Dynamics of the Raw English Fluency Premium for Refugees and Other Immigrants in the U.S.

Abdihafit Shaeye*

Abstract: Previous work has established that U.S. immigrants earn more if they are fluent in English, but a portion of that premium likely reflects biases because fluency is correlated with unobserved factors like self-selection, ability to job-shop, and legal status. Since those unobserved factors represent potentially important barriers to immigrants' economic assimilation, this paper investigates them by looking for variation in the raw fluency premiums earned by different groups of immigrants. One key distinction is between refugees and other immigrants; while refugees have greater legal access to the labor market, non-refugees benefit from greater ability to self-select into both migration and (pre-migration) fluency, and those relative advantages change during the years after individuals migrate. Empirically, non-refugees initially earn a much larger raw wage premium for fluency (even though the effect of fluency on productivity is likely similar between the groups), suggesting that fluency is more strongly correlated with unobserved skills among non-refugees -- as expected given their greater opportunity to self-select. This gap persists in the first years after immigration, even though English-speaking refugees presumably have greater ability to take advantage of refugees' greater latitude in seeking more suitable employment matches. However, the refugees' premium does eventually grow at around the same time that more of them learn English, likely because more capable refugees are more likely to become fluent. Nevertheless, the gap never vanishes, possibly because the less successful non-refugees return-migrate, while that is not a practical option for most refugees. It thus appears that much of the variation in the fluency premium reflects correlations between fluency and other skills, rather than English-speakers' advantage in terms of job search.

JEL Classification: J15, J61, O15

Keywords: Refugees, English Investment, Selection, Job Shopping

*Department of Economics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA, Email: ahshaeye@uwm.edu. I am grateful to many readers who offered helpful comments, especially Scott Drewianka. I am solely responsible for all remaining shortcomings.

1 **Introduction:**

Acquisition of language fluency is one of the important means by which immigrants culturally and economically assimilate. A number of previous studies have shown that immigrants who are fluent in English typically have greater success in the U.S. labor market. Studies that have measured the effect of English fluency on immigrants' wages have generally found that fluent immigrants earn 17- 33 percent more than immigrants with limited English skills (Chiswick and Miller 1995; Bleakley and Chin 2004).

To date, however, most of this work has not examined whether the earnings premium to language proficiency could differ between refugees and non-refugees. This distinction is critical because the selection process into migration is different for refugees, and also because refugees have different rights than other immigrants after they arrive in the new country. These differences could affect the initial premium to language fluency as well as its evolution over time, and the patterns in those differences could shed light on the sources of the raw fluency premium.

It is possible that the previously estimated premiums truly reflect the value of the immigrants' language ability on their actual productivity, but it could also reflect other effects that are correlated with or mediated through language ability; for example, a portion of the raw fluency premium could be due to selection bias or even job shopping. Such biases mostly likely account for the variation in estimated fluency premiums across different groups of immigrants, since it seems likely that the true productivity of English fluency is very similar for everyone.

There are a few reasons why the measured return to fluency might differ between refugees and other immigrants both at arrival and over time. First, the estimated return to fluency may simply be a statistical artifact that reflects a strong correlation between fluency and other unobservable factors that employers value, such as intelligence or ambition. However, such a correlation is likely to be lower among refugees than among other immigrants (Chiswick 2008). If this is an important cause of the initial difference between the groups' fluency premiums, the gap would likely diminish over time as more refugees acquire fluency after immigrating, especially if the more capable or ambitious refugees learned English more quickly (Duleep and Regets 1999, Cortes 2004).

Second, some economic immigrants only migrate after having a job offer which is presumably more common among English speakers. But the difference between the groups' fluency premiums may diminish in the years after immigration if the refugees who are fluent utilize the fewer legal restrictions they face upon arrival and become more successful in climbing the job ladder. However, the convergence rate for the refugees may slow over time, first as the acquisition of fluency expands into the less capable groups of immigrants especially as the non-refugee immigrants gain legal opportunities to seek new jobs.

There are only a few studies that have examined the effect of language skills on refugee wages. In investigating the effect of language fluency on hourly wages of recently-arrived refugees, Shaeye (2015) finds that there is a much smaller premium than previously estimated by other studies which mainly focus on non-refugees. Although that study only investigates the fluency premium of recently arrived refugees, its findings raise new questions such as whether non-refugee immigrants receive a fluency premium from the moment they arrive and whether the refugees eventually receive a fluency premium comparable to the one of non-refugees. Finding higher estimated premiums to English fluency for newly arrived economic immigrants would reinforce the idea that there is something special about refugees, and if I found that the fluency premium grows faster over time for refugees, that result could have implications for the effectiveness of the refugee resettlement polices that are currently in place.

In this analysis, I investigate the difference between refugees' and non-refugees' raw fluency premiums and how that gap evolves over time. Unlike most studies, my goal is not to produce an unbiased estimate of the true effect of English fluency on refugees' and other immigrants' productivity in their jobs. In fact, my empirical strategy is based on the assumption that the true productivity effect is similar for those two groups, and thus that the difference between their estimated fluency premiums mainly reflects different levels of bias. While bias is ordinarily undesirable, in this case I will analyze it in order to gain insight into the relative importance of some difficult-to-observe factors like job-shopping and selection bias that may have large effects on immigrants' economic assimilation and are plausibly correlated with language fluency. Since there are institutional reasons to believe that the importance of those factors will evolve differently over time for refugees and non-refugees on a known schedule, we can infer the size of their effects by examining the dynamics of the difference between refugees' and non-refugees'

raw fluency premiums. Also, if the difference remains large after a very long time, we could then infer that it must represent some permanent unobservable difference between the groups that is correlated with fluency within the groups, possibly caused by differences in the processes by which the groups were originally selected into immigration.

I use the U.S. Census and subsequent waves of the American Community Survey to examine these questions. The advantage of this data is that it has a very large sample size, and the repeated cross-sections allow me to implement a synthetic cohorts approach. Neither data set reports on immigrants' admission status, including whether they are refugees, but I supplement this data with administrative data from U.S. immigration agencies that allow me to compute the share of immigrants from a given country in a given year who were refugees. Variation in that share over time allows me to measure the premiums to fluency among refugees, even in specifications that include country-of-origin fixed effects.

Results for the full sample of all regions of the world show that refugees have a lower fluency premium in the first five years after arrival, as the literature has already found. The fluency premium then grows for refugees for the next fifteen years, while for non-refugees it mainly grows during the last ten years. The two groups' fluency premiums begin to converge, although the difference is never completely eliminated. Furthermore, if I exclude the subsample from regions that are less likely to send refugees, or regions that send a lot of undocumented immigrants, I find that the two premiums do not converge.

The rest of the paper is organized as follows. Section two discusses background and related literature. Section three then presents the theory, and section four describes the data and methodology. Section five presents and discusses the empirical results, and section six summarizes the findings and concludes.

2 Background and Related Literature

Language ability is one of the important mechanisms for the economic and social assimilation of immigrants. Generally, immigrants with good language skills perform better in the labor market in terms of employment and wages compared to those who do not have those skills. Language skills can affect an individual's earnings for a number of reasons. First, those with good language skills may be able to communicate better with coworkers and customers, thereby directly

increasing their productivity and earnings. Second, better language skills may also increase one's earnings indirectly, by allowing him to accumulate more host-country education, enter high-skill occupations, expand networks, and search for jobs more effectively.

There has been a considerable empirical work on the correlation between immigrants' English fluency and their wages in the U.S. The estimated premium could reflect the true effect of language skills on their productivity. It could also reflect other effects that could be mediated through language skills, such as, selection bias, job shopping, other forms of job shopping, post-immigration improvements in language skills or even other investments in human capital. Studies that correct for the possible endogeneity have found large effects on the order of 17- 33 percent (Chiswick and Miller 1995; Bleakley and Chin 2004). Using Census data, Chiswick and Miller (1995) show that immigrants in the US who are fluent in English have earnings about 17 percent (57 percent for instrument variable) higher than immigrants with limited English skills. Similarly, Bleakley and Chin (2004), who exploits the phenomenon that younger children learn languages more easily than older children to construct an instrumental variable for language proficiency, also find that those who speak English earn 22.2 percent higher than non-English speakers (33.3 for IV). However, both of these studies average the immigrants across years of arrival and do not follow them over time as they spend more time in the U.S., and do they not consider the potential differential premiums to fluency between refugee and non-refugee immigrants, so their results are not necessarily generalizable to all immigrant groups in the U.S.

Immigrants can be separated into at least two distinct groups: Refugees and Non-refugees. Refugee immigrants, individuals fleeing persecution in their home country, and non-refugee immigrants, are individuals searching for better jobs or reuniting with family already in the U.S. Refugees are conceptually different from non-refugees and are not selected into either migration or fluency in the same way that non-refugees are (Cortes, 2004). The plausibly exogenous nature of the migration decision of refugees would lead them to experience worse labor market outcomes than the non-refugees in the initial years after migration, but over time this gap might disappear. An analogy can be drawn between the migration decision of an immigrant and labor market decision of a worker. Refugees are like those who were laid off due to plant closing and have to find jobs in different industries, where non-refugees are immigrants who quit their jobs, and usually have been involved in searching for work before they quit.

After arrival, refugees' movements in the labor market are less restricted than those of other immigrants. Many non-refugees have visas that are only valid for a single employer, so they cannot change jobs without overcoming legal barriers, whereas refugees face no such restrictions even in their first years in the U.S. labor market. Many non-refugees eventually obtain similar rights to job mobility, such as when they get a green card or citizenship, and at that point both groups have similar opportunities to job-shop. However, it usually takes the non-refugees at least 6 years and sometimes much longer to obtain those rights. In the years before that time, refugees have better options for finding their way into jobs that are better matched to their skills, and it seems likely that this process would have a larger advantage for the English-speaking refugees.

There are only few studies that have examined the impact of language skills on refugee wages, but these studies focus on narrowly defined groups of refugees. The only exception to my knowledge is the one by Shaeye (2015), which investigates the initial fluency premium for recently arrived refugees. He finds that there is no evidence that English ability is a major determinant of the wages of recently-arrived refugees, even though previous work has found that it is a major determinant of non-refugee immigrants' wages. The findings raise a new set of questions, such as whether non-refugee immigrants receive a fluency premium from the moment they arrive, and if not whether the refugees eventually receive a fluency premium.

3 Conceptual Framework: The Endogenous Choice of Migration and Human Capital Investment

In this section, I discuss the hypotheses that I expect to account for the patterns I observe in the English fluency premium. Consider an econometric framework in which the raw premium to English fluency could differ between the two groups upon arrival for two reasons. First, refugees are not as likely to be matched well with their initial jobs because they are less likely to arrive with jobs (Akresh 2008). In other words, English speaking refugees have had less opportunity to find the employers who most value their fluency. Second, it is also likely that there is a difference in the bias component of the raw fluency premium that reflects the correlation between fluency and the unobserved skills for the two groups. Since refugees' decisions to migrate were not primarily motivated by economic considerations, those with higher levels of unobserved skills have not necessarily moved to a new country where their language skills can best be accommodated. For example, a French-speaking economic migrant would be more likely

to migrate to a French-speaking country, whereas a French-speaking refugee does not necessarily have that luxury.

Formally, say the relationship between wages and human capital skills is as follows:

$$\omega_{ijt} = \beta X_{ijt} + \delta D_{it} + U_{ijt}, \qquad (1)$$

where ω is the log wage, D is a dummy for English skills, X is other relevant covariates, i indexes individuals, j indexes jobs and t represents time. The residuals term can be further decomposed as

$$U_{ijt} = a_i + b_{ij} + e_{ijt} \qquad (2)$$

Here, a_i is the ability of the individual, which is unobserved by the econometrician but relevant to employers; b_{ij} is a measure of the quality of the specific match between worker i and job j (which is also unobservable to the econometrician), and e_{ijt} is the remaining error, which is assumed to be uncorrelated with both a and b.

Under this model, the estimated raw fluency premium δ^{hat} converges to the sum of three components: (a) the true causal effect of language skills on productivity (δ), and biases associated with the correlations between fluency and (b) ability and (c) job match quality:

$$\delta^{hat} \rightarrow \delta + \text{Cov}(D, a_i)/\text{Var}(D) + \text{Cov}(D, b_i)/\text{Var}(D) \qquad \textbf{(3)}$$

Most likely δ is the same for both refugee and other immigrants (though it could differ if, e.g. one group were more likely to be initially employed in jobs that involve interactions with natives).

The covariance between fluency and ability (Cov (D, a_i)) could possibly represent an actual premium to fluency if the employer uses fluency as a signal about a_i, but otherwise if the employer observes a_i directly this component does not represent a causal effect of fluency on wages. It probably does differ between the groups because of economic immigrants' self-selection into migration. For example, those who do not have language skills that are valued in the U.S. but do have high ability are probably better off either staying in their homeland or else moving to somewhere else. Such selection, however, is less likely for refugees (Chiswick 2000).

The covariance between fluency and the job match parameter (Cov (D, b)) could probably represent a premium to language skills insofar as those skills help one search for a job that is a better match, but it is also possible that job match quality varies for other reasons besides fluency. It can also differ between the groups for two reasons. First, some economic immigrants only migrate after having a job offer, and that is presumably more common among English speakers. Second, over time refugees may have a better opportunity to increase b due to having fewer legal restrictions on job mobility.

Let us next look at the dynamics of the raw fluency premium δ over time

$$\frac{\widehat{\partial \delta}}{\partial t} \quad \rightarrow \quad \frac{\partial \delta}{\partial t} + \frac{\partial}{\partial t}\frac{Cov(D,a)}{Var(D)} + \frac{\partial}{\partial t}\frac{Cov(D,b)}{Var(D)} \qquad (4)$$

$\frac{\partial \delta}{\partial t}$ is plausibly same for everyone, and reaches to zero in the long run. The proposed hypotheses are represented by $\frac{\partial}{\partial t}\frac{Cov(D,a)}{Var(D)}$ and $\frac{\partial}{\partial t}\frac{Cov(D,b)}{Var(D)}$.

1) $\frac{\partial}{\partial t}\frac{Cov(D,b)}{Var(D)}$

This trend is presumably much larger factor for refugees at first because they have greater legal right to move between jobs (and they may also be more poorly matched to begin with, even conditional on fluency). However, a similar trend should eventually appear for economic migrants too as they acquire the permanent rights to search (e.g., via a green card).

2) $\frac{\partial}{\partial t}\frac{Cov(D,a)}{Var(D)}$

It seems most likely that this trend is likely to be higher for the refugees ($\frac{\partial}{\partial t}\frac{Cov(D,a)}{Var(D)} > 0$), at least over the first few years after arrival, assuming the higher ability refugees are more likely to acquire language skills faster (Duleep and Regets 1999, Cortes 2004). This seems likely due to both their greater ability and their stronger incentive (assuming these skills are complementary to ability in the labor market). However, this trend could also be higher for non-refugees due to

return migration if those who return their homelands are both less fluent and less successful. (Presumably very few refugees become return migrants)

4 Data

The analysis uses repeated cross-sections of individual level data of immigrants on both the 2000 U.S. Census of Population, and American Community Survey (ACS) from the years 2001 to 2014. I can follow year-of-arrival immigrant cohorts who came to the U.S. between 1993 and 2013. I focus only on male immigrants who are 18 and above and below 66 years of age, are in the labor force, work for wages or salary income, are non-farmer workers, and living outside group quarters. I also exclude those who report that they are self-employed. 238,635 of those in the sample are considered non-refugees, where 59,855 of them are considered refugees.

The advantage of the U.S. census and subsequent waves of the American Community Survey data, though, is that it has a very large sample size, and the repeated cross-sections allow adopting a synthetic cohorts approach. The response to the year of arrival is also more accurate than the responses of earlier censuses.

The legal definition of an immigrant is a person lawfully admitted for permanent residence in the United States." (INS Annual Report: Legal Immigration, 2000). Throughout this paper, however, I use the census definition of immigrants as individuals who were born outside of U.S. with no citizenship at birth, which includes both refugees and non-refugees. Non-refugees include those on student, work, or other temporary visas; persons residing in the country without authorization; legal immigrants; and naturalized citizens. I exclude from the sample those immigrants from some regions that send a lot of undocumented immigrants, particularly Mexico and Central America, as well immigrants from broad geographic regions that send very few refugees to the U.S., including Northern Europe, Australia, and East Asia. In other words, I will restrict the sample to only those who come from Eastern Europe, Southeast Asia, Middle East, South America, the Caribbean, and also Africa.

Ideally I would like to have a panel data on wages and English skills for persons who are clearly identified as refugees, or non-refugees, but this information in unfortunately not included in the Census or any other large data sets. Most of the literature thus imputes refugee status based on immigrants' countries of origin (e.g. Cortes, 2004). The problem with this strategy is that it treats

all immigrants from same country the same way, even if the majority of those arrived in some years were not refugees. Bollinger and Hagstrom (2008) thus propose to identify refugees by using a combination of year of immigration, gender, country of origin, and age at immigration, and I will follow a similar strategy in the analysis below.

4.1. Assignment of Refugee Status

Specifically, I use two different strategies for attributing refugee status, both of which are based on two characteristics: individuals' country of birth and their year of arrival in the U.S. The first method involves determining how many people are likely to have migrated to the U.S. as refugees from a particular country in a particular year, which I estimate as the ratio of the actual number of refugees admitted to the U.S. from a given country in a given year to the total number of people (whether on immigrant or non-immigrant visas) admitted from a that country in that year, as reported in the government's official administrative immigration records. Due to measurement error, this ratio is relatively conservative and might produce more conservative estimates since I do not count asylees among the refugees. On the other hand, I do not observe counts of undocumented immigrants, which would tend to bias the refugee ratio upwards. However, most countries that send large number of undocumented immigrants do not send large number of refugees, so the refugee ratio is still accurately computed at close to zero[1]. Conversely, countries that do send large numbers of refugees to the U.S. usually do not also send large number of undocumented migrants, in part because they are generally located far away from the U.S. The larger problem may be with asylees, especially those from Western Hemisphere countries like Colombia and Haiti. Some people from these countries come to the U.S. without documentation, hoping to apply for asylum after they arrive, and thus are not counted among the non-refugee immigrants, but they are also not counted as refugees unless they achieve refugee status-which is different from asylee status.

To assess whether refugees are correctly identified in the first method, I utilize an alternative identification strategy in which the denominator of the ratio is computed from the Census sample

[1] For example, data from the U.S. Census American Community Survey (ACS) and the Department of Homeland Security (DHS) suggest that between 40 and 60 percent of Mexicans in the U.S. are undocumented, but the refugee ratio is still accurately computed at near zero because there are only a handful (perhaps few hundred per year) of refugees from Mexico.

instead. I estimate the whole sample by dividing the actual number of respondents to their respective sample percentage, which will give us the total estimated population that the census data represents. Only a very small number of country-years are reclassified from refugee-granting to non-refugee-granting countries. So it is not surprising that the estimates of the fluency premium are very similar under the two methods.

To smooth out the possibility of measurement error due to misreporting of the immigration year or a discrepancy between the Census year and the fiscal year (which is what is reported in the immigration data), I work with a three-year moving average across both the year of individuals' immigration and the Census year. However, this change makes almost no difference whatsoever. In practice, so it appears that this sort of measurement error it not quite serious problem anyhow.

4.2. Summary and Data Characteristics

I first present summary statistics in Table 1 and then discuss some other demographic and human capital characteristics by immigrant status for immigrants who come from Eastern Europe, South East Asia and the Middle East, South America, the Caribbean, and Africa. Data and differences in both institutional and cultural channels for immigrant assimilation guide the division of time into three phases. During the initial "adjustment" period (the first years after immigration), refugees have more legal rights than other immigrants and potentially have more opportunities for job-shopping. During the transition period (6-12 years after arrival), non-refugees acquire rights equivalent to the ones that refugees had before. Finally, during an "integration" period (13-20 years after arrival), all immigrants should have similar paths for economic assimilation.

Non-refugees have higher employment rates than refugees during the adjustment period, consistent with the expectation that non-refugees are more likely to have a job ready upon arrival. However, refugee employment rates eventually catch up with those of non-refugees, and the two rates grow similarly afterwards. In general, non-refugees have higher annual and hourly earnings than refugees, and this persists throughout the period of study.

Table 1 Insert Here

Figures 1, 2, 3, and 4 show the ability to speak English, level of education, school attendance and citizenship by years in the U.S. for both refugees and non-refugees. In general, non-refugees have higher fluency rates than refugees throughout the period, but language acquisition is higher

for refugees both during adjustment and transition periods. It is also true that refugees' raw fluency premium grows faster during those two periods. One potential explanation is that there was initially a stronger correlation between fluency and unobserved skills- as represented by higher levels of education - among the non-refugees, but the correlation among refugees grew stronger over time as the most capable refugees learned English more quickly. Figure 4 shows that refugees who are fluent in English are more likely to have high school diploma or higher than non-refugees especially after the adjustment period. Immigrants who are fluent in English are also more likely to be younger in age than those who are not fluent in English across immigration status. This is intuitive because it is easier for young people to acquire English than older immigrants all else equal.

Figures 1 and 2 Insert Here

Besides the importance of education for immigrant wages, it could also be a proxy for the correlation between fluency and unobservable skills. Although the level of education grows in parallel for all immigrants over time, non-refugees have higher education levels than refugees, and they maintain that through the period of study. The correlation between fluency and education is also higher for non-refugees during adjustment period, but refugees close the gap at the end of the adjustment period. The correlation between English and education then grows faster for refugees during the integration period. Conditional on fluency, non-refugees are likely to have higher levels of education than non-refugees during the adjustment period, but refugees over take them thereafter.

Figure 3 Insert Here

As reported in figure 1, non-refugees are more likely to attend school than refugees during the adjustment period. It also true that refugees acquire higher language skills faster during this period than non-refugees, suggesting that non-refugees are more likely to enroll in order to study subjects other than English. School attendance for both groups is quite similar after the adjustment period.

Figure 4 Insert Here

It is also not surprising that refugees are more likely than non-refugees to become naturalized citizens within six years of arrival, but they maintain that advantage throughout the period of study. Fluent non-refugees are more likely than non-fluent non-refugees to become naturalized citizens, which could suggest that those who are unsuccessful among the non-refugees would more likely return-migrate.

Figure 5 Insert Here

5 Empirical Specification and Results

5.1 Empirical Specification

The econometric model tests the relationship between language ability and the wages[2] of male refugees using the standard human capital earnings function in labor economics. I initially use pooled cross-sectional ordinary least squares (OLS) and country fixed effects (FE) regressions to estimate the relationship between English language fluency and wages for three different time periods: adjustment, intermediate and integration. The log of reported yearly wages is the dependent variable, and English speaking ability serves as the main independent variable.

The assumption here is that there is a log-linear relationship between wages and language skills:

$$\omega_{it} = \beta X_{it} + \delta_{\rho} D_i + U_{it}, \qquad (2)$$

Log ω denotes log weekly wages, i is an index for individuals, and t is an index for the different time periods of adjustment and integration phases of the assimilation process. D is the focal explanatory variable, a dummy for English fluency. X is a set of demographic controls such as age, education, native wages, year of census, marriage status, race, citizenship, age at arrival, state of residence and, linguistic distant measure as a proxy for enclaves, and ε_{it} represents the error term. Controlling for these characteristics are important, especially for non-refugees since they are more heterogeneous group in terms of immigration admission class and motive.

I hypothesize that refugees have lower fluency premium upon arrival, but are expected to see higher growth in their premium than non-refugees' over time. I identify the language fluency premium by looking at the difference-in-difference in fluency premium between the two groups

[2] I windorize the wages by dropping from the sample the observations in the bottom and upper 1% tails of the distribution to account for any potential outliers, but the results are quite similar.

in each assimilation phase. I also look the within group dynamics of this fluency over the long-run.

5.2 Empirical Results

5.2.1 Initial premiums to fluency

I first estimate the initial fluency premium for both refugees and non-refugees at the adjustment period and report the results in the first two lines of Table 2. After netting out the differences in human capital characteristics and socioeconomic background, I find that upon arrival non-refugee immigrants earn higher fluency[3] premiums on their hourly[4] wages than immigrants who arrived as refugees. This confirms that the two groups initially earn different premiums for fluency, a finding that is consistent with earlier studies by Shaeye (2015) and Chiswick and Miller (1995).

5.2.2 Evolution of the Fluency Premium

I next look at the dynamics of the fluency premium over time. As the second and third columns of the first section in Table 2 show, preliminary results[5] from OLS estimation indicate that the fluency premium grows during both the transition and integration periods for both groups, and their fluency premiums begin to converge, although the difference is never completely eliminated. The pattern for non-refugee immigrants is driven by the non-refugees from Central America and Mexico. However, the inclusion of those from Central America and Mexico only matters a great deal when I consider the full sample and do not exclude those with less than high school diploma. If I exclude immigrants from Central America and Mexico – which I will do for the rest of the analysis – because of their home countries' geographic proximity to the U.S. and high rates of undocumented immigration, or restrict the sample to those with high school

[3] Similar results emerge if we define "fluent" to include only those who speak English only and those who speak English very well, or if we exclude those who speak only English.

[4] The estimates do not change substantially if we instead measure weekly wages using the "weeks worked last year" variable, even though this procedure most likely introduces additional measurement error.

[5] This result is robust to both restricting the sample to only those with high school diploma and higher and also excluding those from English-speaking countries.

diploma or higher, I find that both refugee and documented non-refugee fluency premiums grow both during both transition and integration period, but the two groups' premiums do not converge (See Table 2). In other words, there is a parallel upward shift of the sample for non-refugees compared to the sample that includes those from Central America and Mexico (steeper slope for non-refugees. This means the two groups' premiums grow hand in hand, but the initial difference persists over time.

When I control for country fixed effects, the difference between the two groups still persists throughout all of the periods. This non-convergence is being driven by non-refugees who come from developed countries such as those from Northern Europe, Canada, Australia, and Japan. However, if I further restrict the sample to only those from Eastern Europe, Southeast Asia, Middle East, Latin America, and Africa - which are the regions that are more likely send refugees - the fixed effects estimates reported in Table 3 begin to converge during the integration period. That is mainly due to changes in the estimates for the non-refugee groups. For example, the estimated premium for non-refugees during the integration period is 39 percent under OLS and 33 percent under fixed effects, whereas the comparable estimates for refugees are 30.0 and 30.5 percent, respectively. This suggests the estimated premium[6] under OLS was biased upwards by between-country difference within the non-refugee group.

The dynamic of the premium and its growth for both groups are robust even when I account for immigrants' occupations, or restrict the sample to those with high school diploma or higher. The estimated fluency premium, which is reported in Table 3, falls if I control for occupation, or separate those who have at least a high school diploma or higher into those with associates degree, and those who have bachelor degree or higher. However, those changes are essentially uniform across time periods and regions of origin, so the differences in the fluency premiums largely remain intact and most of them are no longer statistically significant. This makes sense because the fluency premium is most likely be mediated through occupation, and there could be a strong, positive relationship between earnings and the occupational English requirements. The

[6] I censor those who make minimum wage from the left to check if results are sensitive to minimum wage, but the results remain unchanged.

fluency premium is also capturing part of the education return (especially for higher levels of education), suggesting that it may serve as a proxy for the correlation between earnings and unobserved skills.

I then look at the different regions individually and report the estimates[7] in Table 4. Most results are similar to those from the pooled sample without fixed effects. However, there are differences in the initial premium between refugees from different regions, with larger estimated fluency premiums for those from Eastern Europe and Cuba than for those from Asia and Africa. The premiums earned by the refugees from the former sets of countries are also more similar to those earned by non-refugees from the same countries than are those from elsewhere in the world. The differences between the two groups nearly vanish when I control for fixed effects, except for Asia[8].

One possible explanation for that difference is that the events that cause refugees to come to the U.S. from Eastern Europe or Cuba are generally most relevant for only a relatively modest subgroup of the population, whereas the refugee-creating events in Asia and Africa have generally been more severe and more pervasive. Those who are most strongly affected by big events likely constitute a less selected sample. The countries that produce those big events are also more likely to be less developed than are the countries that experience more targeted shocks. It could also have something to do with the fact that refugees from Africa and Asia are more likely to spend a long time in refugee camps before they are resettled, and during that period they are less likely to expand their formal education or to increase their work experience.

5.3 Discussion of Results

Let me revisit the hypotheses I have laid out already and talk about how well they explain the results of the fluency premium for the two groups and its growth pattern over time. The different hypotheses would predict different dynamics in the evolution of the groups' fluency premiums, although the net effect remains ambiguous ex ante. Selection on both who migrates and on who acquires language skills would predict that there should be a higher premium to fluency for non-refugees at first (i.e., a larger premium shortly after the time of immigration), but that refugees might see higher growth later (a steeper trend in the estimated fluency premium) as the more

[7] See graphical representation of results in the appendix
[8] The non-convergence in the Asia region is an anomaly and needs further investigation.

capable among them acquire English and thus become more selected over time. On the other hand, selection on who can job-shop predicts the growth in the premium to fluency would be higher for refugees in the first few years after immigration if that is important for refugees, but the growth in the fluency premium for the non-refugees might eventually catch up with it once they acquire more legal rights. Selection on who decides to return to their home country would predict greater acceleration in the fluency premium for non-refugees later as the non-successful among them, who are more likely to be non-English speakers, return-migrate.

Results confirm that non-refugees receive a higher initial fluency premium, that is in part because they are more strongly selected into fluency (as seen by their higher initial correlation between fluency and education) and also because they are more likely to arrive in the U.S. with a job already in hand. Results also confirm that refugees experience higher premium growth over time. Figure 15 suggests that refugees with high school diploma or higher are more likely to be fluent in English than those without those credentials during transition and integration periods, which could have contributed to the premium growth of refugees.

Figure 15 Insert Here

The job-shopping hypothesis would predict that refugees would be able to shop for jobs right upon arrival since they have broader legal rights than non-refugees. Results show that the premium for non-refugees does not substantially grow during the adjustment period. This is not surprising since non-refugees have fewer rights to change employers during this period. In view of their greater legal rights to search for employment, it is somewhat surprising that the fluency premium for refugees does not increase faster during the adjustment period. One potential explanation is that it may be a consequence of the refugee self-sufficiency policy, which promotes early employment and thus mainly enrolls refugees in low-skilled jobs - which could mean an occupational downgrade for some of them. This also makes it difficult for refugees to transition into higher-skilled ones during this period.

Return migration hypothesis would predict that non-refugee immigrants especially the less successful among them -who are more likely to be non-English speakers- would decide to return to their home country and thus would make their premium grow further during the integration period as they even become more selected.

6 Summary

I develop an econometric framework in which the raw premium to English fluency could differ between refugees and non-refugees. I exploit individual level data of immigrants on both the 2000 U.S. Census of Population, and American Community Survey (ACS) to test these hypotheses. Consistent with previous work, I find that non-refugees initially earn a larger premium for English fluency, most likely because they are more strongly self-selected into migration on the basis of fluency and because they are more likely to arrive with a job to which they are at least somewhat well-matched. However, the refugees' premium grows during both adjustment and integration periods, probably because the correlation between unobserved skills, and English is getting stronger as more capable refugees learn the language over time. Although the premium grows faster for refugees during the integration period, it does not catch up with the premium of non-refugees, who even become more selected as the less successful among them return to their home country. The fixed effect specification reveals that between-country differences within the non-refugee group drive the non-convergence between the two groups.

In other words, although refugees have more legal rights at first and there is an increase in their fluency levels over time, differences in selection remain the most important factor for the two groups' premiums difference throughout the period of study.

An extension for future research is exploring whether other human capital skills such as schooling follow similar pattern, or whether the differential premium to skills across the two groups can explain the wage-gap between refugees and non-refugees documented in the literature.

References

Akresh, Ilana Redstone. "Occupational trajectories of legal US immigrants: Downgrading and recovery." Population and Development Review 34.3 (2008): 435-456.

Bleakley, Hoyt, and Aimee Chin. "Language skills and earnings: Evidence from childhood immigrants*." Review of Economics and Statistics 86.2 (2004): 481-496.

Bollinger, Christopher R., and Paul Hagstrom. "Food stamp program participation of refugees and immigrants." Southern Economic Journal (2008), 74(3): 665-692.

Chiswick, Barry R., and Paul W. Miller. "The endogeneity between language and earnings: International analyses." Journal of Labor Economics (1995): 246-288.

Chiswick, Barry R. "Are immigrants favorably self-selected? An economic analysis." An Economic Analysis (March 2000). IZA Discussion Paper 131 (2000).

Cortes, Kalena E. "Are refugees different from economic immigrants? Some empirical evidence on the heterogeneity of immigrant groups in the United States." Review of Economics and Statistics 86.2 (2004): 465-480.

Department of Homeland Security (DHS), Immigration Statistics Data,1991-2014. Available at https://www.dhs.gov/immigration-data-statistics.

Department of State (DS), Immigrant Visa Statistics, 1991-2014.Available at www.travel.state.gov

Duleep, Harriet Orcutt, and Mark C. Regets. "Immigrants and human-capital investment." The American Economic Review 89.2 (1999): 186-191.

Shaeye, Abdihafit H. "The Unexpectedly Small Wage Premium for English Fluency among Recent U.S. Refugees" Working Paper, University of Wisconsin in Milwaukee (2015).

Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2015.

Table 1-A
Refugees

	Full Sample (N= 59,855)		English-Speakers (N= 353,87)		Non-English Speakers (N= 24,468)	
Variables	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**
Log Hourly wage (conditional on employment)	4.68	0.71	4.76	0.73	4.54	0.65
Employment	0.68	0.49	0.74	0.43	0.66	0.47
High School and Above	0.60	0.47	0.66	0.47	0.50	0.50
School Attendance	0.26	0.44	0.37	0.48	0.11	0.31
Age	37.3	17.7	31.8	15.4	45.2	17.8
Citizenship	0.36	0.48	0.43	0.49	0.26	0.44

.

Table 1-B-Non-Refugees

	Full Sample (N= 357,461)		English-Speakers (N= 283,102)		Non-English Speakers (N= 74,359)	
Variables	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**
Log Hourly wage (conditional on employment)	4.88	0.80	4.95	0.81	4.59	0.66
Employment	0.917	0.27	0.92	0.26	0.89	0.30
High School and Above	0.68	0.46	0.71	0.45	0.56	0.49
School Attendance	0.32	0.47	0.37	0.48	0.14	0.35
Age	33.7	16.4	31.4	14.8	42.8	18.8
Citizenship	0.32	0.46	0.35	0.47	0.218	0.413

Notes: Regression controls include age, education, race, Hispanic status, whether attending school, native wage measure, network measure, and state and year dummies.

***, **, ** indicate significance at 1 percent, 5 percent, and 10 percent levels.

	Table 2: Regressions on Log Hourly Wages			Sample Size	R^2
Method/Period	0-5	6-12	13-20		
OLS -Whole Sample					
Fluency-Non-Refugees	0.226*** (0.05)	0.262*** (0.07)	0.324*** (0.08)	742396	0.626
Fluency-Refugees	.084*** (0.034)	0.179*** (0.04)	0.297*** (0.048)	34863	0.605
Difference	-0.142*** (0.045)	-0.083*** (0.072)	-0.028 (0.094)	777259	0.625
Fixed-Effects- Whole Sample					
Fluency-Non-Refugees	0.196*** (0.038)	0.208*** (0.053)	0.261*** (0.068)	742396	0.645
Fluency-Refugees	0.170** (0.020)	0.220*** (0.020)	0.318*** (0.024)	34863	0.607
Difference	-0.025 (0.044)	0.011 (0.058)	0.056 (0.075)	777259	0.643
OLS-Excluding Central America, and Mexico					
Fluency-Non-Refugees	0.370*** (0.043)	0.443*** (0.057)	0.528*** (0.069)	409986	0.636
Fluency-Refugees	0.118** (0.032)	0.192*** (0.26)	0.289*** (0.03)	34863	0.605
Difference	-0.252*** (0.059)	-0.250*** (0.069)	-0.238 *** (0.084)	444849	0.635
Country-Fixed Effects- Excluding Central America, and Mexico					
Fluency –Non-Refugee	0.30*** (0.027)	0.37*** (0.058)	0.477*** (0.085)	409986	0.656
Fluency-Refugees	0.171*** (0.023)	0.213*** (0.025)	0.297*** (0.022)	34863	0.607
Difference	-0.128*** (0.039)	-0.157*** (0.072)	-0.180*** (0.102)	444849	0.656

	Table 3: Regressions on Log Hourly Wages			Sample Size	R^2
Method/Period	0-5	6-12	13-20		
OLS -Specific Regions					
Fluency-Non-Refugees	0.277*** (0.028)	0.313*** (0.027)	0.388*** (0.026)	219875	0.625
Fluency-Refugees	0.147*** (0.022)	0.213*** (0.018)	0.301*** (0.022)	34863	0.605
Difference	-0.130 *** (0.036)	-0.10*** (0.032)	-0.086*** (0.034)	258638	0.623
Fixed-Effects- Specific Regions					
Fluency-Non-Refugees	0.225*** (0.023)	0.268*** (0.025)	0.331*** (0.025)	219875	0.651
Fluency-Refugees	0.169*** (0.020)	0.219*** (0.018)	0.306*** (0.015)	34863	0.607
Difference	-0.025 (0.044)	0.011 (0.058)	0.056 (0.075)	258638	0.632
Controlling for Education					
Fluency-Non-Refugees	0.198*** (0.038)	0.252*** (0.048)	0.302*** (0.054)	492388	0.641
Fluency-Refugees	0.115*** (0.029)	0.172*** (0.026)	0.246*** (0.027)	26336	0.601
Difference	-0.083* (0.04)	-0.080 (0.052)	-0.056 (0.061)	518724	0.640
Controlling for Occupation					
Fluency –Non-Refugee	0.10*** (0.020)	0.15*** (0.025)	0.202*** (0.025)	219875	0.703
Fluency-Refugees	0.069*** (0.022)	0.099*** (0.017)	0.142*** (0.016)	34863	0.663
Differences	-0.037 (0.024)	-0.052* (0.029)	-0.059 (0.045)	258638	0.70

Notes: Specific regions include Eastern Europe, South East Asia, and the Middle East, South America, and the Caribbean, and Africa.

Regression controls include age, education, race, Hispanic status, whether attending school, native wage measure, network measure, and state and year dummies. ***, **, ** indicate significance at 1 percent, 5 percent, and 10 percent levels.

	Table 4: Regressions on Log Hourly Wages			Sample Size	R^2
Method/Period	0-5	6-12	13-20		
OLS -Eastern Europe					
Fluency-Non-Refugees	0.361*** (0.040)	0.361*** (0.052)	0.413*** (0.047)	50721	0.655
Fluency-Refugees	0.220*** (0.063)	0.293*** (0.056)	0.355*** (0.040)	5414	0.673
Difference	-0.141** (0.071)	-0.067 (0.032)	-0.057 (0.057)	56135	0.656
Fixed-Effects- Eastern Europe					
Fluency-Non-Refugees	0.318*** (0.034)	0.332*** (0.049)	0.384*** (0.049)	50721	0.655
Fluency-Refugees	0.223*** (0.020)	0.298*** (0.060)	0.344*** (0.036)	5414	0.673
Difference	-0.094 (0.070)	-0.033 (0.075)	-0.040 (0.054)	56135	0.662
OLS- South America and Caribbean					
Fluency-Non-Refugees	0.227*** (0.037)	0.231*** (0.030)	0.307*** (0.030)	84875	0.618
Fluency-Refugees	0.178 (0.006)	0.221 (0.007)	0.286*** (0.009)	14854	0.562
Difference	-0.051 (0.037)	-.010 (0.027)	-0.022 (0.031)	99729	0.605
Fixed-Effects- South America and Caribbean					
Fluency –Non-Refugee	0.230*** (0.039)	0.215*** (0.031)	0.279*** (0.029)	84875	0.618
Fluency-Refugees	0.173*** (0.006)	0.221*** (0.007)	0.288*** (0.010)	14854	0.562
Differences	-0.057 (0.039)	-0.005 (0.027)	-0.009** (0.028)	99729	0.609

Notes: Regression controls include age, education, race, Hispanic status, whether attending school, native wage measure, network measure, and state and year dummies.

***, **, ** indicate significance at 1 percent, 5 percent, and 10 percent levels.

	Table 5: Regressions on Log Hourly Wages			**Sample Size**	**R^2**
Method/Period	0-5	6-12	13-20		
OLS -South East Asia and Middle East					
Fluency-Non-Refugees	0.318*** (0.039)	0.383*** (0.032)	0.49*** (0.035)	50896	0.631
Fluency-Refugees	0.069 (0.049)	0.140*** (0.044)	0.231*** (0.032)	11136	0.635
Difference	-0.248*** (0.073)	-0.243*** (0.050)	-0.259*** (0.048)	62032	0.630
Fixed-Effects- South East Asia and Middle East					
Fluency-Non-Refugees	.277*** (0.028)	0.32*** (0.039)	0.417*** (0.048)	50896	0.638
Fluency-Refugees	0.101*** (0.045)	0.148*** (0.026)	0.244*** (0.026)	11136	0.636
Difference	-0.176 (0.056)	-0.172 (0.031)	-0.173 (0.063)	62032	0.636
OLS- Africa					
Fluency-Non-Refugees	0.192*** (0.040)	0.311*** (0.050)	0.419*** (0.044)	32822	0.620
Fluency-Refugees	0.097** (0.048)	0.122 (0.010)	0.315*** (0.096)	3459	0.605
Difference	-0.094 (0.071)	-0.189** (0.012)	-0.104** (0.010)	36281	0.633
Fixed-Effects- Africa					
Fluency –Non-Refugee	0.127 (0.021)	0.240 (0.046)	0.327*** (0.031)	32822	0.619
Fluency-Refugees	0.1119 (0.048)	0.106** (0.010)	0.317*** (0.092)	3459	0.639
Differences	-0.008 (0.058)	-0.134 (0.124)	-0.010 (0.096)	36281	0.622

FIGURE 1. FLUENCY PROFILES FOR THE POOLED SAMPLE

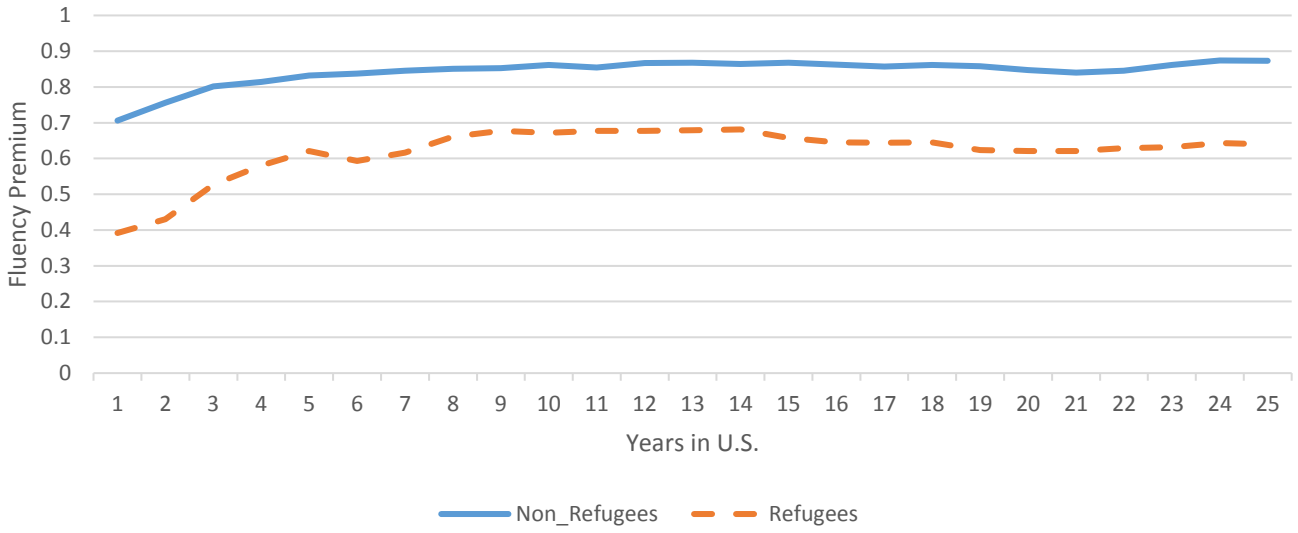
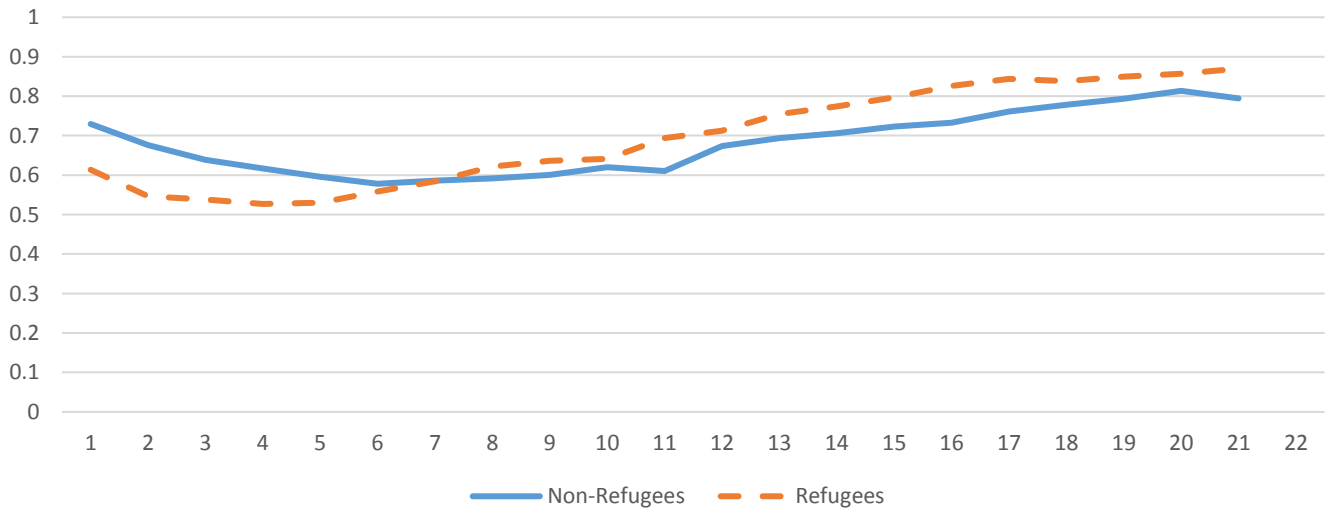


FIGURE 2. EDUCATION PROFILES OF THOSE FLUENT FOR THE POOLED SAMPLE



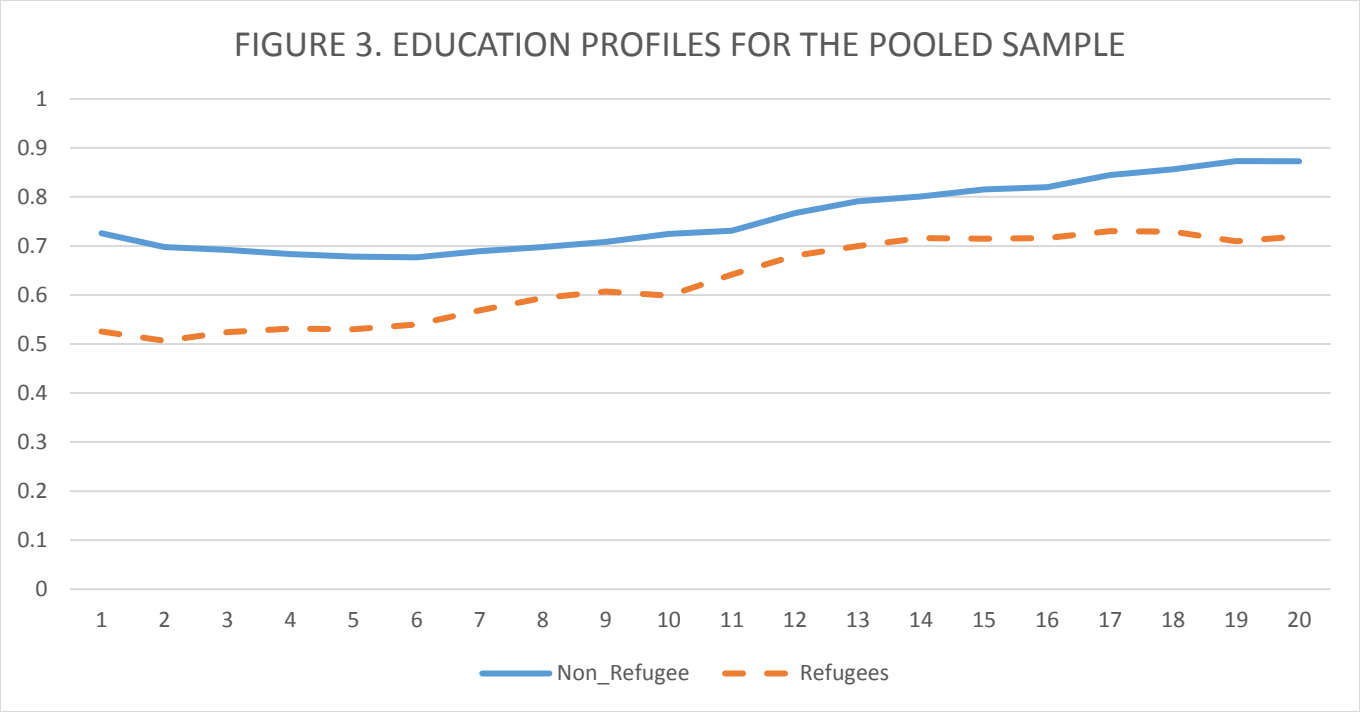


FIGURE 3. EDUCATION PROFILES FOR THE POOLED SAMPLE
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Non_Refugee
Refugees

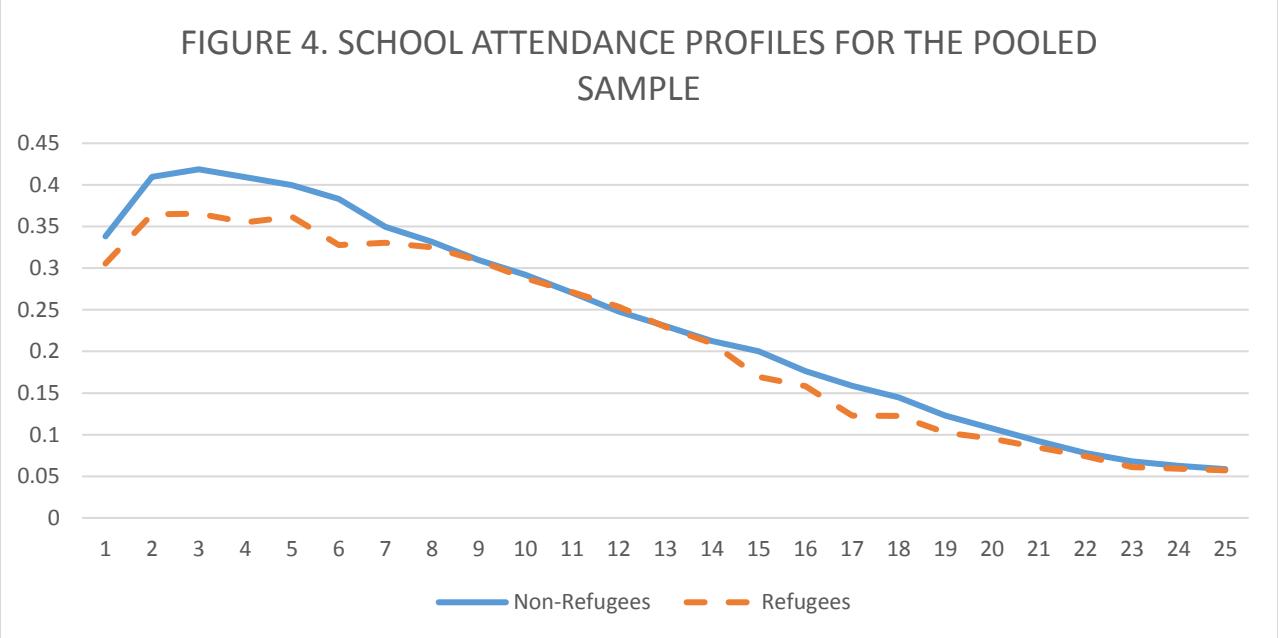
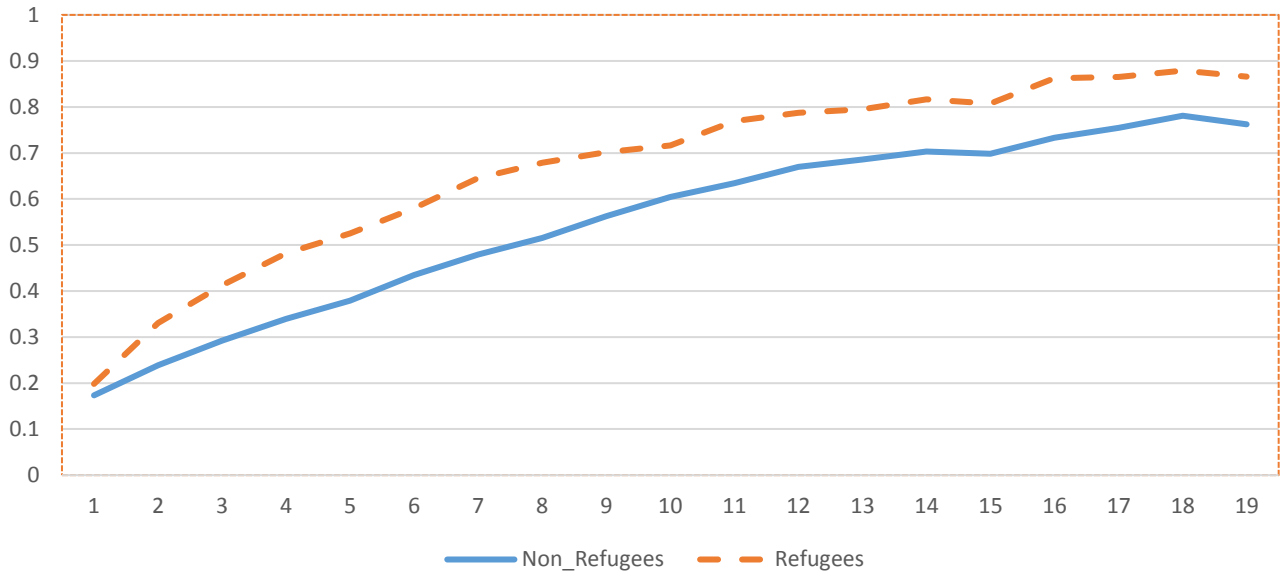


FIGURE 4. SCHOOL ATTENDANCE PROFILES FOR THE POOLED SAMPLE
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Non-Refugees
Refugees

FIGURE 5. CITIZENSHIP STATUS PROFILES FOR THE POOLED SAMPLE



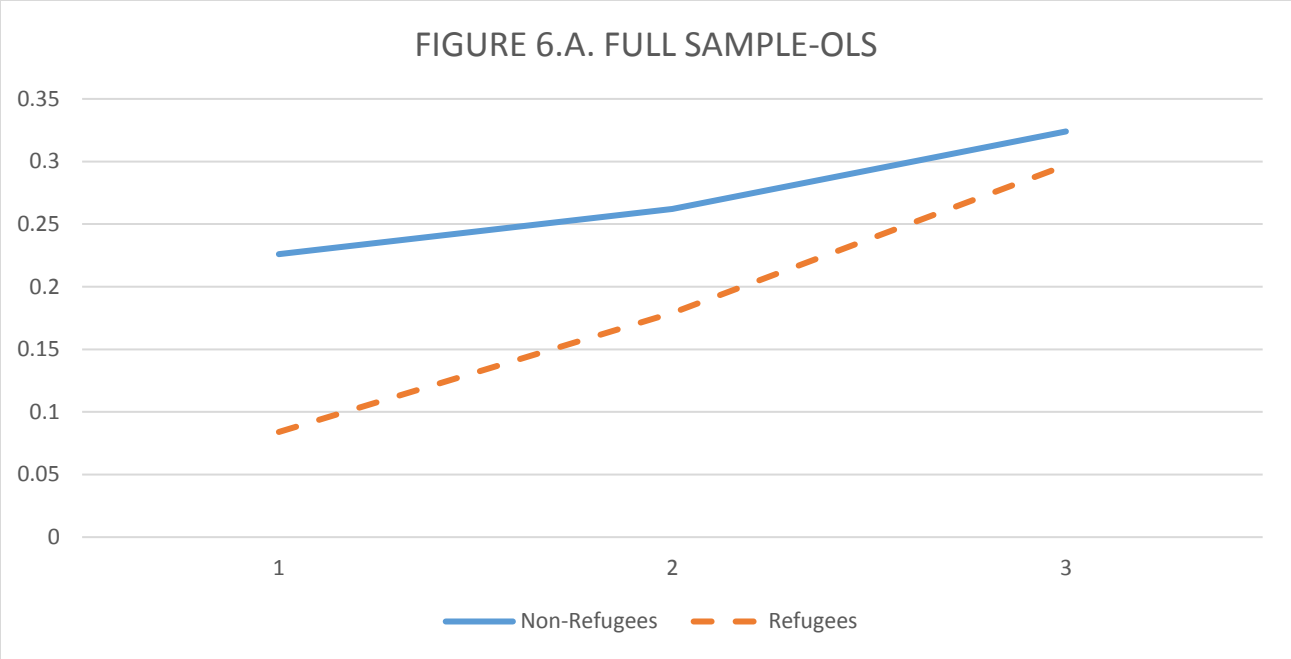


FIGURE 6.A. FULL SAMPLE-OLS
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non-Refugees
Refugees

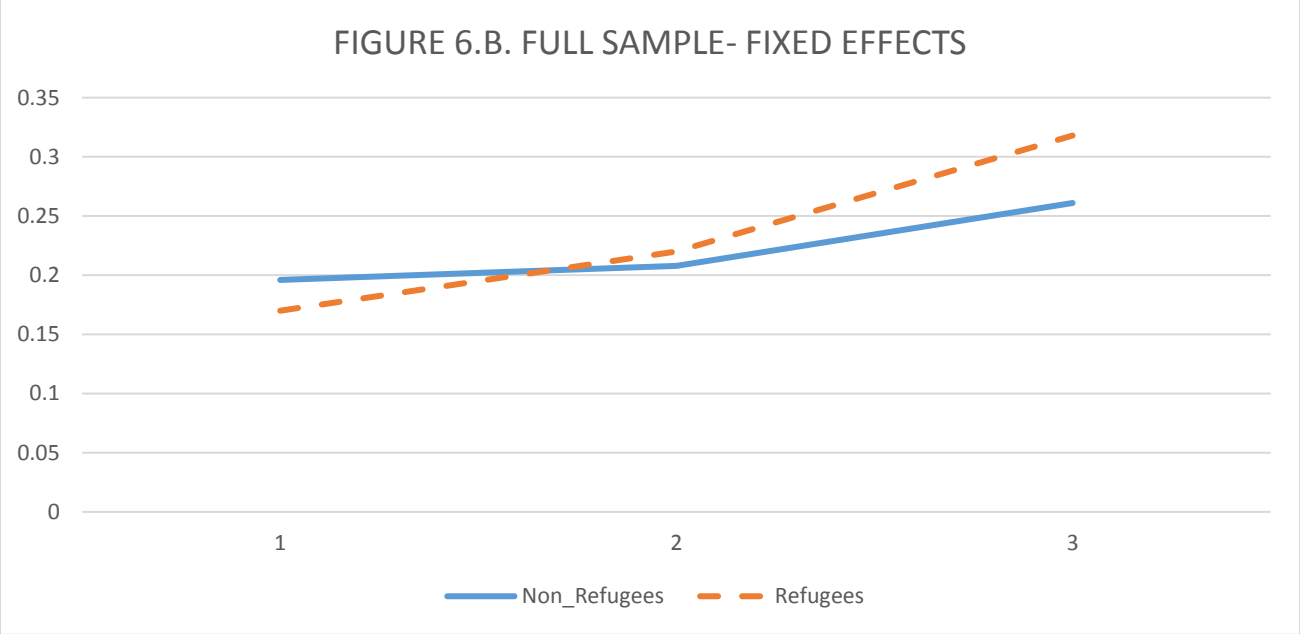


FIGURE 6.B. FULL SAMPLE- FIXED EFFECTS
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non_Refugees
Refugees

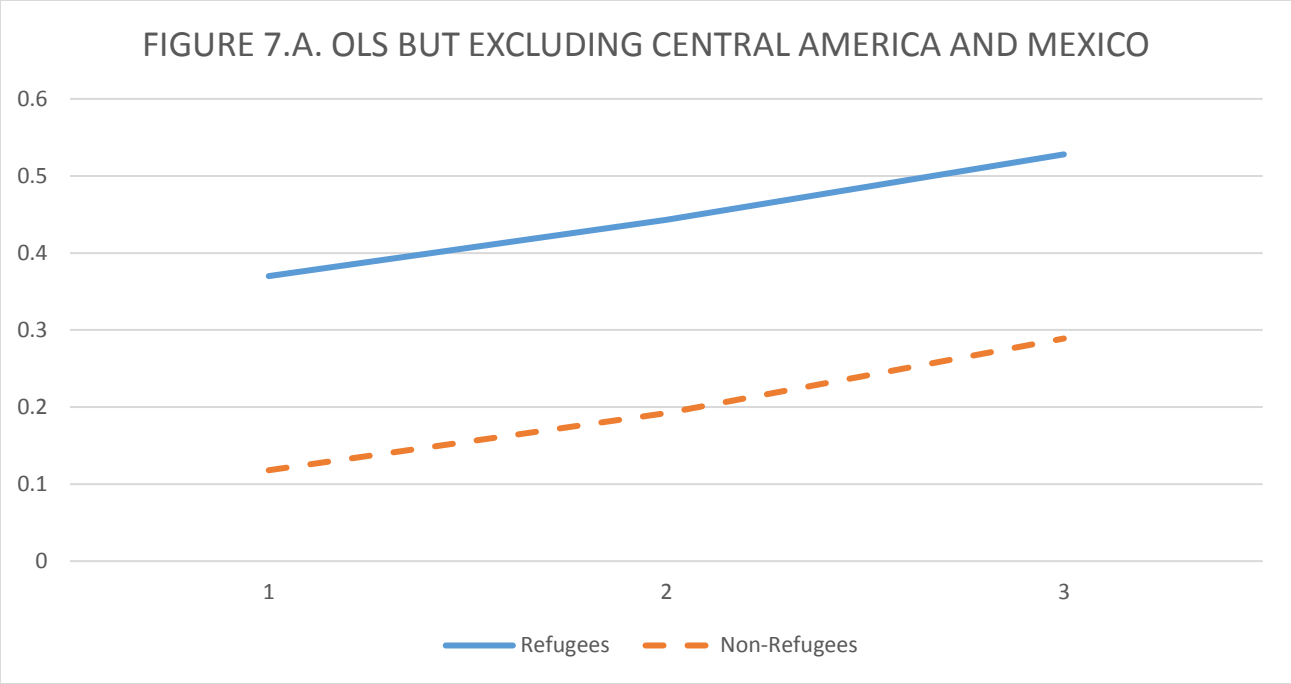


FIGURE 7.A. OLS BUT EXCLUDING CENTRAL AMERICA AND MEXICO
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
Refugees
Non-Refugees

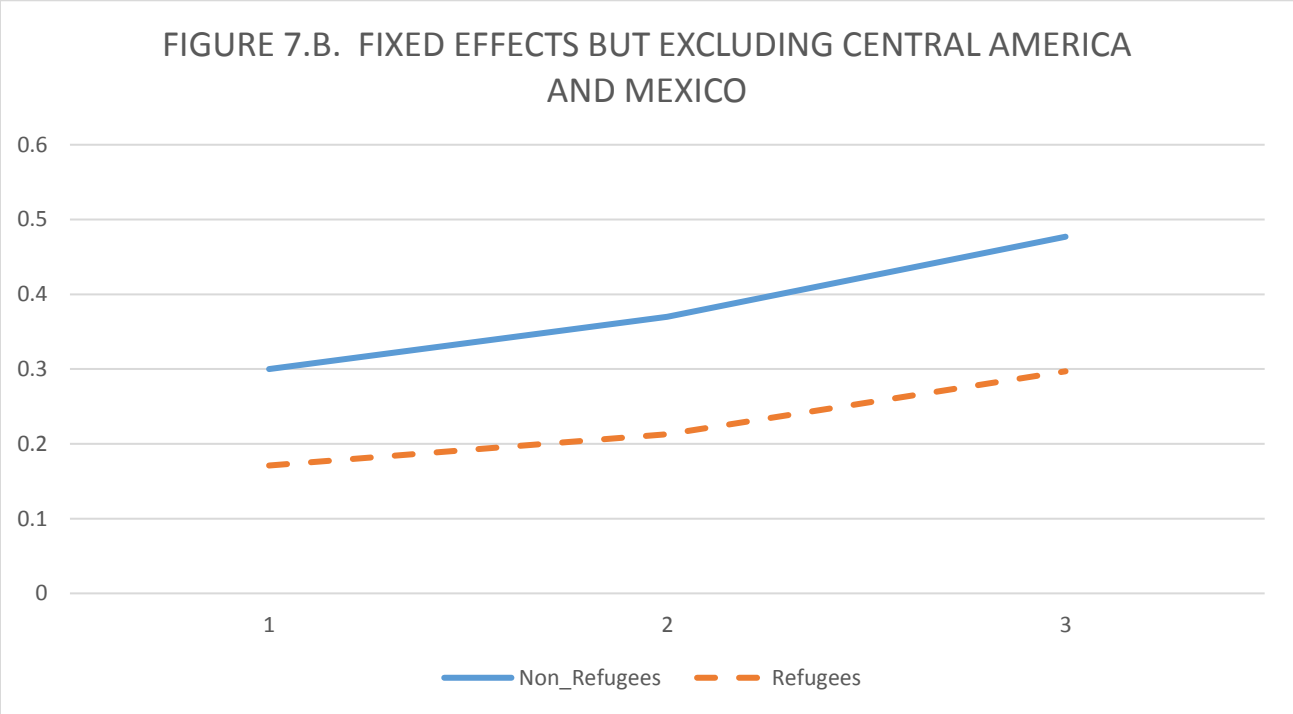


FIGURE 7.B. FIXED EFFECTS BUT EXCLUDING CENTRAL AMERICA AND MEXICO
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
Non_Refugees
Refugees

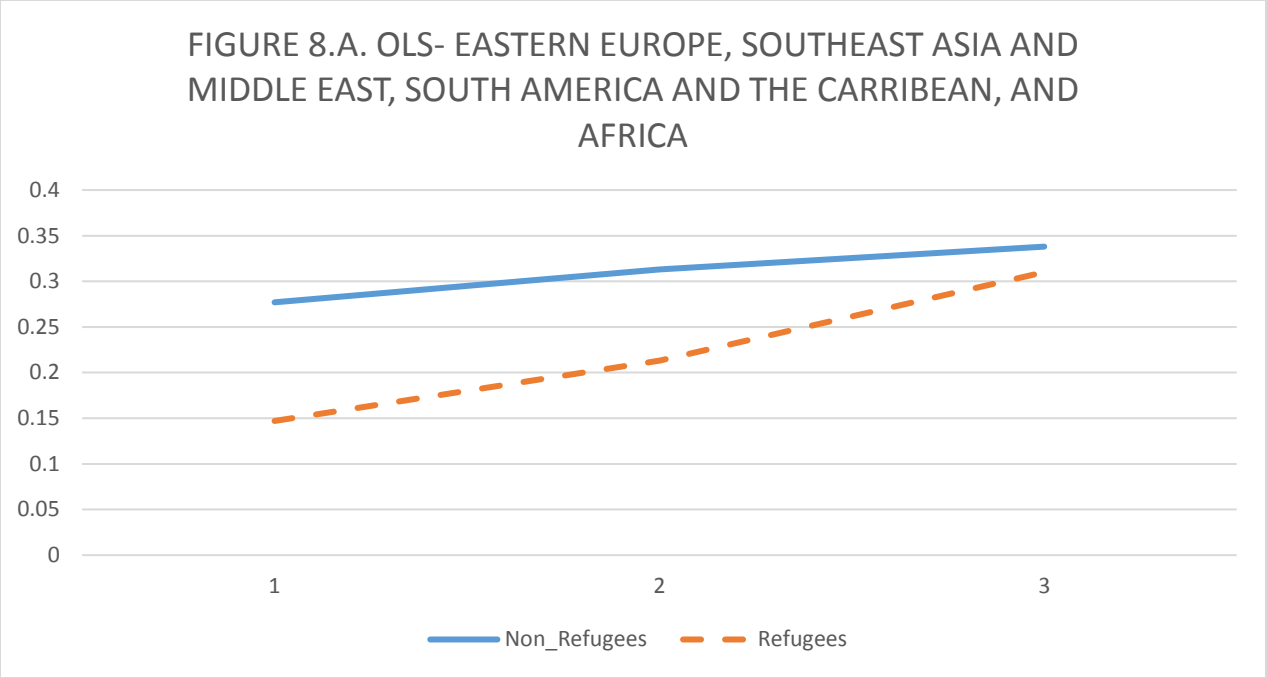


FIGURE 8.A. OLS- EASTERN EUROPE, SOUTHEAST ASIA AND MIDDLE EAST, SOUTH AMERICA AND THE CARRIBEAN, AND AFRICA
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non_Refugees
Refugees

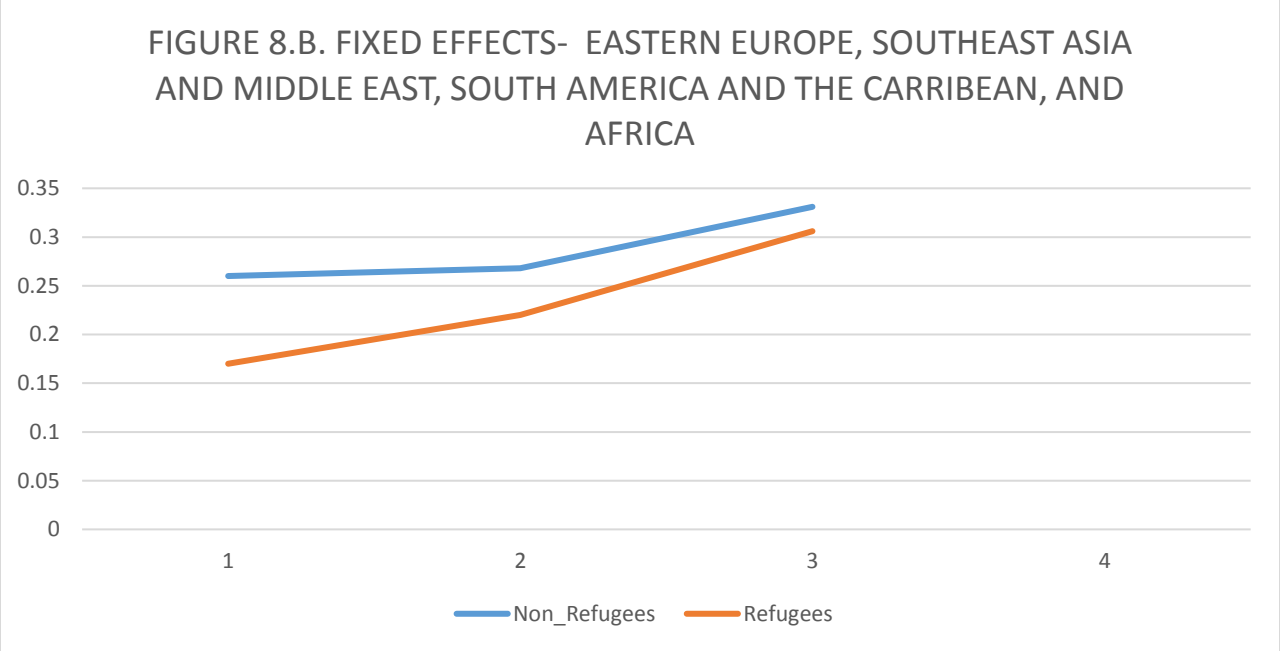


FIGURE 8.B. FIXED EFFECTS- EASTERN EUROPE, SOUTHEAST ASIA AND MIDDLE EAST, SOUTH AMERICA AND THE CARRIBEAN, AND AFRICA
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
Non_Refugees
Refugees

FIGURE 9. EASTERN EUROPE, SOUTHEAST ASIA AND MIDDLE EAST, SOUTH AMERICA AND THE CARRIBEAN, AND AFRICA-by_Education

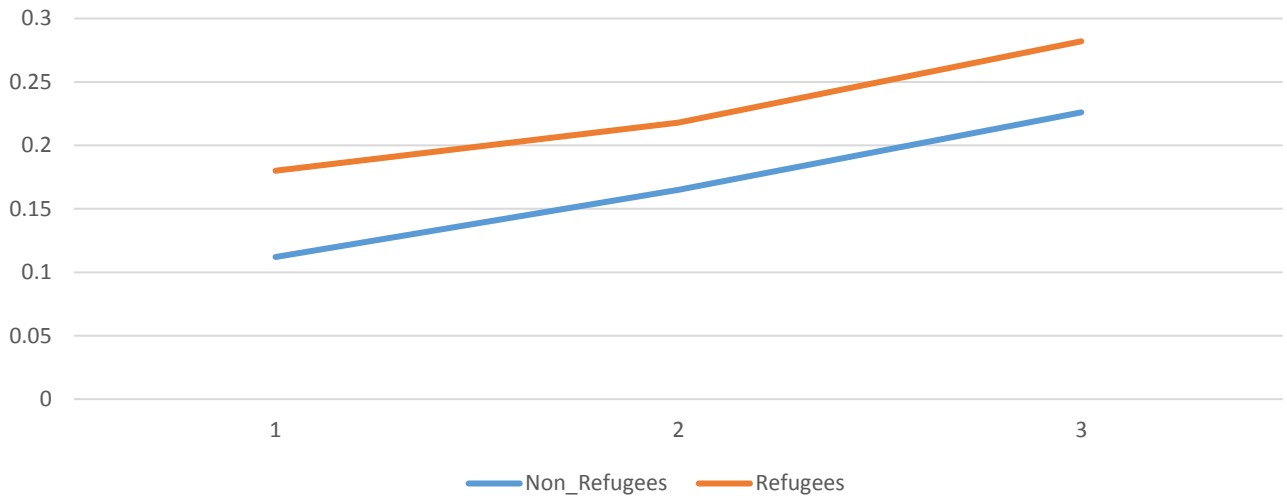
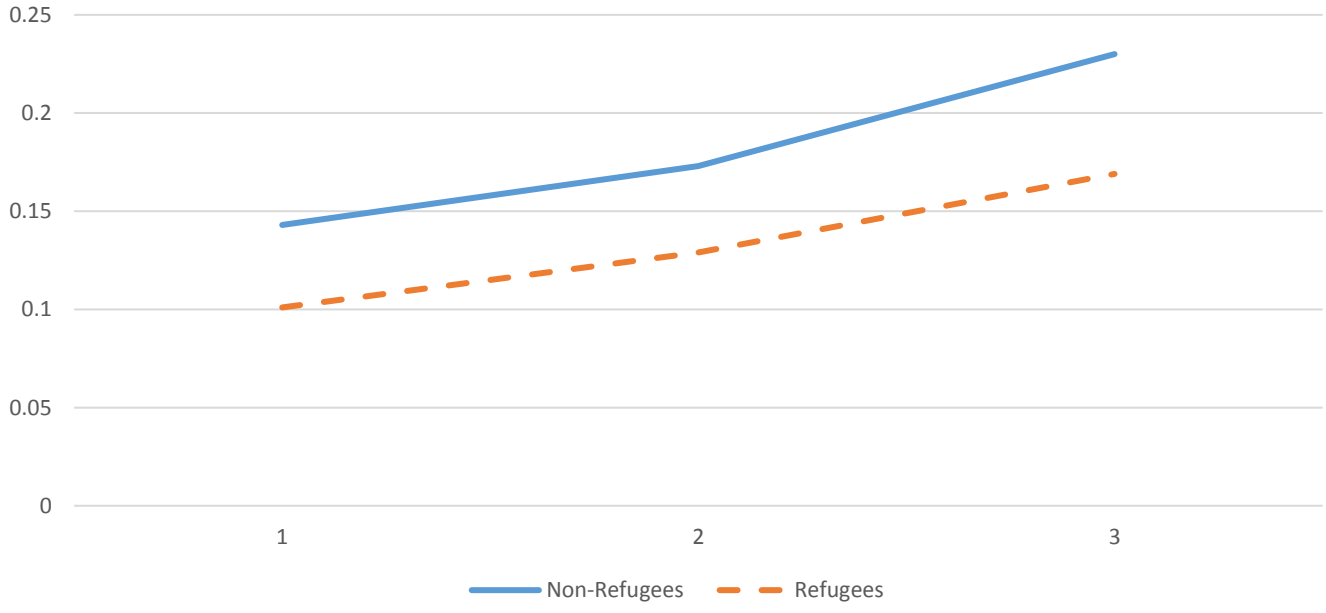


FIGURE 10. EASTERN EUROPE, SOUTHEAST ASIA AND MIDDLE EAST, SOUTH AMERICA AND THE CARRIBEAN, AND AFRICA AFTER CONTROLLING FOR OCCUPATION



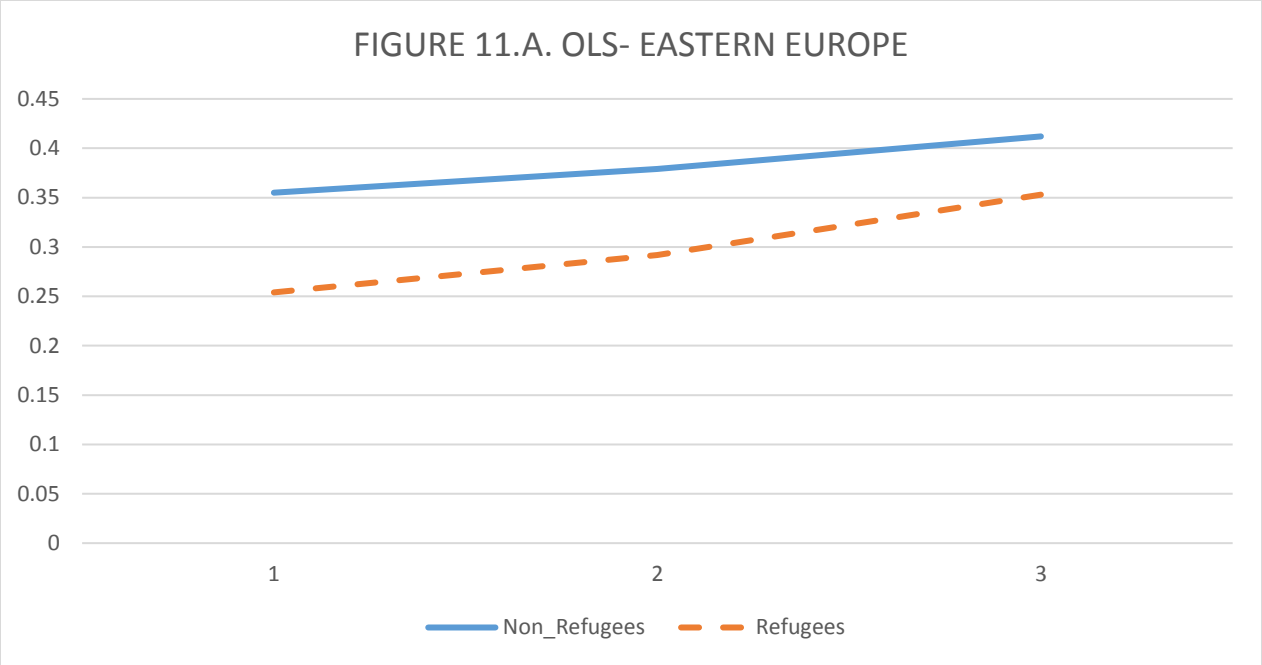


FIGURE 11.A. OLS- EASTERN EUROPE
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non_Refugees
Refugees

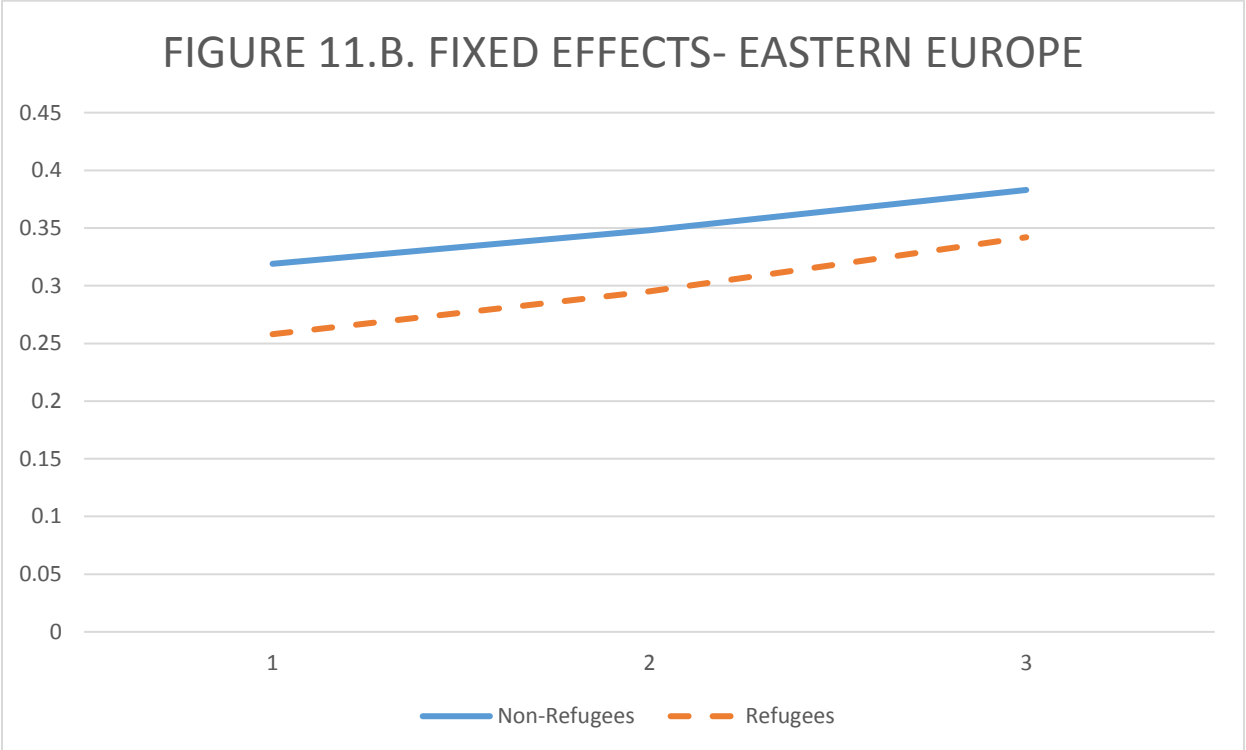


FIGURE 11.B. FIXED EFFECTS- EASTERN EUROPE
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non-Refugees
Refugees

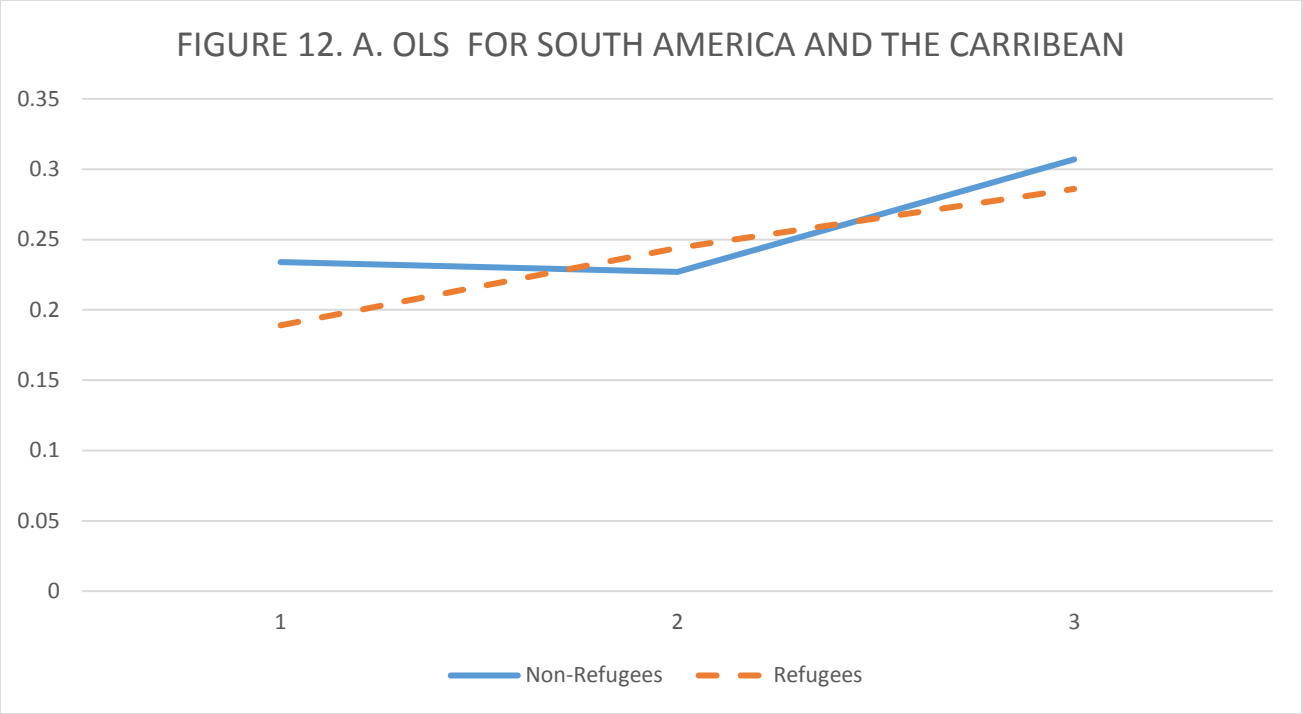


FIGURE 12. A. OLS FOR SOUTH AMERICA AND THE CARRIBEAN
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non-Refugees
Refugees

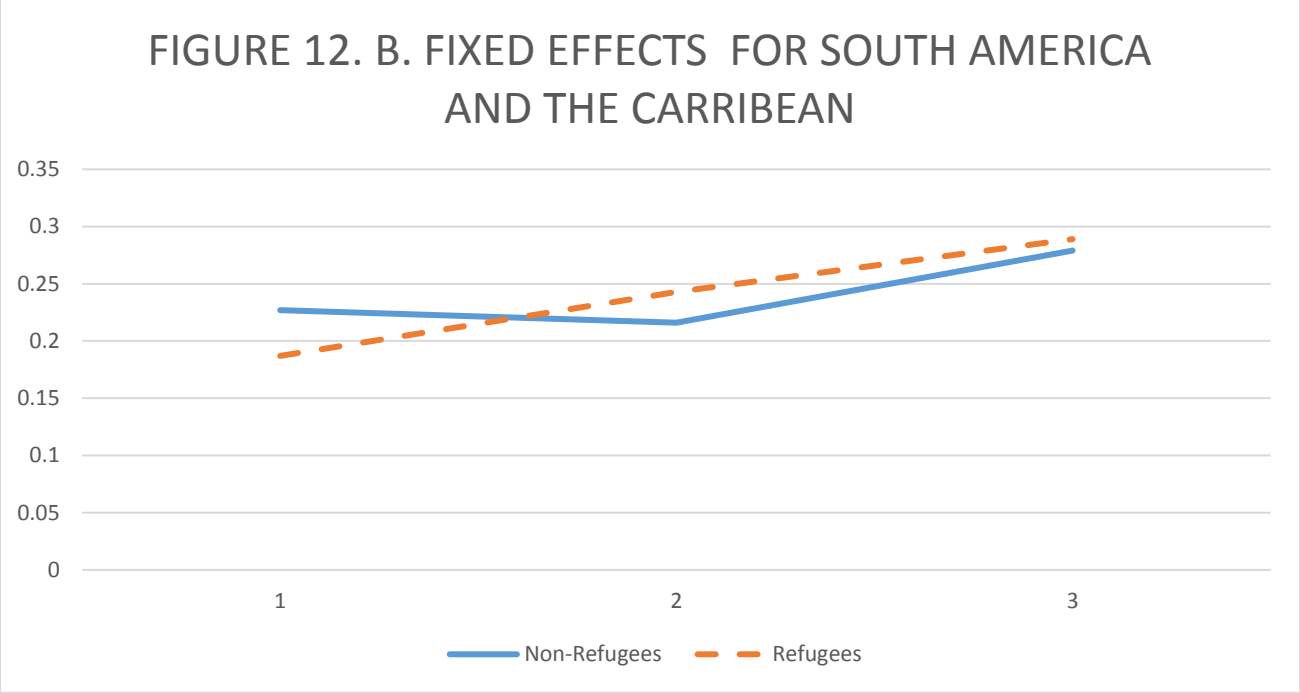


FIGURE 12. B. FIXED EFFECTS FOR SOUTH AMERICA AND THE CARRIBEAN
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non-Refugees
Refugees

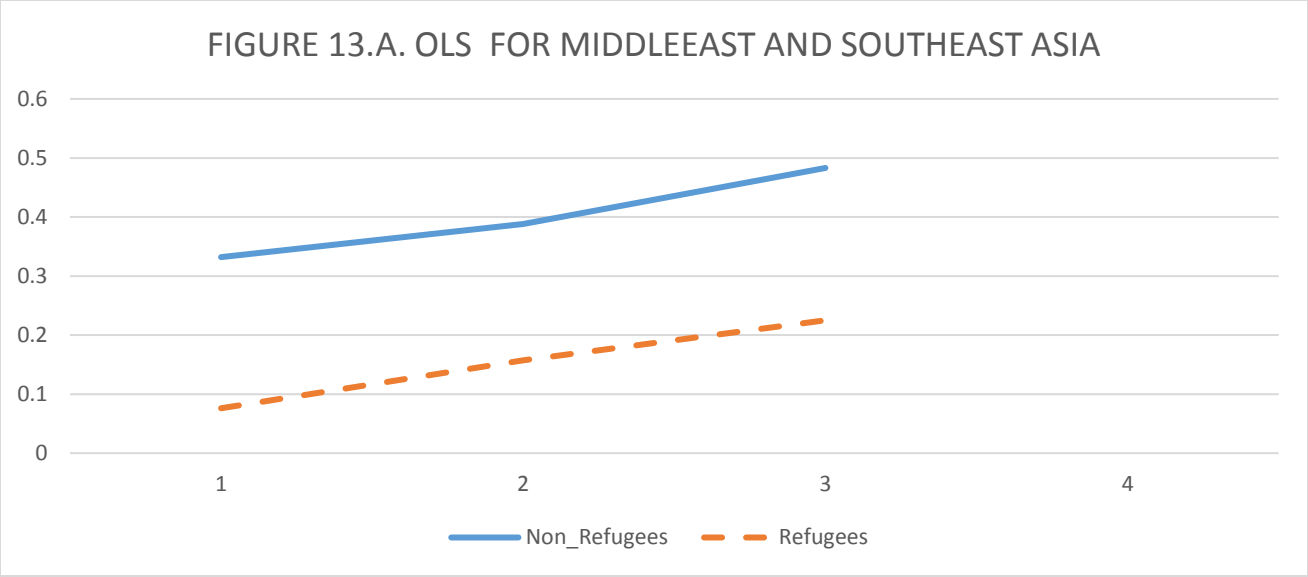


FIGURE 13.A. OLS FOR MIDDLEEAST AND SOUTHEAST ASIA
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
Non_Refugees
Refugees

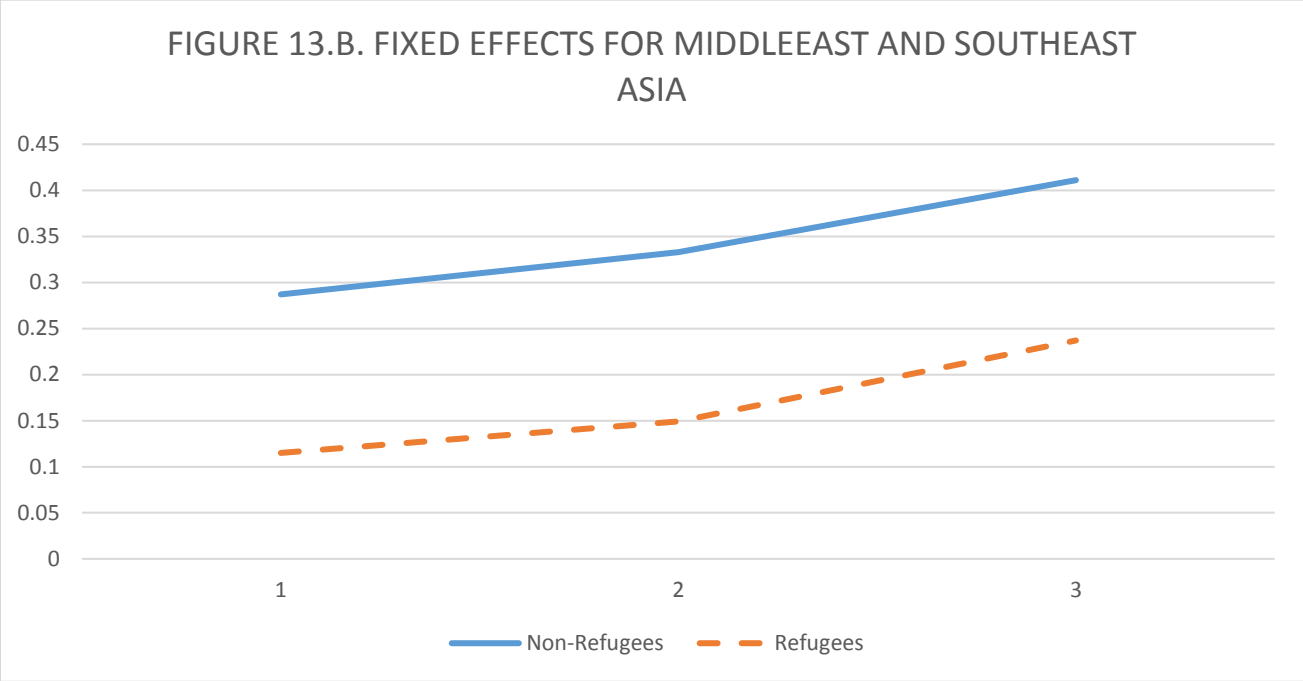


FIGURE 13.B. FIXED EFFECTS FOR MIDDLEEAST AND SOUTHEAST ASIA
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non-Refugees
Refugees

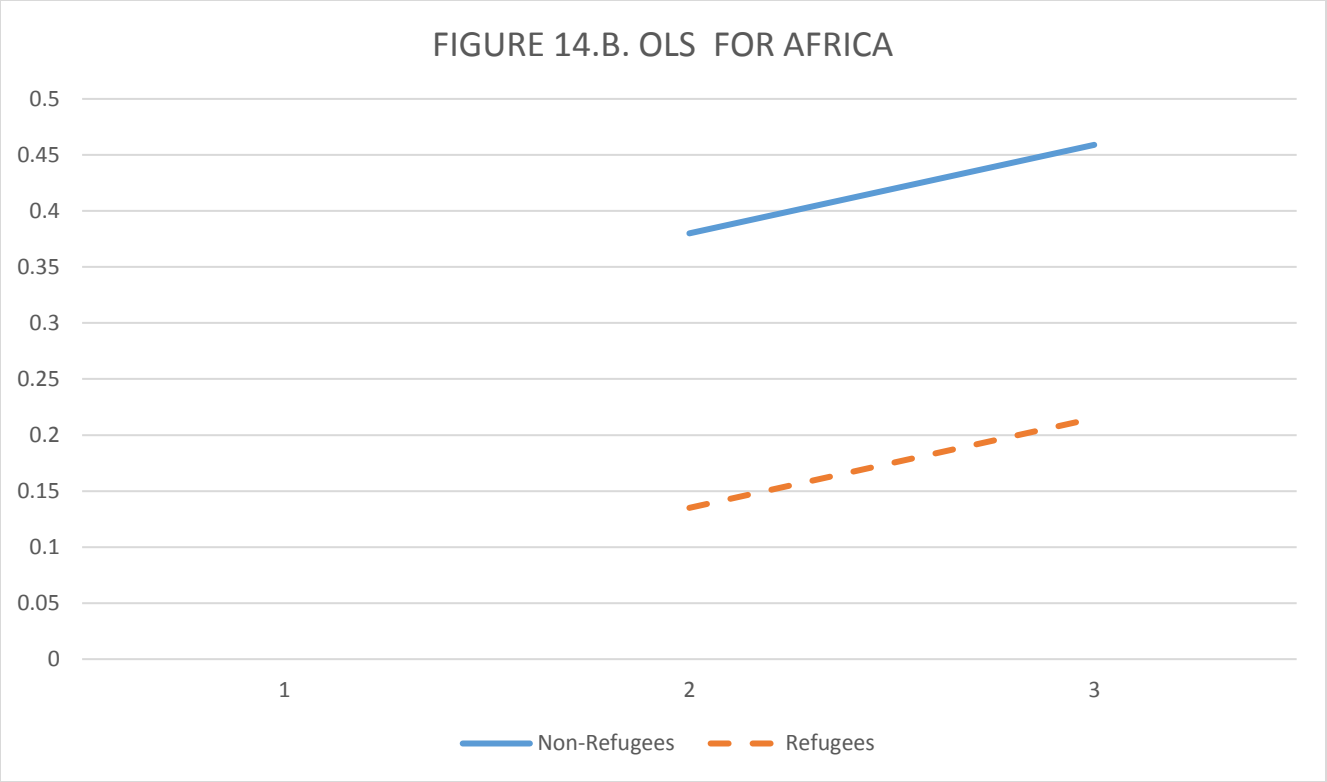


FIGURE 14.B. OLS FOR AFRICA
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non-Refugees
Refugees

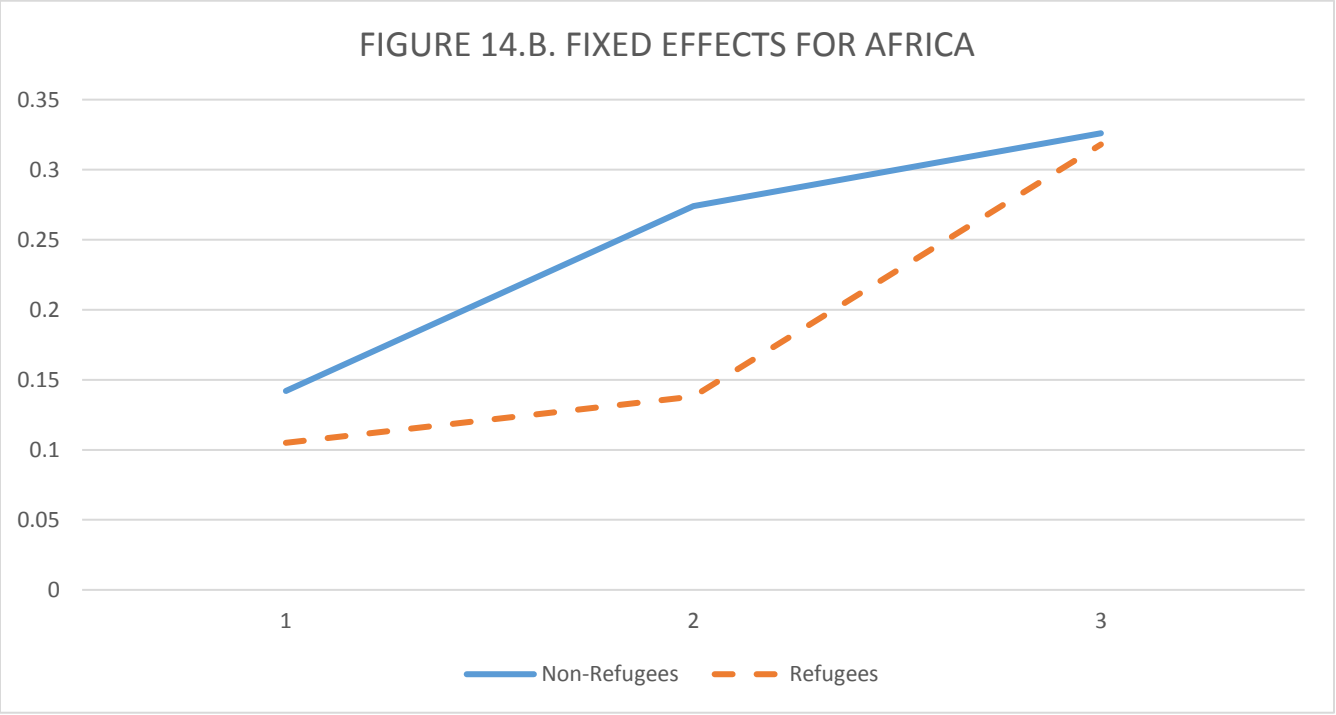


FIGURE 14.B. FIXED EFFECTS FOR AFRICA
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
Non-Refugees
Refugees

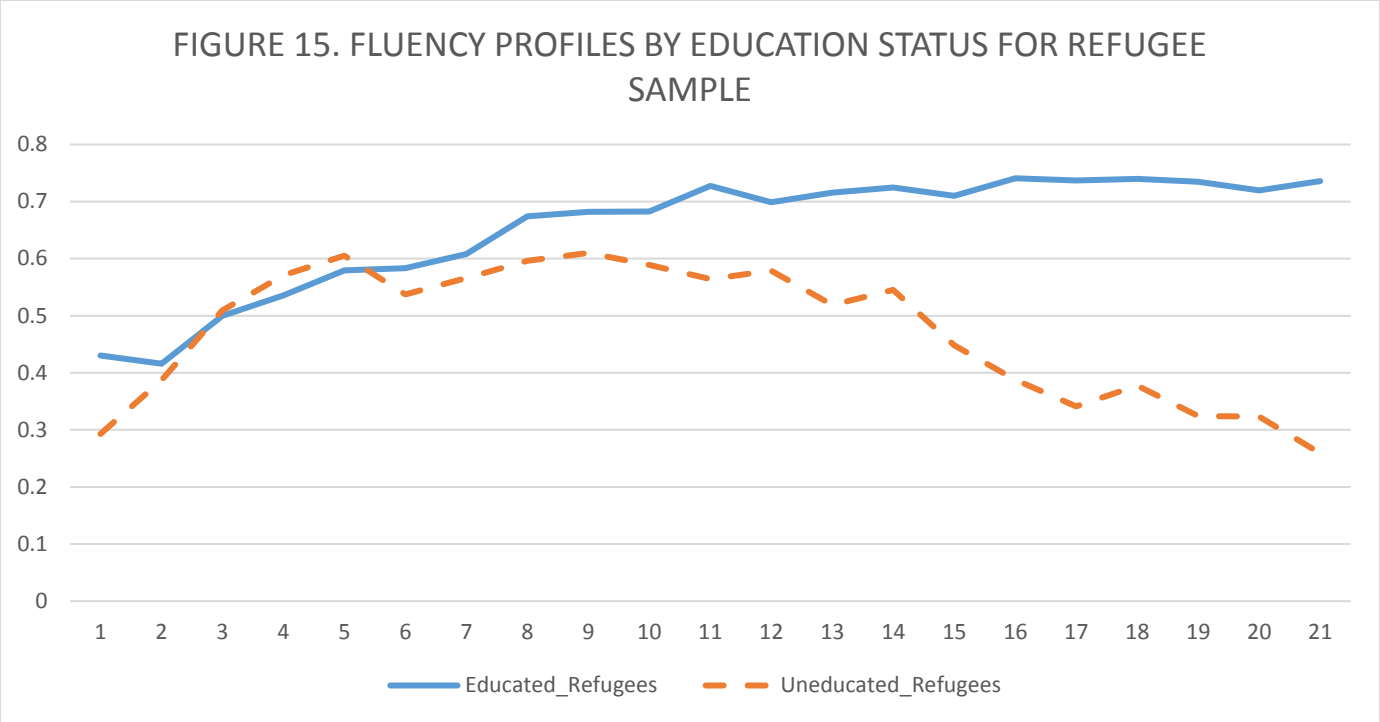


FIGURE 15. FLUENCY PROFILES BY EDUCATION STATUS FOR REFUGEE SAMPLE
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Educated_Refugees
Uneducated_Refugees