased and inconsistent? Even though Schoellman (2016) justifies that the unanticipated migration in Chapter VI and that early-age migration behaves naturally as exogeneous treatment to refugees, can it be that the entire refugee groups are still subject to selection bias? Maybe it just happens to be the case that some specific regions within the those countries have resources or proximity advantage for refugees to ‘escape’ and migrate to the U.S.?

For extension, though not directly showed in my report, I would consider merging quarterly marine weather data during that migration period and instrument the age-at-arrival variable just in case that the case (iii) endogeneity does not arise. I expect the IV regressions to be over-identified cases ($l \geq k$). The relevance condition is obvious: bad marine weather deters age at arrival. Allowing for heteroskedasticity, I would proceed with GMM overidentification test for exogeneity condition (check if $\mathbb{H}_0 : \mathbb{E}[Ze] = 0$ holds). If J stats fails to reject the null, we are good to go.

References

Schoellman, T. (2016). Early childhood human capital and development. *American Economic Journal: Macroeconomics*, 8(3), 145–74. <https://doi.org/10.1257/mac.20150117>