THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

An Empirical Investigation of Social Capital and International Migration: The Role of Language, Culture and Diasporas

NICHOLAS SHEA
SPRING 2021

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Economics
with honors in Economics

Reviewed and approved* by the following:

Bee-Yan Roberts
Professor of Economics and Asian Studies
Thesis Supervisor

Russell Chuderewicz
Professor of Economics
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

This thesis aims to obtain a better understanding of the role that social capital, language, culture and diaspora networks play in determining migration rates. Using data from the World Bank Global Bilateral Migration Database, and a new set of language variables constructed by Jacques Melitz and Farid Toubal (2014), I conduct a gravity analysis of bilateral migrant flows between 1960 and 2000. Overall, the results produce an image of the migration decision that is in some ways consistent with a traditional theoretical conception of migration as an investment decision, in some ways consistent with previous migration studies studying the role of diasporas, language and culture, but also poses new empirical questions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to sincerely thank Professor Roberts for consistently inspiring my curiosity and challenging me as a student. I am grateful to have had the opportunity to study development economics in her classes, and I could not have completed this thesis without her guidance as a professor and mentor. I would also like to thank Professor Tybout for guiding me through the thesis process. The insights that he provided were invaluable the development of this work.

Chapter 1

Introduction and Literature Review

Introduction

What is Social Capital?

In contemporary social science, the concept of "social capital" has become a subject of renewed research interest. Robert D. Putnam, a leading researcher on social capital, concisely defines the concept as, "social networks and the associated norms of reciprocity and trustworthiness" (Putnam, 2007, 137). This thesis will use this definition for its purposes. By this definition, social networks are effectively a type of economic capital, and like physical capital and human capital, social networks have value (Putnam, 2007). The value added by social capital is derived from the reciprocity and trustworthiness that healthy social networks produce. The literature on cooperation generally comes to the consensus that cooperation beyond the family or small group is not the default inclination of rational decision makers, and is therefore difficult to achieve on the societal scale (Collier, 2013). The provision of public goods and redistribution of resources to the less fortunate fundamentally rely on the ability of people to cooperate successfully on a societal level. This cooperation is difficult to attain and fragile to maintain, making it highly desirable and valuable (Collier, 2013). The reciprocity and trustworthiness gained from social networks are essential to overcoming these fundamental difficulties and

enabling cooperation on a large scale. In this way, social capital has tangible value, as it enables the desirable externalities that are made possible by cooperation (Collier, 2013; Putnam, 2007).

As Putnam's (1993) work in Italy has shown, social capital is a real and valuable resource that offers tangible benefits to societies. Like other forms of non-physical capital, such as human capital, social capital must be developed and maintained under the correct conditions. Whereas human capital is developed through endeavors that result in productive knowledge gain, social capital is developed through endeavors that result in gains in solidarity, or what Collier (2013) refers to as "mutual regard." Mutual regard refers to a blend between empathy, respect, and trust, and is a key prerequisite to the formation of successful social networks (Collier, 2013; Putnam, 2007). Mutual regard is reciprocal, meaning that an individual feels mutual regard for others in their social network more strongly when they trust that others in the social network feel the same mutual regard for them (Putnam, 2007). People typically have a stronger sense of mutual regard for others when they feel that they are "like themselves" and share some common identity or unifying characteristics (Collier, 2013; Putnam, 2007). The idea of a shared common identity is particularly important, and a shared cultural, ethnic, or historical identity can strongly reinforce a sense of mutual regard (Putnam, 2007).

Mutual regard offers two important benefits that characterize successful societies: willingness to finance transfers to the less fortunate, and willingness to cooperate to provide public goods (Collier, 2013, 57). Economists and social scientists generally understand that cooperation beyond the family or small group is not the default inclination of rational decision makers, and is therefore difficult to achieve on the societal scale (Collier, 2013). Mutual regard is a powerful force in overcoming these natural inclinations. If people have mutual regard for one another, have empathy for the less fortunate, and have faith that their own goodwill will be

reciprocated, they will be much more willing to contribute their own resources to enable cooperation. People will be more inclined to contribute their own resources to participation if they can reasonably presume that their actions will be reciprocated. This mutual expectation of reciprocity is founded on the trust and understanding that their social network is characterized by mutual regard (Collier 2013, 57). It is through a society's ability to cooperate and accomplish these two tasks that the main utility gained from social capital is derived. For these reasons, societies that are well endowed with mutual regard, tend to also be well endowed with social capital, and are in this way better off than those that are poorly endowed.

Diversity and Social Capital

Social capital is determined by mutual regard, which tends to be strengthened by homogeneity and common identity. Intuitively, operating under this understanding, one would expect an increase in ethnic or cultural diversity to have some effect on social capital. There are two general hypotheses in social science that predict what this effect will be: the contact hypothesis and the conflict hypothesis (Putnam, 2007). The contact hypothesis predicts that when members of two groups are brought into closer proximity to one another, they will become more comfortable with each other over time and levels of mutual trust will increase (Putnam, 2007). A famous study surveying American soldiers in WWII addressed the question of the contact hypothesis. In the study, white soldiers were asked whether they would be comfortable serving in the same platoons as African American soldiers. The results indicated that white soldiers that had previously served in mixed race platoons were much more comfortable with the idea, lending credibility to the contact hypothesis (Putnam, 2007). Alternatively, the conflict

hypothesis predicts that closer proximity between groups will lead to increased mutual suspicion and lower levels of mutual trust.

How Migration May Affect Social Capital

In a study involving a large American sample, Putnam investigated the effect of immigration on trust among community members. An unfortunate, yet significant, finding was that as the fraction of immigrants in a community increased, levels of mutual trust between immigrants and natives decreased (Collier, 2013, 74). Unfortunately, most empirical results on this issue also tend to come to similar conclusions (Putnam, 2007). Here, Putnam concisely articulates the main premise of conflict theory, "On this theory, the more we are brought into physical proximity with people of another race or ethnic background, the more we stick to 'our own' and the less we trust the 'other'." The following works, cited by Putnam, account for some of the most relevant research establishing empirical consensus on the prevalence of conflict theory (Blumer 1958; Blalock 1967; Giles & Evans 1986; Quillian 1995, 1996; Brewer & Brown 1998; Taylor 1998; Bobo 1999; Bobo & Tuan 2006) (Putnam, 2007, 142).

Putnam (2007) also found a new result that is even more unnerving. Not only did a higher fraction of immigrants decrease trust between groups, it also decreased trust *within* groups. In other words, not only did social networks fail to include immigrants, the presence of immigrants damaged the already existing social networks among natives (Collier, 2013). Putnam explains that this phenomenon, known as "hunkering down," likely occurs because, as a community becomes more heterogeneous, the sense of shared identity among members begins to deteriorate. As this identity deteriorates, people begin to lose faith that their community is characterized by

reciprocal mutual regard and, fearing their goodwill may not be reciprocated, become more hesitant to engage in social networks and provide their resources for common goals. The combination of decreases in both inter and intra group trust have a significant negative effect on social capital. In the short run, there is reason to believe that, through the combination of decreases in both inter and intra group trust**,** the increase in diversity that accompanies immigration could have an adverse effect on mutual regard and social capital in the host society. Fortunately, this effect is likely alleviated in the long run as migrants assimilate with the native population and a new encompassing identity forms (Putnam, 2007). The importance of building a new shared identity cannot be overstated. Returning to the case of the white soldiers in WWII that were comfortable serving in mixed race platoons, it was likely not simply that these soldiers had been exposed to diversity that caused them to be more comfortable, as contact theory would suggest. Instead, their comfort was likely due to the new, *encompassing identity* crafted by the unifying mission of the US Army, and the common struggle that all soldiers endured together (Putnam, 2007). Because assimilation and the formation of a common identity are needed to overcome these effects, any factors that may slow assimilation, such as cultural distance or linguistic distance, could plausibly exacerbate them. On the contrary, people who are culturally or linguistically closer to the host population may assimilate faster and more easily, and thus overcome these barriers more quickly (Collier, 2013).

How Social Capital May Affect Migration

Social networks not only provide significant external benefits to society, they also provide significant benefits to individuals that take part in them (Putnam, 2007). Research

indicates that ethnic and cultural homogeneity are conducive to the maintenance of higher levels of mutual regard and the formation of social networks (Putnam, 2007). This is true at the level of societies as a whole, and it is also true at the level of the individual. It will be easier for an individual to join and take advantage of social networks if they themself are ethnically or culturally similar to the group they are trying to join. If this is the case, it should be much easier for culturally similar migrants to adjust to life in the host country and begin realizing the gains of migration by finding a good job. Easier integration and settlement represent a decrease in the cost of migration, whereas easier integration into the labor market represents an increase in the returns to migration. Other things being equal, one could reasonably infer that migrants will want to move to a society in which they can more easily join social networks and take advantage of the associated economic benefits. Additionally, even without considering the economic benefits, migrants may simply prefer to move to a country that is similar to their home country because they enjoy living in a familiar environment. Because of both of these effects, it is plausible to presume that migrants will prefer to immigrate to countries that are culturally, ethnically, and linguistically similar to their own.

Goal of This Thesis and Intended Contribution

Mutual regard or solidarity is conducive to maintaining social capital, and prevails among populations that are homogenous: culturally, linguistically, religiously, or otherwise (Putnam, 2007). I hypothesize that, all other things being equal, migrants prefer destination countries that are culturally, linguistically, and religiously proximate to their own, due in part to the expected ease of social integration, in part to expected ease of labor market integration, and in part to

migrants' living preferences. The goal of this thesis is to test this hypothesis by conducting an empirical analysis investigating the relationship between the predictors of social capital and the rate of international migration.

When conducting empirical analyses of the predictors of international migration flows, the common approach is to estimate a gravity model (Collier & Hoeffler, 2014; Hatton & Williamson, 2005; Karemera, Oguledo, & Davis 2000; Mayda, 2007). Following from the previous literature, I will employ a similar strategy and estimate a gravity model explaining bilateral migrant flows. This study will analyze flows of migrants from low- and middle-income World Bank classification countries to Organization for Economic Cooperation and Development (OECD) countries in each decade between 1960 and 2000.

Drawing from previously completed work, this model will include a range of relevant push and pull predictor variables that are known to affect the migration decision and rates of migration. The novel contribution of this study is the inclusion of variables measuring the determinants of social capital. These take the form of bilateral indices that measure cultural, linguistic, and religious similarity between country pairs. Although these variables have been used frequently in other areas of literature, particularly in international trade, most of them have not been used for the purpose of explaining migration, and certainly not with the intention of understanding social capital's role in migration. Introducing these variables in a gravity analysis of migration will provide a better understanding of the role that social capital plays in the migration decision.

Why this Goal is Relevant and Interesting

Migration from poor countries to rich countries has increased significantly in the past several decades, and has become the defining trend in global migration (Collier, 2013). Because this type of migration is so prevalent, and appears to only be increasing, there has been much controversy over crafting policy to address it, as many experts disagree fundamentally on the consequences of this migration (Borjas, 2015; Clemens, 2011; Clemens, 2019; Collier, 2013). Regardless of the controversy, there is little doubt that this large-scale migration will play an important role in global development. In the past decade, prominent economists have argued that increased global migration could lead to potentially massive gains in global income, offering an unprecedented opportunity to alleviate global poverty (Clemens, 2011). Unfortunately, there are many impediments that either discourage migration in the first place, or diminish the returns to migration that does occur. For example, Alicía Adserà (2015) finds that language barriers cause migrants to be underemployed and earn sub-optimal wages, eroding the gains to migration. Adserà recommends that governments prioritize language learning programs for migrants to help realize these lost gains.

Though there are some short run impediments that result from increased diversity and diminished social capital, there is reason to believe that these can be overcome with intelligent policy based on a thorough understanding of the fundamental dynamics of migration (Collier 2013). It is important to better understand the nature of these impediments where they exist, and to use this understanding to design policy to alleviate them. There is significant reason to believe that migration is largely beneficial in the long run, but naturally, there are temporary costs and conflicts that result from a large inflow of heterogeneous people (Clemens, 2019; Putnam, 2007).

A central goal of migration research should be to better understand these impediments and conflicts, and do what is necessary to alleviate them. By doing this the costs of migration can be minimized, the benefits maximized, and a more prosperous world for future generations can be created.

The Role of Language in Migration

In previous gravity studies on international trade and migration, most authors used singular dummy variables for common official language to capture the effect of language (Collier & Hoeffler, 2014; Mayda, 2007; Melitz & Toubal, 2014; Karemera et al., 2000). However, this practice fails to account for the various causal pathways within the topic of language, and likely conflates multiple distinct and relevant effects. Melitz and Toubal (2014) address this historic oversimplicity in the international trade literature and construct a set of more nuanced variables that they believe better capture the distinct effects of language. Their chosen variable specifications are as follows: common spoken language (CSL), common native language (CNL), common official language (COL), and linguistic proximity of native languages (LP). For the purposes of studying international trade, the interpretation of the significance of these variables is as follows. If CSL is significant in the presence of the CNL it shows that ease of communication is important beyond ethnicity and trust. If COL is significant in the presence of CSL and CNL, it could show the importance of "institutionalized support for translation from a chosen language into the others that are spoken at home." If LP is significant in the presence of CSL, CNL, and LP, it could reflect the importance of "the ease of obtaining translations and interpreters when native languages differ and without any public support, and perhaps also the

influence of ethnic rapport between groups when their native languages differ." Melitz and Toubal (2014) argue that including these distinct measures in a gravity analysis of international trade is important in uncovering the most thorough understanding of how language influences trade. To date, no gravity analyses of migration have used this new set of variables. I will use this set of variables in my analysis to develop a more nuanced understanding of the role of language in the migration decision and how it may relate to social capital.

Previous Studies' Treatment of Language

In this paragraph I briefly summarize the methods and results related to language of recent gravity studies of migration on which this paper is based. Mayda (2007) builds on previous gravity studies and empirically investigates the determinants of migration flows into fourteen OECD countries between 1980 and 1995. Her study generates interesting results for the effects of average income, income inequality, demographic factors, and changes in destination countries' policies. Like most studies of this type, Mayda's includes a variable to account for the effect of language. This variable is specified as a dummy variable equal to one if a common language is spoken in the sending country and host country (Mayda, 2007). She finds that, although the effect of common language is positive as expected, it is not statistically significant. As described above, this single dummy variable approach may fail to capture the distinct effects within the concept of language, which may explain the surprising result of a lack of significance. Collier and Hoeffler (2014) conduct an empirical investigation building on the work of Mayda. They estimate a gravity model for the determinants of migration flows from low- and middle-income countries into high income OECD countries between 1960 and 2000, focusing on the

effects of cultural distance, and the size of the diaspora. To account for the effect of language, the authors use a measure of linguistic proximity of native language derived from a world language tree compiled by Fearon (2003). They find that linguistic distance does significantly impede migration. Collier and Hoeffler interpret this result as showing the effect of logistical challenges, as linguistic distance may impede information flows, and migrants may expect to have lower earning potential if they lack the linguistic skills necessary for a higher paying job. However, the authors are omitting other language variables such as common spoken language, which one would expect to have a greater logistical effect on migration. This thesis will include a more complete and encompassing set of language variables. By doing so, it will be possible to check the robustness of previous results and develop a better understanding of the role of language in migration.

How Language Affects Migration

There are several ways in which language likely influences the migration decision. For the purposes of this thesis, I am interested in two main categories of effects, ones that are driven by social capital, and ones that are driven by logistical challenges or other factors. In terms of social capital effects, recall that language, particularly native language, is an important determinant of cultural or ethnic identity. Because they share a native language with a destination country, a migrant might reasonably expect that they will more easily be able to forge a common identity with the local population and benefit from inclusion in social networks. In this way, one would expect CNL to be a social capital variable with a positive effect on migration. Alternatively, there are many ways in which language can have an effect on migration

for practical reasons. For instance, suppose migrants do not share a native language with a host country, but are proficient in an acquired language spoken in the host country. Though the sense of common identity may not be as strong as if they shared a native language, these migrants could reasonably expect that being familiar with a spoken language in the host country would be beneficial for them for practical reasons. A common spoken language would likely enable migrants to earn higher wages, more easily navigate the migration process, and have better access to information, reducing uncertainty. In this way, I expect CSL to be a primarily logistical variable with a positive effect on migration. It would be difficult to gain much insight on the relative importance of these two types of language effects from Collier and Hoeffler (2014) or Mayda's (2007) results. Because both authors included only one variable for the effect of language, any and all effects from either of these categories would be captured in a singular coefficient, making it impossible to analyze their relative importance. The variables derived by Melitz and Toubal (2014) offer the ability to measure the effects of the different fundamental attributes of language. In the paragraphs below, I describe how I expect each of the language variables to affect migration, and which category these effects fall into.

Expected Effects of Language Variables

I expect that the variable that has the strongest effect on social capital and mutual regard is CNL. Sharing a native language represents the cultural and ethnic component of language, and offers the strongest source of common identity of all of the language variables. One would expect that CNL would significantly bolster mutual regard and social capital. Significance of CNL in the presence of the other language variables may provide evidence that the ethnic component of language and its effect on social capital is important to the migration decision. When a native language is not shared, LP offers a measure of closeness in countries' native

languages. I expect that groups having similar native languages would be naturally inclined to have higher mutual regard for one another than those with distant native languages. If LP is significant in the presence of the other language variables, this could provide further evidence that the ethnic component of language and its effect on social capital are important to the migration decision. It is worth noting that significance of either of these variables may also be due to the preferences of migrants and not necessarily be fully caused by effects of social capital.

While CNL and LP likely capture the linguistic components of ethnicity and trust and the corresponding effects of social capital, CSL and COL likely capture mainly logistical challenges. Once the ethnic/trust components of CNL and LP are controlled for, the coefficient on CSL will only capture the ways in which a shared spoken, non-native language affects the migration decision. Of particular importance is the increased availability of information for migrants, which reduces the uncertainty of migration. This has been shown to be an important factor in the migration decision, as migrants are much more willing to migrate when they feel more informed of likely outcomes (Banerjee & Duflo, 2019). CSL also reduces logistical difficulties, making travel and navigation of the migration process less difficult and costly. Migrants likely expect to have better labor market outcomes when they speak the language of the destination country, so CSL also may increase the incentive to migrate in this way. Although the primary effects of CSL are likely logistical in nature, it is possible that it could still affect social capital. When the native language of migrants is distant, it is possible that sharing common spoken language could have positive effects on mutual regard and the formation of social networks. It is then possible that migrants could expect that their social capital outcomes could be better in a country with a shared CSL, and that this could impact the migration decision. However, I expect that CSL

influences migration primarily via logistical effects, while CNL and LP influences migration primarily through social capital effects. The forthcoming empirical analysis sheds additional light on the relative importance of each of these variables, and will allow for more sophisticated inference to be drawn on potential causal pathways of language on the migration decision.

The Role of Culture in Migration

Previous Studies' Treatment of Culture Variables

Many gravity studies modeling migration flows, including the two on which this thesis directly builds on, have accounted for cultural factors in some way (Collier & Hoeffler, 2014; Mayda, 2007). In addition to the common language variable, Mayda includes a "cultural variable" in the form of the former colonial relationship between countries in her regression. Country pairs that were previously engaged in a colony-colonizer relationship can reasonably be expected to have influenced one another culturally, and developed some form of cultural rapport. The expected effect of a former colonial relationship on migration flow would be positive. Surprisingly, Mayda finds that this variable was not significant. While this result is interesting, it cannot be taken as evidence that cultural variables are not important to the migration decision. Notably, aside from common language, Mayda's analysis lacks a measurement of cultural distance between origin and destination countries beyond the colonial link. This thesis introduces a holistic measure of cultural distance to better understand the role that culture plays in the migration decision.

Collier and Hoeffler (2014) also include a former colonial relationship variable. They find that the effect of this variable is only weakly significant and is quantitatively minor, a similar result to Mayda. However, Collier and Hoeffler investigate the role of culture more deeply, by including a measure of genetic distance pioneered by Spolaore and Wacziarg (2009, 2013) as a proxy for cultural distance. This measure captures the length of time since two populations became separated from each other, and can be used as a measure of relatedness of populations. While researching the adoption of new technologies, Spolaore and Wacziarg (2009) have shown that their cultural distance measure can be a barrier to the knowledge flows between cultures. Collier and Hoeffler suggest that this same knowledge barrier could be an impediment to migration flows. Interestingly, the authors find that genetic distance has no significant direct effect on the rate of migration, and that "Migration from poor countries to rich ones appears not to be impeded by the gap in cultures" (2014, 23). A reader examining Collier and Hoeffler's methods will likely find it surprising that the authors chose to use the abstract concept of genetic distance as their proxy of cultural distance, especially given the focus of their paper on cultural effects. In fact, the authors themselves acknowledge this, stating, "The contrast between the effects of cultural and linguistic difference is striking. Based on the regression presented in column 5, a one standard deviation increase in cultural distance increases the migration flow by 27 percent, whereas a one standard deviation increase in linguistic distance reduces the migration flow by 16 percent. Clearly, these distinctive effects warrant further research, including complementary studies at the micro-level." (Collier & Hoeffler, 2014, 28). This thesis uses a more conventional measure of cultural distance. If my results mirror Colliers, this would provide further evidence for his interpretation of the role of cultural distance on migration, and may suggest that genetic distance is a reasonable proxy for cultural distance in studies of migration.

How Culture Affects Migration

This thesis is primarily concerned with with three relevant effects of cultural distance between the origin and destination countries on the migration decision: those driven by logistical factors, those driven by migrants' taste, and those driven by migrants' expectations relating to social capital outcomes. The main logistical factor, which Collier and Hoeffler addressed in their paper, is the barrier to knowledge transfer that may be caused by cultural distance (2014). As mentioned earler, they found this effect to be insignificant when measured by genetic distance. An important goal of this thesis is to evaluate whether replacing Collier's genetic distance measure with a more conventional measure of cultural distance affects these results.

Another way that cultural distance could affect the migration decision is through migrant tastes. Aw, Lee, and Vandenbussche (2019) investigate the role of consumer taste in shaping demand for international exports using a control function composed of several relevant bilateral variables. One of these variables is a bilateral index of cultural distance. Aw et al. assert that culture is an important component of taste, and that it is a greater determinant of similarities in taste than physical distance. For instance, despite Australia being physically closer to China, and despite vegemite, Australian food tastes are clearly closer to American food tastes than Chinese food tastes. Following this logic, culturally determined food tastes likely influence demand for exports between countries. Similarly, culturally determined tastes may also influence migrants' preference of destination country. Just as people's consumption preferences depend heavily on their culturally determined tastes, migrants' preferences of destination country may depend on their culturally determined tastes. Migrants may simply prefer to live in an environment that offers a similar lifestyle to what they are previously familiar with.

The last expected effect is related to social capital. Migrants may think that it will be more difficult for them to integrate into beneficial social networks in countries culturally distant from their own, which could influence the migration decision. Although this section only addresses ways that cultural distance directly affects the migration decision, there are additional ways in which cultural distance likely affects the migration flows through its interaction with the diaspora, which will be discussed in the next section.

My Cultural Distance Variable Specification

Unlike previous studies like Mayda (2007) which include no direct measurement of cultural distance, or studies like Collier and Hoeffler (2014) which use an unconventional measure of cultural distance, this thesis will use a well-established measure of cultural distance frequently used in the gravity literature. In 1980, Geert Hofstede pioneered the measurement of national cultures and their effect on work related values. Hofstede identified 6 relevant dimensions of every culture: power distance, individualism versus collectivism, uncertainty avoidance, masculinity versus femininity, long versus short term orientation, and indulgence versus restraint (Hofstede, 1980). Kogut and Singh (1988) developed a method of turning Hofstede's measures into a single bilateral index of cultural distance, which has been commonly used in the gravity literature. Aw et al. (2019) uses this methodology to construct a bilateral index of cultural distance for use in their study. I will use the same bilateral index of cultural distance used by Aw et al. (2019) as my cultural distance variable.

The Role of the Diaspora in Migration

How the Diaspora Affects Migration

As social capital theory suggests (Putnam, 2007), the ability to take advantage of social networks in the host country is beneficial for migrants. Factors that increase access to social networks should tend to lower the cost of migration and positively affect migration rates. The diaspora, or stock of migrants in a host country sharing a country of origin, provides migrants with a social network that may play a key role in determining migration rates. Beine, Docquier, and Ozden (2011) investigate the effect of diaspora networks on international migration rates. They argue that diasporas mitigate migration costs by helping migrants navigate linguistic barriers, new social norms, and a new job market, as well as by lowering legal barriers through family reunification programs (Beine et al., 2011). Beine et al. find empirical evidence that larger diaspora networks have a strong, positive effect on the size of international migration flows. This result solidifies the intuition that the availability of a robust social network in the host country should make migration less costly and positively influence migration rates.

Collier and Hoeffler (2014) draw similar conclusions to Beine et al. in that migration costs are lowered by support from diaspora networks, and that this has a powerful effect on migration rates. In addition to affirming this basic relationship, Collier and Hoeffler find that the diaspora affects migration rates through several important interactions. The authors find the interaction between diaspora size and income per capita in the country of origin to be significant and negative. Their interpretation is that migrants in lower income countries are more constrained by the investment costs of migration, and that support from a larger diaspora

substitutes for income in overcoming the investment costs. They find the interaction between physical distance and the diaspora to have a significant positive effect on migration rates. Their interpretation is that greater distance directly increases the costs of migration, and that as distance increases, the support of the diaspora becomes more important in offsetting these costs. They also find a significant negative interaction between the diaspora and a former colonial relationship. Their interpretation is that a former colonial relationship substitutes for the information flow provided by diasporas, making a large diaspora network less important when a former colonial relationship is present. Overall, Collier and Hoeffler find strong evidence that the diaspora plays an important role in offsetting migration costs and driving migration.

The most interesting effect of the diaspora that Collier and Hoeffler (2014) investigate is the effect of the interaction between the diaspora and cultural distance. Though one would expect cultural distance to be an impediment to migration, Collier and Hoeffler found that cultural distance alone did not have a significant effect on migration rates. Interestingly, Collier and Hoeffler instead find the interaction between cultural distance and the size of the diaspora to have a significant and positive effect on migration. This result suggests that a larger diaspora attracts more new migration when its members are culturally distant from the host population. Their interpretation of this result is complex but impactful, and builds on the theory discussed extensively in *Exodus: How Migration is Changing Our World* (Collier, 2013, 27). Collier and Hoeffler point out that as the diaspora accumulates, migration costs are lowered and migration increases. Migration increases the size of the diaspora, while the assimilation of migrants into mainstream society decreases the size of the diaspora. The authors postulate that a wider cultural distance causes migrants to develop social ties to natives more slowly, leading to a slower rate of assimilation, leading to a faster accumulation of the diaspora, ultimately leading to a higher rate

of subsequent migration. Hence, “the wider is cultural distance, the more rapidly migration will accelerate” (Collier & Hoeffler, 2014, 24). Collier and Hoeffler take their empirical results as evidence for this interpretation, explaining that, “whereas the direct effect of cultural distance on migration is small and insignificant, its interaction with the diaspora has a large and significant effect. As postulated, the wider is the gap in cultures the more useful is the accumulated stock of immigrants to subsequent migrants” (Collier & Hoeffler, 2014, 24). Curiously, Collier and Hoeffler did not find a significant interaction between the diaspora and linguistic distance. They take this as evidence that linguistic distance may not be an important determinant of the assimilation rate, and that absorption may be mainly a function of culturally determined frequency of social interaction. The interactions between cultural distance and the diaspora and their interpretations are consistent with Putnam’s (2007) theory of social capital in two ways. The result that a larger diaspora network significantly offsets migration costs is consistent with the idea that social networks are valuable enough that they would reasonably influence decision making. The interpretation that cultural distance slows the development of social ties and, consequently, the rate of assimilation is consistent with the idea of conflict theory and that increased diversity without a new encompassing identity is an impediment to the formation of social networks (Collier, 2013, 57, 231; Putnam, 2007).

This Study’s Treatment of the Diaspora

Following from Collier and Hoeffler (2014) and Beine et al. (2011), my study will further investigate the role of the diaspora as a mitigator of migration costs and a driver of migration rates. To do so, this study will interact the size of the diaspora with variables that measure

impediments to migration, including income per capita at the country of origin, physical distance, and a former colonial relationship. As was previously discussed in the "Role of language" section of this paper, Collier and Hoeffler used only a single linguistic proximity variable, limiting their ability to capture the distinctive roles of language. It is possible that the components of language relating more to the practical transfer of information, such as common spoken language, may substitute for the information transfer provided by a large diaspora. If this is the case, a common spoken language could reduce the importance of diaspora networks. To investigate this possible relationship, this study interacts the size of the diaspora with CSL. Analyzing these effects will test the robustness of previous empirical results and provide additional insight on the role of the diaspora as a cost mitigating force.

In addition to focusing on the cost mitigating effects of the diaspora, this study will also revisit Collier's theory of the effects of cultural distance on the assimilation of the diaspora and the acceleration of migration (Collier, 2013, 27). Collier and Hoeffler (2014) found that while the effect of cultural distance alone was insignificant, the effect of the interaction between the size of the diaspora and cultural distance to be highly significant, which they interpreted in support of Collier's theory as previously described. However, as was noted in the "Role of culture" section of this paper, Collier and Hoeffler used genetic distance, an unorthodox measure, as a proxy for cultural distance. Because this genetic distance measure is used in the interaction term, the validity of this result depends on whether genetic distance is in fact a reasonable proxy for cultural distance. The authors themselves acknowledge the need for further research in this dimension. To test the robustness of Collier and Hoeffler's result, this study uses a measure of cultural distance based on the widely used Hofstede cultural dimensions, and interacts this variable with the size of the diaspora (Hofstede, 1980; Kogut and Singh, 1988). To

my knowledge, Collier and Hoeffler were the first to explicitly empirically analyze this relationship between cultural distance and the diaspora, and in doing so they produced compelling results. Replicating their methods with a conventional measure of cultural distance will provide additional evidence for or against the potentially powerful culture-diaspora dynamics theorized by Collier.

Chapter 2

Detailed Description of the Data

Empirical Specification

My model analyzes the flow of migrants from low- and middle-income countries to high income countries from 1960-2000. The following regression represents the unrestricted version of my model, which includes all variables and interaction terms, as well as decade dummies.

$$migrant\ stock_{ijt} - migrant\ stock_{ijt-1}$$
$$= \alpha + \beta_0 migrant\ stock_{ijt-1} + \beta_1 \ln Distance_{ij} + \beta_2 CSL_{ij} + \beta_3 COL_{ij} + \beta_4 CNL_{ij}$$
$$+ \beta_5 LP_{ij} + \beta_6 CD_{ij} + \beta_7 CRE_{ij} + \beta_8 \ln GDP_{it-1} + \beta_9 \ln GDP_{jt-1} + \beta_{10} \ln Pop_{it-1}$$
$$+ \beta_{11} \ln Pop_{jt-1} + \beta_{12} Polity_{it-1} + \beta_{13} Polity_{jt-1} + \beta_{14} Hostility_{it-1}$$
$$+ \beta_{15} Hostility_{jt-1} + \beta_{16} landlocked_i + \beta_{17} landlocked_j + \beta_{18} island_i + \beta_{19} island_j$$
$$+ \beta_{20} contiguity_{ij} + \beta_{21} Colony_{ij} + \beta_{22} Seventies_t + \beta_{23} Eighties_t + \beta_{24} Nineties_t$$
$$+ \beta_{25} Colony_{ij} : \ln GDP_{it-1} + \beta_{26} Colony_{ij} : \ln GDP_{jt-1} + \beta_{27} Colony_{ij} : Polity_{it-1}$$
$$+ \beta_{28} \ln Distance_{ij} : \ln GDP_{jt-1} + \beta_{29} migrant\ stock_{ijt-1} : \ln GDP_{it-1}$$
$$+ \beta_{30} migrant\ stock_{ijt-1} : \ln GDP_{jt-1} + \beta_{31} migrant\ stock_{ijt-1} : Polity_{it-1}$$
$$+ \beta_{32} migrant\ stock_{ijt-1} : Polity_{jt-1} + \beta_{33} migrant\ stock_{ijt-1} : \ln Distance_{ij}$$
$$+ \beta_{34} migrant\ stock_{ijt-1} : Colony_{ij} + \beta_{35} migrant\ stock_{ijt-1} : CSL_{ij}$$
$$+ \beta_{36} migrant\ stock_{ijt-1} : COL_{ij} + \beta_{37} migrant\ stock_{ijt-1} : CNL_{ij}$$
$$+ \beta_{38} migrant\ stock_{ijt-1} : LP_{ij} + \beta_{39} migrant\ stock_{ijt-1} : CD_{ij} + e_{ijt}$$

Where i denotes the country of origin, j denotes the country of destination, and t denotes time including years 1970, 1980, 1990, and 2000. The regression expresses the change in migrant stocks as a function of initial stocks (Diaspora), the natural logarithm of physical distance between countries, language variables (CSL, COL, CNL, LP), cultural distance (CD), common religion (CRE), the natural logarithm of lagged per capita income levels (ln GDP), the natural logarithm of lagged population (ln Pop.), lagged political stability (Polity), lagged hostility levels between countries, whether a country is landlocked, whether a country is an island, whether two countries are contiguous (Contiguity), and former colonial relationship (Colony). The regression also includes decade dummies, and numerous interaction terms. I interact the diaspora with various push and pull variables, language variables and cultural distance. I also interact former colonial relationship with various push and pull variables. Models one through three in the results table selectively omit certain variables, which I will describe in detail in the next section.

The remainder of this section provides a detailed discussion of each of the variables included in this model.

The Dependent Variable

The conceptual dependent variable that I wish to analyze is the inflow of new migrants from the country of origin to the country of destination in each decade. Like Collier and Hoeffler (2014), I use data from The World Bank Global Bilateral Migration Database for my dependent variable. Özden, Parsons, Schiff and Walmsley (2011) completed the Global Bilateral Migration Database, currently publicly available from the World Bank (World Bank, 2011). The authors discuss the creation of this dataset in their article "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960–2000." The construction of this database was a massive undertaking, during which the authors brought together 3,500 individual census and population records from the United Nations Population Division's Global Migration Database, covering over 230 destination countries (Özden et al., 2011). The resulting dataset provides the first comprehensive look at global bilateral migration for the second half of the twentieth century, and opens the door for more advanced empirical investigation of the causes and consequences of international migration in this period. For a thorough discussion of issues encountered in constructing the dataset, including how destination countries classify migrants, geopolitical changes in the second half of the 20th century leading to changing international borders, the aggregation of source data to geographic categories rather than country of origin, and dealing with omitted and missing data see Özden et al. (2011).

The data provide stocks of migrants by country of origin living in destination countries in the years 1960, 1970, 1980, and 1990, where migrants are defined as people who were born in one country, but currently live in another (Ozden et al., 2011). The data are arranged in country pairs, which are directed dyads, meaning that country pair ij is not the same as country pair ji (Collier and Hoeffler, 2014). While the data provide migrant stocks at points in time, the variable of interest is the migrant flows between these points in time. Mimicking the methodology of Collier and Hoeffler (2014), migrant flows are approximated by taking the difference between the stock of migrants at time t and t-1, such that migrant flow for a given country pair in a given period is defined as migrant_stockijt - migrant_stockijt-1. Unfortunately, the difference in the stocks between time t-1 and t is not exactly equal to the actual inflow of new migrants, as some of the migrants from time t-1 will have died, returned to their home country, or migrated to a new country altogether by time t (Collier & Hoeffler, 2014). Because of this, this method will likely understate migrant flows. To account for this, I adjust the data for migrant mortality using a similar method to Collier and Hoeffler (2014). This method involves adding the estimated number of migrants that died during the decade to the original measure of migrant flow, which yields a measure of migrant flow adjusted for mortality. Unfortunately, I do not have any information on return or secondary migration, so the flow values may still be understated to some degree.

In cases where there were no people who were born in country i that lived in country j in year t, a zero is recorded for migrant_stock for country pair ij for that year. This occurs somewhat frequently in the data, raising concerns of sample selection bias. Collier and Hoeffler (2014) discuss the option of dropping these zeros and using only positive observations. To test for sample selection bias, Collier and Hoeffler included a Heckman correction, using the

existence of diplomatic relations in 1960 as their selection variable. The authors found the inverse Mills Ratio to be insignificant, and concluded that they could not find any evidence for sample selection bias. Additionally, they point out that many of these zeros could actually represent a very small number of migrants that were not reported in census data due to confidentiality issues. Hence, because it is desirable to analyze zero or small stocks of migrants that do not increase over time, Collier and Hoeffler decide to retain the zero observations in their sample. Because Collier and Hoeffler did not find that doing so posed any significant empirical cost, and I also would like to analyze non-increasing small or zero stocks of migrants, I elect to do the same and retain the zero observations. For a more involved discussion of possible issues associated with using these data, review Beine et al. (2011b, 2014).

Explanatory Variables

The explanatory variables that are included in my models represent pushes and pulls that make migration either easier/more attractive or more difficult/less attractive. There are three main categories of explanatory variables: language variables, culture variables, and control variables related to the migration decision. These are drawn from three main data sources, which I will describe in detail as follows.

Language Variables

As described in the previous section, a major focus of this study is to provide a more advanced understanding of how language impacts the migration decision. To do this, I use a new,

more sophisticated set of language variables pioneered by Melitz and Toubal (2014). In the paragraphs that follow, I will provide background on the construction and features of these variables as it is relevant to this thesis.

Common official language is the measure that most previous studies have used as their binary control for common language. As this variable has been constructed many times before, Melitz and Toubal (2014) note that COL was the simplest variable to construct. To construct this measure, the authors used data on official languages from the CIA World Factbook, but also referenced a broader range of sources in complex cases (CIA World Factbook, 2009). Though there are many countries that have numerous official languages, the authors chose to limit the number of languages considered per country to two. When a country had more than two official languages, they kept the two that were most important to world trade. This was done both to simplify the construction of the variable, and because the authors were interested only in official languages that are relevant and understandable to anyone within a country. The measure resulting from this exercise includes 19 official languages, which are listed in the table below.

Common spoken language and common native language are the first measures that differ significantly from traditional language variables, and play an important role in this study. CSL measures the extent to which a country's residents are able to speak the same language as the citizens of another country. CNL measures the extent to which a country's residents share the same native language as the citizens of another country. Melitz & Toubal (2014) constructed CNL and CSL from the same datasource, the EU Survey in November-December 2005, which covers 28 current EU members as well as Turkey (Special Eurobarometer 243, 2006). The authors constructed CSL by summing the percentage responses to the survey question "Which languages do you speak well enough in order to be able to have a conversation, excluding your

mother tongue (… multiple answers possible)". Similarly, they constructed CNL by recording percentage responses to the survey question "What is your maternal language." Supplemental data was also drawn from a variety of sources. To be included as a CSL or CNL, the authors required that a language be spoken by at least 4% of the population of at least two countries. Survey respondents were able to list up to 3 spoken languages in addition to their native language. Because of this, if a pair of respondents from their respective countries in a country pair were to be proficient in the same two or three non-native languages, they would be double counted in the authors summation. The authors found this to be a common issue, and made an adjustment to account for it. The resulting data yielded 42 CSL languages, of which 19 are also COLs. Every CSL is also a CNL in the sample. These additional languages are again provided in the table below.

The last linguistic measure that I am interested in considering in this study is linguistic proximity of native languages between countries. For Melitz and Toubal (2014), the purpose of such a measure was to account for ease of communication between countries that share no official, spoken, or native language, and also for the role of ethnic rapport when native languages differ. In the context of my study, migrants may be able to more easily understand and acquire host country languages that are linguistically proximate to their own native language. The ethnic rapport associated with speaking a similar native language to the host population may also have implications for social capital outcomes. It is therefore desirable to account for linguistic proximity in my study.

When assembling data to construct their measures of LP, Melitz and Toubal (2014) again consider a maximum of only 2 native languages per country. In cases where there are 2 native languages, the authors adjusted their relative percentages to sum to 1, which is the same score

that is given to a single native language. For example, “Switzerland shows 0.74 for German and 0.26 for French, Bolivia 0.54 for Spanish and 0.46 for Quechua.” The authors used 89 native languages in calculating LP. If a country exhibited a very high index of linguistic diversity according to *Ethnologue*, meaning that there were no clearly dominant native languages in that country, the authors assigned a zero to that country by necessity. Because of this, the authors assigned 31 zeros.

Melitz and Toubal (2014) then used this data to construct two different measures of LP using differing methods, the results of which they designated LP1 and LP2. LP1 was constructed using a similar method to Fearon (2003), whose measure Collier and Hoeffler used in their migration study, and has also been used in studies on a broad range of topics. LP1 was calculated based on the *Ethnologue* language tree, which consists of trees, branches, and sub-branches. The authors assigned the following values for LP1: 0 when two languages belong to separate trees, 0.25 when two languages belong to different branches of the same tree, 0.50 when two languages belong to the same branch, and 0.75 when two languages belong to the same sub-branch. However, the authors found that LP1 proves problematic when comparing languages belonging to different trees, as it assumes, for instance, that two languages in the same branch of one tree are equally close to one another as two languages from a different tree, which is not necessarily the case.

Instead of a language tree, LP2 is based on the Automated Similarity Judgment Program’s (ASJP) measurement of similarity between 200 words (Melitz & Toubal, 2014). The authors normalized both LP1 and LP2 such that their averages would equal exactly one. Because of this, the coefficients on LP1, LP2, and COL are all exactly comparable to one another. In their analysis of bilateral trade, Melitz and Toubal (2014) included versions of their model with each

of these measures individually. I will take a similar approach, and will include LP1 and LP2 in different versions of my model.

Cultural Distance Variable

One of the main goals of this study is to analyze the role of cultural distance, and how it interacts with the diaspora, in determining migration rates. In contrast to Collier and Hoeffler (2014), I choose to use a more conventional measure of cultural distance. In his now famous 1980 book *Culture's Consequences*, Geert Hofstede (1980) pioneered the measurement of national cultures and their effect on work related values based on 6 relevant dimensions. These cultural dimensions have since become standard in a wide variety of research contexts. Kogut and Singh (1988) developed a method of turning Hofstede's measures into a single bilateral index of cultural distance, which has been commonly used in the gravity literature. I use the Kogut and Singh index as the cultural distance variable in my models.

Though the Kogut and Singh index has been used very widely in international business research over the years, there has been some recent criticism of using this methodology to quantify cultural distance. In their article *Why We Should Stop Using the Kogut and Singh Index*, Palitha Konara and Alexander Mohr (2019) argue that the index is incorrectly specified, and inadvertently captures squared cultural distance. The authors claim that this misspecification may cause empirical results that utilize the original measure to be misleading. Others have suggested that the concept of cultural "distance" is flawed fundamentally. Oded Shenkar (2012) argues that cultural distance is conveniently used due to the intuitive "background metaphor" of geographic distance; however, distance may be a poor metaphor for "capturing the essence of cultural

differences, and their impact on international business phenomena." Instead, Shenkar argues that the metaphor "friction" between cultures may more accurately capture the relevant forces at play in cultural interactions, and proposes the development of new measurements based on this concept. While this does present a concern for the validity of this study's results, it is not within the scope of this project to correct for the apparent issues with the Kogut and Singh measure. Furthermore, a great number of previous studies have used this construct in empirical analyses, including recent work such as Aw et al. (2019), meaning that my results will be comparable to others that use this measure. I would encourage those conducting future gravity studies of migration to consider the benefits of correcting for the misspecification of the Kogut and Singh index, and possibly moving towards alternative measures of differences between cultures.

Control Variables

Economic/Demographic

In addition to the main variables of interest, this study also includes several control variables that may influence the migration decision. The control variables used in this study are drawn from *The Dynamic Gravity Dataset* constructed by Tamara Gurevich and Peter Herman (2018), which provides "extensive country and country pair information for a total of 285 countries and territories, annually, between the years 1948 to 2016." Here I will name each variable and briefly describe how it was constructed. For my measure of income per capita I use the variables "gdp_wdi_cap_const_o" and "gdp_wdi_cap_const_d," which provide the real GDP per capita in 2010 US$ in year t for the origin and the destination country respectively (Gurevich

& Herman, 2018). This variable is based on data from the 2016 World Development Indicators (WDI), available from the World Bank (World Bank, 2016). The variables "pop_o" and "pop_d" provide the population in millions of country_o and country_d respectively (Gurevich & Herman, 2018). Data for the population variable were sourced from the Penn World Tables (PWT), version 9.0 (Gurevich & Herman, 2018; Feenstra et al., 2015).

Geographic

When constructing their geographic distance measure the authors note that many countries cover large geographic areas, and that economic activity may be concentrated in certain areas within their borders (Gurevich & Herman, 2018). To correct for this, the authors employ a methodology developed by Mayer and Zignago (2005) to construct a measure for distance based on the distance between pairs of cities, weighted by the proportion of the country's population residing in each city, in kilometers. By doing this, the authors intended to more accurately capture the distance economic activity must travel between two countries. I use this "distance" variable as the measure for geographic distance in my study. Most of the data used to calculate the distance measure was collected from Simplemaps.com's "World Cities Database" (Gurevich & Herman, 2018; Simplemaps, 2015). Contiguity is a binary indicator equal to 1 if country_o and country_d border one another in year t. Landlocked_o and landlocked_d are binary indicators equal to 1 if country_o or country_d does not border an ocean or body of water connected to an ocean. Island_o and island_d are binary variables equal to 1 if country_o or country_d does not share a land border with any country or territory. The variables contiguity,

landlocked, and island were based on information primarily collected from maps at the United States Library of Congress.

Political

The three political variables that I include in my study are polity, hostility, colony_of_destination_ever. Polity is a variable denoting political stability based on the Polity IV Project (Marshall et al., 2016). The polity_o and polity_d variables score origin and destination countries on an ordinal scale from -10 to 10 on their level of autocracy to democracy, with -10 being the most autocratic and 10 being the most democratic (Gurevich & Herman, 2018; Marshall et al., 2016). Hostility variables are constructed based on data from The Correlates of War Project's (CoW) Militarized Interstate Disputes dataset (MIDB v4.01) (Palmer et al., 2015). The variable hostility_level_o denotes the level of hostility of country_o towards country_d in year t on a scale of 1-5, while hostility_level_d denotes the reverse. The meanings of each level of the scale are as follows: 1 - no militarized action taken, 2 - some form of threat made, 3 - some show of force or alert of force, 4 - some use of substantial force, 5 - complete war. Data for the variable "colony_of_destination_ever" are again drawn primarily from the CoW Project (Correlates of War Project, 2017). This variable simply denotes whether the country of origin has ever been a colony of the destination country.

Chapter 3

Empirical Results and Discussion

Model Specifications

I now turn to the estimation of my models. I estimate 4 different versions of my model, all of which are OLS regressions with robust standard errors. Results for each model are reported in the table below. As mentioned earlier, the model in column 4 is the unrestricted model, which includes all variables. The model in column 1 does not include cultural distance, or the language variables. The model in column 2 includes the language variables, but does not include cultural distance. The model in column 3 includes cultural distance, but does not include the language variables. My sample for cultural distance included significantly fewer country observations than my sample for the language variables. Thus, the models in columns 3 and 4, which included cultural distance, omitted countries with missing observations for cultural distance. This reduced the number of observations from 4257 to 1568, unfortunately narrowing the scope of validity for these models. The following paragraphs provide an explanation of the reasoning behind the various control methods and interaction terms included in the models.

Migration Restrictions and Fixed Effects

I control for time effects by including decade dummies. It is important to note that regulations on immigration by host countries are likely to have important effects on the size of migrant flows. As Collier and Hoeffler (2014) note, if immigration regulations were exogenous to demand for migration, took the form of quotas, and were enforced with complete success, the

size of the quota would fully explain the size of migrant flows. In this case, the decisions of migrants would only affect the composition of migrant flows, but the size would be limited to exactly the quota set by the host country. However, Collier and Hoeffler (2014) further note that migration restrictions rarely take the form of absolute numerical limits, and are rarely fully effective. Therefore, the volume of migration into a particular country likely is affected by migrants' demand to move to that country, but it is also likely affected by that country's restrictions on migration that are not absolute numerical limits. It is therefore desirable to control for the effect of host country restrictions on migration. Mayda (2007) does this by including a variable, which "increases by one (decreases by one) if in that year the destination country's immigration policy became less (more) restrictive, zero otherwise" (12). Collier and Hoeffler (2014) acknowledge that it would be desirable to directly control for migration restrictions, and suggest that the use of data from Koopmans et al (2012), who conclude that immigrants' rights differ significantly among ten Western European countries, to control for differences in restrictiveness would be optimal. However, using this data would have greatly limited their sample size, and because there is currently no similar data source covering a broader selection of countries, they were unable to include any direct measure of restrictions in their analysis. Instead, Collier and Hoeffler accounted for restrictions on migration in two ways. First, they included a variable indicating a former colonial relationship. They interpret a former colonial relationship as a proxy for more relaxed migration restrictions, with the intuition being that former colonizers would be expected to impose lighter restrictions on migration from former colonies. They also included a variable indicating whether a country of origin is landlocked. When a country is landlocked, the only way to migrate is via direct flight, which may be cost prohibitive, or by traveling through other countries by land. Collier and Hoeffler point out that, if

migrants must travel through other countries to reach the destination, not only must they face destination country restrictions, but they must also face the restrictions imposed by all intermediate countries. Although these measures may not be as closely related to destination country restrictions as I would like, I unfortunately have no other way of controlling for restrictions on migration, so I include both of these measures as proxies for migration restrictions in my model.

Interactions

As Collier and Hoeffler (2014) suggest, the decision to migrate is a complex one, and is influenced by a wide range of influences. Because of this, it is useful to investigate interactions between explanatory variables. Mayda (2007) points out that, according to the theory, in years where restrictions are eased, the effect of pull factors should become more positive, and the effect of push factors should become more negative. To empirically test this theory, Mayda interacts key push and pull variables with her variable measuring restrictions on migration, and finds results to be consistent with this theory, although she finds that pull factors to be more important than push factors. While Collier and Hoeffler do not use a variable directly measuring restrictions on migration (a constraint), they do include several variables that measure forces that may influence the migration decision by either mitigating the burden of a constraint to migration, or strengthening the effect of a pull variable. Namely, these variables are a former colonial relationship and the size of the diaspora. Collier and Hoeffler interact these two variables with numerous important impediments to and facilitators of migration. For a former colonial relationship, the intuition is that increased information flows, or expectations set during

education should have a positive effect on potential migrants' expected outcomes when migrating to the formerly colonizing country, relative to the actual average income in that country. For the diaspora, the intuition is that the support of a larger diaspora network should make constraints to migration less binding, and may substitute for other pull variables that would otherwise be more important. While I unfortunately do not have access to a variable directly measuring migration restrictions, I interact the size of the diaspora and former colonial relationship with numerous push and pull variables, with the goal of investigating the role of these variables as cost mitigators and substitutes for pull variables. Because my study is primarily focused on the effects of the distinctive effects of language, cultural distance, and diaspora networks, I interact the size of the diaspora with all of my language variables, and my cultural distance variable. With these considerations in mind, I now estimate my models. Results are shown in the table below.

Table 1 Regression Results

	DV: Adjusted Migrant Flow			
	migrant_flow_adj			
	(1)	(2)	(3)	(4)
Diaspora	-4.532***	-2.175***	-4.208***	0.645
	(0.332)	(0.321)	(0.663)	(0.540)
ln Distance	-6,831.228	-4,878.999	4,091.181	10,073.660
	(28,880.160)	(26,520.040)	(64,761.560)	(50,312.800)
CSL		4,209.419		-1,395.934
		(4,841.090)		(15,085.530)
COL		4,125.678*		16,612.890
		(2,469.825)		(10,138.800)
CNL		-51,972.080***		-112,024.400***
		(8,015.897)		(21,034.870)
LP2		-1,107.970		1,123.161
		(1,047.301)		(2,271.125)
Cultural Distance			-2,049.261	-581.357
			(1,849.690)	(1,469.757)
Common Religion	4,417.932	5,676.654	10,179.610	2,897.630
	(3,590.057)	(3,704.969)	(8,557.947)	(7,719.576)
ln GDP Origin t-1	-2,448.553***	-1,155.145	-5,782.720**	-332.456
	(737.608)	(710.112)	(2,354.932)	(1,824.232)
ln GDP Dest. t-1	-3,327.576	-3,505.866	3,480.209	12,304.900
	(24,993.260)	(22,953.710)	(56,980.640)	(44,333.150)
ln Pop. Origin t-1	-0.726	470.421	194.351	979.626
	(460.573)	(422.969)	(1,642.139)	(1,335.690)
ln Pop. Dest. t-1	-314.872	136.108	-1,485.353	1,187.017
	(567.183)	(519.960)	(1,532.819)	(1,193.496)
Polity Origin t-1	62.355	54.199	374.724	54.853
	(112.970)	(105.816)	(287.663)	(233.810)
Polity Dest. t-1	680.992***	432.065**	1,460.978**	769.878*
	(236.261)	(218.004)	(566.512)	(439.350)
Hostility Origin t-1	8,732.803**	4,707.196	38,508.930***	-6,629.283
	(3,996.944)	(3,656.484)	(13,786.440)	(10,737.490)
Hostility Dest. t-1	-8,407.009**	-6,259.415**	-41,432.040***	-10,803.070
	(3,466.304)	(3,168.767)	(11,443.290)	(8,855.684)
Landlocked Origin	-2,177.693	-2,173.521		
	(2,147.712)	(1,968.732)		
Landlocked Dest.	-697.347	-1,049.298	72.026	-616.089
	(2,468.240)	(2,268.873)	(6,497.656)	(5,028.126)
Island Origin	1,925.496	-337.542	4,265.385	471.951
	(2,199.268)	(2,076.697)	(5,866.555)	(5,488.335)
Island Dest.	2,492.331	796.798	7,326.611	1,787.460
	(1,852.861)	(1,701.019)	(4,746.467)	(3,719.166)
Contiguity	161,239.400***	141,619.500***	176,173.600***	28,966.130
	(13,788.030)	(12,823.620)	(29,251.360)	(24,229.540)
Colony	325,258.700***	461,263.300***	684,192.900**	789,419.500***
	(105,436.500)	(97,259.270)	(284,448.100)	(219,765.600)

70s Decade Dummy	4,749.118*	4,932.846**	11,933.960*	9,037.153*
	(2,600.529)	(2,375.283)	(6,109.822)	(4,685.337)
80s Decade Dummy	45.307	1,765.332	2,429.479	4,033.562
	(2,765.963)	(2,530.842)	(6,763.305)	(5,191.528)
90s Decade Dummy	-793.664	1,083.832	-2,939.936	1,246.368
	(3,007.192)	(2,755.019)	(7,560.977)	(5,818.597)
Colony:ln GDP Origin t-1	-15,038.800***	-8,604.859**	-46,541.940***	-26,573.290***
	(4,273.218)	(4,086.924)	(12,193.190)	(10,303.100)
Colony:ln GDP Dest. t-1	-21,038.190**	-40,672.680***	-32,060.880	-55,925.140***
	(9,757.569)	(9,109.627)	(26,989.820)	(21,159.690)
Colony:Polity Origin t-1	1,554.022***	284.121	3,903.726**	-3,204.275**
	(596.288)	(551.984)	(1,792.543)	(1,394.126)
ln Distance:ln GDP Dest. t-1	305.715	277.168	-542.290	-1,500.514
	(2,842.494)	(2,609.416)	(6,368.482)	(4,949.026)
Diaspora:ln GDP Origin t-1	0.298***	0.124***	0.307***	-0.040*
	(0.014)	(0.015)	(0.024)	(0.024)
Diaspora:ln GDP Dest. t-1	0.078**	0.101***	0.121*	0.031
	(0.037)	(0.036)	(0.069)	(0.056)
Diaspora:Polity Origin t-1	-0.006***	0.001	-0.008***	0.010***
	(0.002)	(0.002)	(0.003)	(0.002)
Diaspora:Polity Dest. t-1	-0.234***	-0.245***	-0.284***	-0.267***
	(0.016)	(0.015)	(0.030)	(0.023)
Diaspora:ln Distance	0.545***	0.388***	0.462***	0.320***
	(0.012)	(0.013)	(0.025)	(0.021)
Diaspora:Colony	-0.739***	-0.264***	-0.589***	-0.585***
	(0.027)	(0.060)	(0.072)	(0.109)
Diaspora:CSL		-0.763***		0.674***
		(0.093)		(0.149)
Diaspora:COL		-0.088		-0.327***
		(0.069)		(0.110)
Diaspora:CNL		2.732***		6.418***
		(0.098)		(0.205)
Diaspora:LP2		0.124***		-0.344***
		(0.027)		(0.041)
Diaspora:Cultural Distance			0.123***	-0.045**
			(0.023)	(0.020)
Constant	77,484.500	55,236.040	10,211.530	-91,151.600
	(253,601.800)	(232,976.400)	(578,804.600)	(449,817.500)
Observations	4,257	4,257	1,568	1,568
R^2	0.805	0.838	0.823	0.896
Adjusted R^2	0.803	0.836	0.819	0.894
Residual Std. Error	42,892.840 (df = 4226)	39,163.520 (df = 4218)	66,740.350 (df = 1536)	51,163.780 (df = 1528)
F Statistic	580.493*** (df = 30; 4226)	572.118*** (df = 38; 4218)	230.156*** (df = 31; 1536)	339.131*** (df = 39; 1528)
Note:				*p<0.1; **p<0.05; ***p<0.01

Empirical Results and Discussion

Overall, the results produce an image of the migration decision that is both consistent with a traditional theoretical conception of migration as an investment decision, as well as with previous migration studies studying the role of diasporas, language, and culture. However, the results also pose new empirical questions.

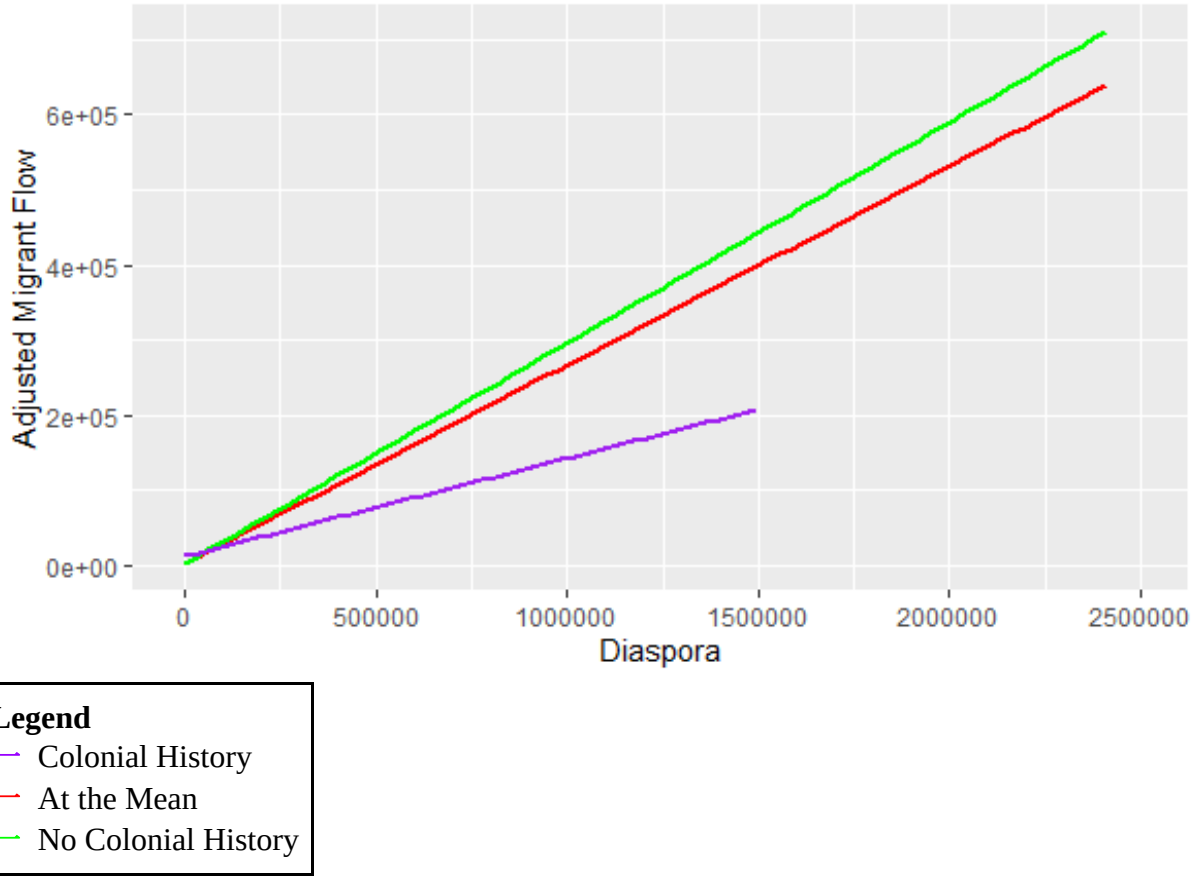
Colonial History

I find that colonial history has a significant, positive and large main effect. When understanding a former colonial relationship as a force that increases information flows, and sets potential migrants' expectations of the host country via education, this is an intuitive result. It would be reasonable to expect that the increase in information flows and familiarity with the host country associated with a former colonial relationship would psychologically ease the uncertainty associated with migration, which has been found to be an important factor in willingness to migrate (Banerjee & Duflo, 2019). However, the overall effect of each variable must also take into account the interactions, of which there are several with former colonial relationship that are significant.

Consistent with Collier and Hoeffler (2014), former colonial relationship has a negative and significant interaction with the size of the diaspora. An interpretation could be that the access to information provided by a colonial history may substitute for the information that would be provided by diaspora networks. An important way that the diaspora may influence migration is the support it offers migrants in circumventing migration restrictions, through family reunification programs and other similar pathways. If colonial history is interpreted as a proxy to

more relaxed migration restrictions, and interpretation of this result could be that when migration restrictions are lower, the diaspora becomes less important in helping migrants overcome these barriers. The graph below shows simple linear models of migrant flows and the diaspora, and illustrates how colonial history and the size of the diaspora interact in affecting migrant flows.

Figure 1 Colonial History, the Diaspora and Migrant Flow



The Diaspora

In the first three versions of my model, the main effect of the diaspora is significant and negative. This result for the main effect of the diaspora is counter intuitive, as one would expect that a larger diaspora would make subsequent migration easier, which should increase

subsequent migrant flows. However, because my regression explains net flows, and not stocks, it is difficult to interpret this result. For instance, even if diasporas have a large effect on migration, the inflows they induce may have already taken place in the previous period. Because of this, I am not able to accurately interpret the meaning of the coefficients on the diaspora itself. Regardless, it is rather unlikely that diasporas would directly reduce migration, as there is little to no theoretical or empirical backing for such an idea. While the main effect of the diaspora presents something of an empirical puzzle, the various interactions with the diaspora provide additional interesting insights.

In the first three models, the diaspora has a significant, positive interaction with GDP per capita at the origin; however, the effect becomes negative in the fourth model. When considering low-income migrants coming from low-income countries, one might expect lower income levels to represent a barrier to financing the costs of migration. Therefore, one would expect the diaspora to have a negative interaction with income at the origin, as the benefits of the diaspora might substitute for higher incomes in financing migration costs. However, because immigrant populations tend to be bimodal in terms of education and income levels, meaning that many very highly educated and very poorly educated people tend to migrate, it is not obvious how income levels of source countries should affect overall migration probabilities. For instance, highly educated migrants with high incomes should not have trouble financing migration costs, so the diaspora would not be useful to them in this way as it would be for poorly educated migrants. Overall, due to migrant heterogeneity, it is difficult to draw inferences based on the results for income levels in the country of origin.

In the first three models, the diaspora has a significant, positive interaction with GDP per capita at the destination; however, the interaction becomes insignificant in the fourth model.

While host countries with higher incomes per capita likely do generally offer better economic opportunities to migrants, there may be obstacles that prevent migrants from taking advantage of these opportunities. In many places, migrants often struggle to economically integrate, and often earn incomes that do not necessarily reflect the average wage of the host country. Perhaps, a large diaspora provides migrants with the social connections and support that they need to successfully economically integrate, and fully take advantage of the labor market opportunities provided by a higher income country. Diasporas may help migrants that would otherwise be stuck at the bottom make a 'leap' into economic success in the host society. The higher the average income, the greater the reward of making this leap.

Consistent with Collier and Hoeffler (2014), the diaspora has a significant, negative interaction with polity at the destination. Increased political stability and freedom at the destination increases the incentive to migrate, which is consistent with the positive coefficient on the main effect of polity at the destination. An interpretation of the negative interaction could be that migrants may have an easier time succeeding in countries with high political stability and freedom. This increased ease of success might substitute for the support offered by the diaspora, thus explaining the negative interaction.

There is a significant and positive interaction between the diaspora and physical distance. Distance is perhaps the most fundamental impediment to migration, and should decrease migrant flows all other things equal. The positive interaction between the diaspora and physical distance supports the idea that the diaspora is a powerful mitigator of the costs of migration. This interpretation is consistent with Collier and Hoeffler (2014).

Language Variables and Cultural Distance

The language variables rarely had significant main effects, but produced interesting results through their interactions with the diaspora. Common spoken language had an insignificant direct effect, suggesting that the improvement in communication abilities when there is an overlap in spoken languages may not be of great importance to the migration decision. Common official language had the expected sign, but was only weakly significant in the second model. A surprising, yet robust result is that common native language had a significant and negative main effect. This result was not expected, as one would expect that greater overlap in native language between countries would increase ethnic rapport, improve social capital outcomes for migrants, increase the expected return to migration, and increase migrant flows. This poses an empirical puzzle which warrants further investigation. Linguistic proximity had an insignificant main effect, suggesting that similarity in native languages when native languages differ may not influence ethnic rapport in a way that is important to migration. Cultural distance had a negative, but insignificant effect on migrant flows. A negative coefficient would be intuitive, as one would expect greater distance between cultures to make it more difficult for migrants to integrate into social networks in the host country. Unfortunately, my cultural distance variable lacked observations for many countries, including most of Africa, which greatly weakens the strength of the analysis. It is possible that results for cultural distance would be different if a more complete sample were used, and I encourage future researchers to pursue this. Nonetheless, Collier and Hoeffler (2014) also found their measure of cultural distance, genetic distance, to be mostly insignificant, so cultural distance may simply not have a

significant direct effect on migrant flows. I now turn to the interactions between the diaspora, the language variables, and cultural distance.

Interactions Between the Diaspora, Language Variables, and Cultural Distance

The Diaspora and Common Official Language

The interaction between the diaspora and COL is negative, but is only significant in the fourth model. A negative interaction between COL and the diaspora could suggest that the benefits of a common official language substitute for the benefits of a larger diaspora.

The Diaspora and Common Native Language

While the main effect of CNL is negative and significant, the diaspora has a positive and significant interaction with CNL. Below is a graph showing the interaction between the diaspora and CNL.

As it is important to the upcoming discussion, I will provide a brief recap of Collier's theory of how migration accelerates through diaspora dynamics. Diasporas, which Collier considers to be unassimilated stocks of migrants living in the host country, provide valuable social networks that ease subsequent migration. Therefore, a larger diaspora will lead to larger migrant inflows in the subsequent period. In turn, these new inflows cause the diaspora to grow even larger. Through this feedback loop of a growing diaspora, migration has a tendency to accelerate. Collier theorizes that the only way that this acceleration can be limited is through forces that cause the diaspora to decay. The primary source of decay for the diaspora is

assimilation of its members into mainstream society, the idea being that this makes them less useful to potential migrants. Therefore, according to this theory, anything that may slow the assimilation of the diaspora is bound to cause migration to accelerate (Collier, 2013). Collier theorizes that cultural distance between migrants and the host society may slow the rate of assimilation, meaning that when cultural distance is greater, diasporas will assimilate more slowly and migration will accelerate.

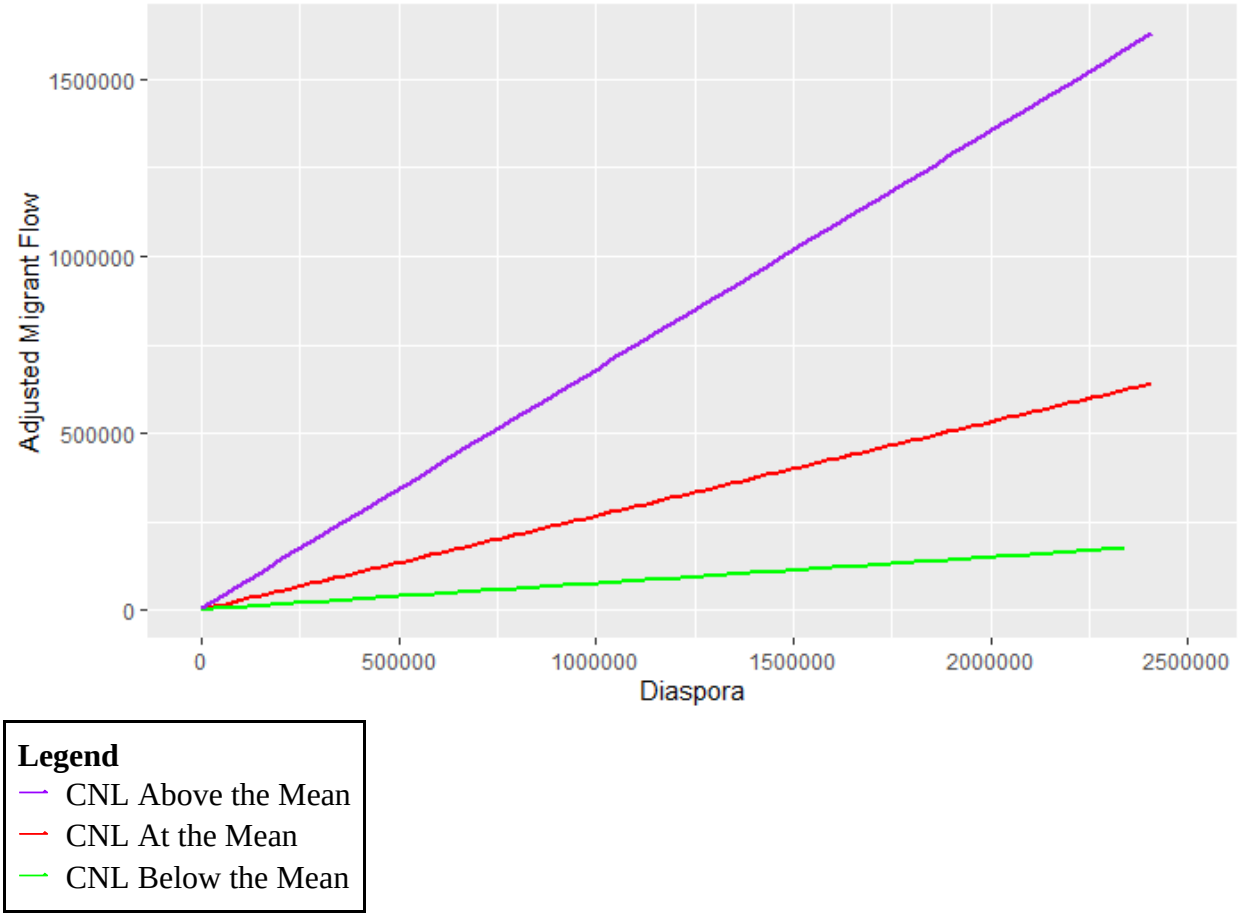
One of Collier and Hoeffler's (2014) most notable results was that they found a significant, positive interaction between cultural distance, proxied by genetic distance, and the size of the diaspora. Their result suggests that, when cultural distance is greater, the diaspora is even more effective at attracting more migrants. They interpreted this as evidence supporting their theory that cultural distance slows the decay of the diaspora, and that variables that slow the decay of the diaspora will ultimately increase migration.

While greater cultural distance should represent a *decrease* in ethnic rapport, greater CNL should represent an *increase* in ethnic rapport. The fact that CNL has a positive interaction with the diaspora is not consistent with Collier's theory. If Collier's theory were indeed correct, CNL would be expected to *increase* the rate at which the diaspora assimilates, which should cause the diaspora to shrink, and subsequent migration to decrease. Instead, my result indicates the opposite. According to my result, when ethnic rapport is stronger (high CNL), the diaspora attracts even more migration than when ethnic rapport is weaker (low CNL). This could be interpreted in a variety of ways. Perhaps when migrants share a native language with a host country, they are very successful at forging social networks with the local population, and this helps them more quickly establish themselves as economically successful members of the host society. Perhaps this well connected, economically well-off diaspora is more helpful to

subsequent migrants than a less successful diaspora would be, and perhaps CNL is one key to such success.

While it is not possible to draw any definite conclusions based on this evidence alone, this result opens new questions on the role of language, culture, and ethnic rapport in determining assimilation rates. Additionally, this result warrants the further investigation of whether the "assimilation" of migrants with the host population necessarily makes them immediately less useful to subsequent migrants from the country of origin. If they do provide less benefit to subsequent migrants after assimilating, how much less helpful do they become? How quickly do they become less helpful and what factors determine how fast this happens? The idea of "assimilation" itself warrants further study. Perhaps it is possible to assimilate to some degree, yet also identify with the local diaspora community, or maintain connections in the home country. A more nuanced understanding of assimilation could have potentially important implications for theoretical understanding of how diasporas and culture determine migration rates, and I would encourage future researchers to explore these topics in greater detail.

Figure 2 Common Native Language, the Diaspora and Migrant Flow



The Diaspora and Linguistic Proximity

While CNL measures ethnic rapport between people that share a native language, linguistic proximity of native languages theoretically measures linguistic ethnic rapport between people who do not share a native language. Likewise, one would expect LP to function in the same way as CNL, and yield results of the same sign. Interestingly, the interaction between LP and the diaspora is significant and positive in model two, but becomes significant and negative in model four after controlling for cultural distance. While a positive coefficient on the interaction between LP and the diaspora would have a similar interpretation to a positive coefficient on the interaction between CNL and the diaspora described above,

a negative interaction would be more consistent with Collier's theory. It is possible that the decrease in ethnic rapport associated with distance between native languages could result in slower assimilation, potentially preventing the diaspora from shrinking and increasing migration.

The Diaspora and Cultural Distance

The third model yields a significant *positive* interaction between the diaspora and CD, while the fourth model yield a significant *negative* interaction. A significant positive interaction would be consistent with the results of Collier and Hoeffler (2014), and would provide further evidence for Collier's theory that greater cultural distance slows the assimilation of the diaspora. However, model three omits CSL, COL, CNL and LP, none of which except for LP were included in Collier's analysis, while model four includes them. After controlling for the language variables, and their interactions with the diaspora, the interaction between CD and the diaspora becomes negative in model four. While CNL is positively associated with ethnic rapport, CD is negatively associated with ethnic rapport. Therefore, a negative interaction between cultural distance and the diaspora is intuitively consistent with a positive interaction between CNL and the diaspora. An interpretation similar to that of the positive interaction between CNL and the diaspora would again be that perhaps a diaspora that is culturally proximate to the host country would be more successful at forging social networks with the local population, helping them more quickly establish themselves as economically successful members of the host society, and making them more useful to subsequent migrants. Regardless of the interpretation, this negative interaction further calls Collier's results into question, and suggests that more research should be

done to understand the relationship between variables affecting ethnic rapport, assimilation, and the diaspora.

The Diaspora and Common Spoken Language

Common spoken language primarily represents the role that language plays in increasing the ease of communication between migrants and the host population, but also may have some social capital effects. As such, higher CSL should theoretically make it easier to migrate and acclimate upon arrival. In the second model, which omits cultural distance, the interaction between CSL and the diaspora is significant and negative. However, after controlling for cultural distance, the interaction between CSL and the diaspora becomes significant and positive. A negative interaction could be interpreted as the benefits of the diaspora substituting for the benefits of CSL. A negative interaction could also be interpreted as evidence that high CSL could increase the rate of assimilation, cause the diaspora to decay more quickly, and decrease migration. However, the fact that this interaction becomes positive and significant after controlling for cultural distance raises concerns for the robustness of this interpretation. If the interaction is positive, this suggests that higher CSL causes a larger diaspora to become more effective at attracting migration. Again, this result calls for a more rigorous investigation of how linguistic and cultural factors, especially those that might affect ease of communication, affect assimilation.

Limitations

There are several notable limitations to this empirical analysis. The most significant limitation to my analysis is that my sample for cultural distance covered significantly fewer countries of origin than the rest of my sample. Thus, I was forced to omit countries with missing cultural distance observations from models three and four, which included cultural distance. It is difficult, therefore, to analyze the effect of controlling for cultural distance, as the models that did not include cultural distance have a significantly different sample size from those that did.

Although destination country migration restrictions are important, I was only able to control for them through proxies that are loosely related to the underlying concepts. It would be beneficial to future research to construct a dataset of migration restrictions covering all countries. Consistent with the theory, Mayda (2007) found that the share of the country-of-origin population that is young is an important predictor of bilateral migrant flows. I did not control for this in my models, which may be a source of bias. My analysis also did not include any measure of income inequality or generosity of the host country's welfare system towards migrants, which may be important measures of the economic incentive to migrate beyond simple income levels.

Finally, my dataset contained no data on migrants' education levels, and treated migrant stocks as homogenous pools of people. This made some results difficult to interpret, as highly educated migrants tend to migrate for different reasons than poorly educated migrants, and migrant stocks tend to be bimodally distributed between the highly educated and poorly educated in reality. Data on the demographic composition of migrant stocks, such as individual level education and refugee status data, would greatly strengthen this type of analysis, and should be pursued in future research.

Policy Implications and Conclusion

Suggestions for Policy Makers Attempting to Predict Future Migration

It is clear that the demand for migration currently exceeds the migration that actually occurs. This is due in large part to the migration policies of destination countries, which impose some degree of restrictions which make it difficult or more costly for people to immigrate. Some economists argue that relaxing these migration restrictions represents an unprecedented opportunity to alleviate global poverty, while others are more skeptical, believing that large increases in global migration will lead to costs that will offset the gains (Borjas, 2015; Clemens, 2011). Regardless of the correct way to think about this issue, the reality is that, because demand outstrips supply, the relaxation of migration restrictions is likely to induce an inflow of new migration. If policy makers are going to consider relaxing migration policy, they will need to be able to predict the size of the inflows that they will be inducing as accurately as possible. This thesis generates results that in some ways offer a better understanding of what drives migration, which policy makers may find useful when predicting future flows. In this thesis, I test the hypothesis that migrants prefer destination countries that are culturally, linguistically and religiously proximate to their own, due to the expected ease of social integration, the expected ease of labor market integration and migrants' own living preferences.

Overall, this thesis finds little evidence that cultural similarities play an important role in the decision to migrate. Based on these results, policy makers should not necessarily expect to induce larger inflows when relaxing migration restrictions on culturally proximate countries. An important finding of this thesis is that the size of the diaspora in destination countries matters. In

particular, a larger diaspora functions as a mitigator of the costs of migration. Based on this result, policy makers should consider that relaxing migration restrictions on countries that would have been expected to face higher barriers to migration (longer distance, no colonial relationship, etc.) may induce higher than expected inflows of new migrants if there is a large diaspora present. Generally, policy makers should be aware that diasporas can have significant effects on migration by either substituting for something that would facilitate migration, or offsetting the costs of certain barriers to migration.

Final Word on Cultural Differences and Migration Acceleration

A central concern raised by Collier (2013) was that, given the presence of a diaspora in a destination country, migration may accelerate to unsustainable levels, especially if there is a greater cultural distance between the host and sending country. While it is not immediately clear whether cultural differences between migrants and the host society impose significant costs in the form of social capital externalities at a given level of migrant stocks, this thesis finds evidence casting doubt on the hypothesis that cultural differences cause migration to accelerate by reducing the rate of assimilation of members of the diaspora. Collier (2013) suggested that, in order to prevent runaway migration, countries should consider imposing greater migration restrictions on more culturally distance countries, as these migrants would have a greater tendency to avoid assimilation, accumulate into large diasporas, and accelerate migration. The results of this thesis suggest that concerns of "runaway migration" driven by cultural differences may be overstated. While more empirical research should be done to better understand this relationship, in the meantime, policy makers should carefully consider whether it is actually necessary to impose greater restrictions on more culturally distant countries, as this idea may not be supported empirically, and may lead to unnecessary restrictions on migration. If migration is observed to be

accelerating, it is possible that this process is being driven by other factors entirely. In instances where acceleration is observed, further research should be conducted to more concretely understand the forces driving the acceleration.

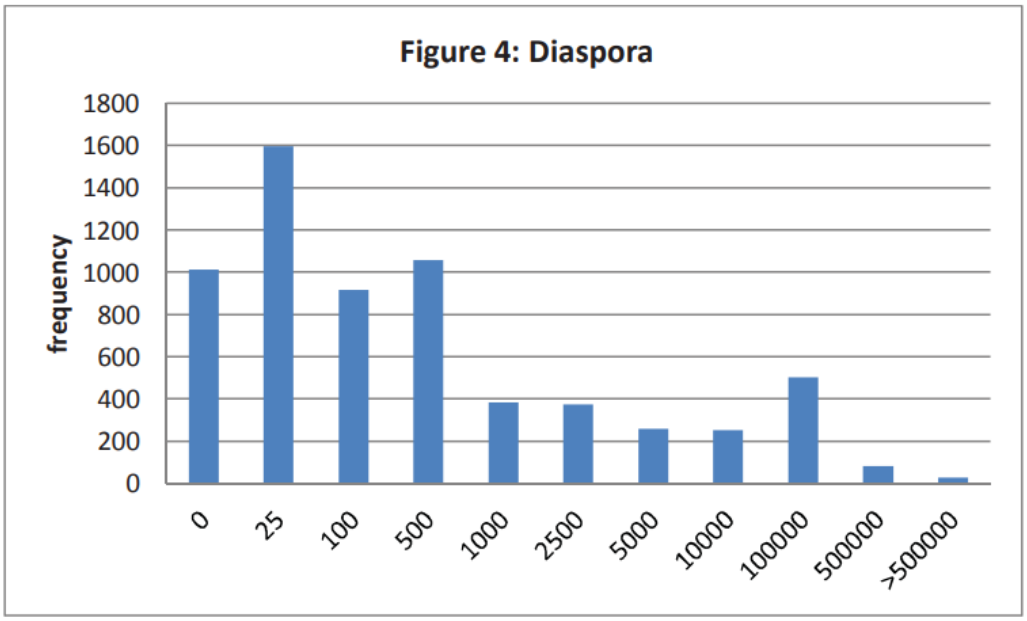
Appendix

Summary Statistics and Robustness Checks

Table 2 Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
migrant_stock	8,080	14,050.650	147,316.800	0	9	1,304.8	9,367,910
migrant_stock_lag	8,080	9,235.702	94,889.570	0	3	577.2	4,662,233
migrant_flow	8,080	4,814.946	69,839.380	-820,763	0	496.5	4,705,677
migrant_flow_adj	8,080	5,461.448	73,118.550	-769,244.800	0.000	546.314	4,915,477.000
c	2,160	2.408	1.068	0.105	1.643	3.127	5.443
csl	7,728	0.108	0.189	0.000	0.001	0.114	1.000
col	7,728	0.112	0.316	0.000	0.000	0.000	1.000
cnl	7,728	0.018	0.107	0.000	0.000	0.000	0.990
lp1	7,728	0.704	1.190	0.000	0.000	1.946	5.838
lp2	7,728	0.663	0.731	0.000	0.000	0.894	5.463
cre	7,728	0.141	0.194	0.000	0.008	0.205	0.912
gdp_wdi_cap_const_o	6,620	3,006.799	3,844.257	171.303	676.319	3,687.392	26,419.130
gdp_wdi_cap_const_lag_o	5,540	2,683.099	3,480.831	171.303	696.044	3,308.772	26,419.130
gdp_wdi_cap_const_d	7,526	31,914.570	13,365.300	8,770.833	22,542.820	38,000.690	81,732.490
gdp_wdi_cap_const_lag_d	6,642	25,663.170	11,587.130	4,506.319	17,663.570	32,276.090	63,881.690
distance	7,700	7,836.116	3,808.600	306.052	4,916.822	10,172.210	19,314.750
pop_o	7,060	40.151	141.857	0.221	3.253	23.421	1,269.975
pop_lag_o	6,240	35.977	126.413	0.221	3.059	19.899	1,154.606
pop_d	7,700	38.499	56.573	2.820	5.962	56.222	282.896
pop_lag_d	7,100	36.028	51.901	2.365	5.244	53.523	252.848
polity_o	7,640	-1.681	6.637	-10.000	-7.000	5.000	10.000
polity_lag_o	7,020	-3.009	6.339	-10.000	-7.000	2.000	10.000
polity_d	7,700	9.254	3.209	-9.000	10.000	10.000	10.000
polity_lag_d	7,100	8.735	4.155	-9.000	10.000	10.000	10.000
contiguity	7,700	0.002	0.050	0.000	0.000	0.000	1.000
landlocked_o	7,700	0.182	0.386	0.000	0.000	0.000	1.000
landlocked_d	7,700	0.100	0.300	0.000	0.000	0.000	1.000
island_o	7,700	0.135	0.342	0.000	0.000	0.000	1.000
island_d	7,700	0.200	0.400	0.000	0.000	0.000	1.000
hostility_level_o	7,700	0.016	0.252	0.000	0.000	0.000	5.000
hostility_level_lag_o	7,100	0.018	0.263	0.000	0.000	0.000	5.000
hostility_level_d	7,700	0.020	0.284	0.000	0.000	0.000	5.000
hostility_level_lag_d	7,100	0.022	0.299	0.000	0.000	0.000	5.000
colony_of_destination_ever	7,700	0.028	0.165	0.000	0.000	0.000	1.000

Figure 3 Size Distribution of Diasporas (Collier & Hoeffler, 2014)



The following charts are from Melitz and Toubal (2014)

Table 3 List of Official, Spoken and Native Languages

Official, spoken and native languages	Other spoken and native languages
Arabic	Albanian
Bulgarian	Armenian
Chinese	Bengali
Danish	Bosnian
Dutch	Croatian
English	Czech
French	Fang
German	Finnish

Greek

Italian

Malay

Persian (Farsi)

Portuguese

Romanian

Russian

Spanish

Swahili

Swedish

Turkish

Fulfulde

Hausa

Hindi

Hungarian

Javanese

Lingala

Nepali

Pashto

Polish

Quechua

Serbian

Tamil

Ukrainian

Urdu

Uzbek

Table 4 Correlations Between Language Variables

Common official language	Common spoken language	Common native language	Linguistic proximity (tree)	Linguistic proximity (ASPJ)	
Common official language	1.0000				
Common spoken language	0.5587	1.0000			
Common native language	0.5399	0.6791	1.0000		
Linguistic proximity (tree)	– 0.1634	0.1489	– 0.0980	1.0000	
Linguistic proximity (ASPJ)	– 0.2284	0.1173	– 0.1586	0.8384	1.0000

I include the following list of countries in my study. They are the same countries included by Collier and Hoeffler (2014). The origin countries are low- and middle-income countries and the destination countries are high income OECD countries, as specified in the 2000 and 2001 World Development Reports from the World Bank.

Origin Countries:
Afghanistan, Albania, Algeria, Angola, Argentina, Bahrain, Bangladesh, Benin, Bolivia, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Cape Verde, Central African Republic, Chad, Chile, China, Columbia, Comoros, Democratic Republic of the Congo, Republic of Congo, Côte d'Ivoire, Cuba, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Fiji, Gabon, Gambia, Ghana, Guatemala, Guinea, Guinea Bissau, Guyana, Haiti, Honduras, Hungary, India, Indonesia, Iran, Iraq, Jamaica, Jordan, Kenya, Korea, Laos, Liberia, Libya, Madagascar, Malawi, Malaysia, Mali, Mauretania, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Romania, Rwanda, Saudi Arabia, Senegal, Sierra Leone, Solomon Islands, Somalia, Sri Lanka, Sudan, Syria, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, Uruguay, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

Destination Countries:
Australia, Austria, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States.

Figure 4 Destination Countries

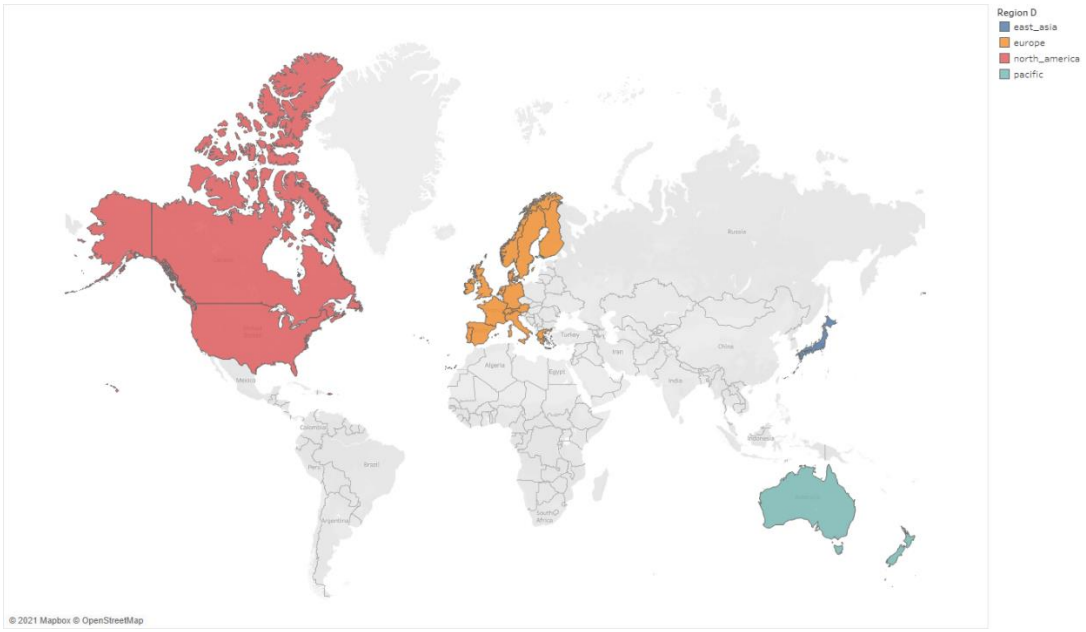


Figure 5 Origin Countries

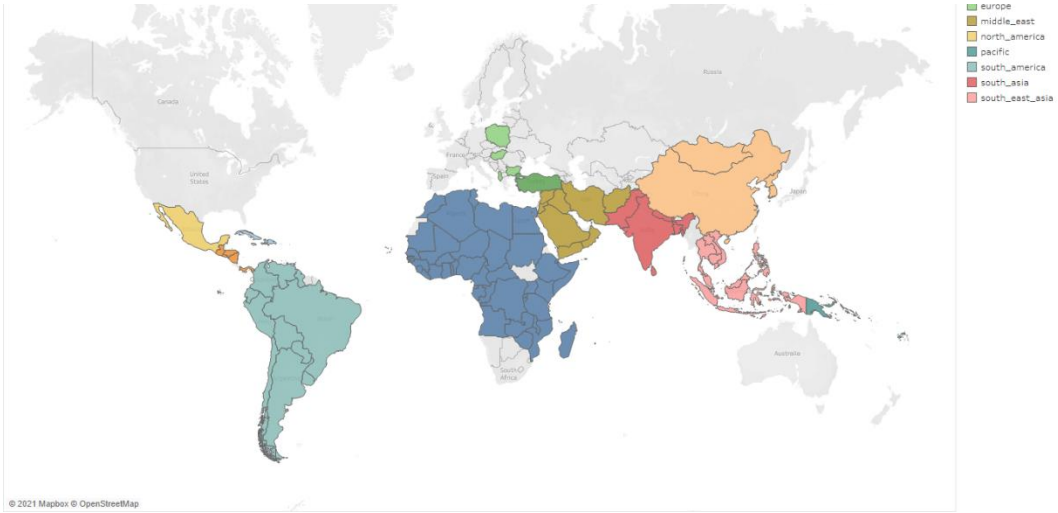
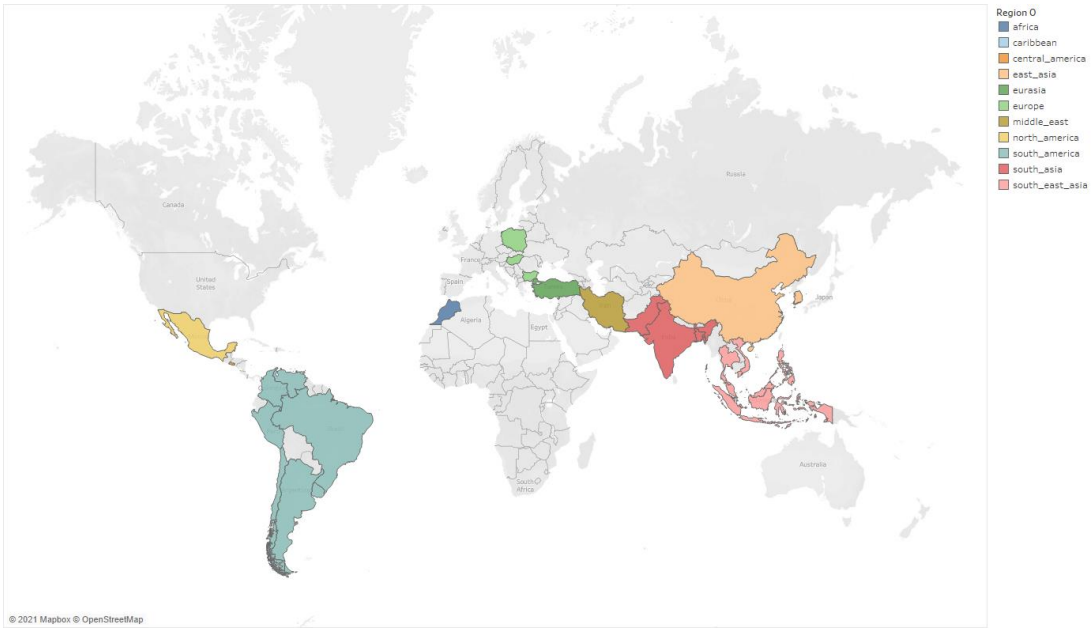


Figure 6 Origin Countries with Non-Null Cultural Distance Observations



Robustness Checks

Table 5 Regression Results: Standard Errors Clustered by Country of Origin

	DV: Adjusted Migrant Flow			
	migrant_flow_adj			
	(1)	(2)	(3)	(4)
Diaspora	-4.532***	-2.175***	-4.208***	0.645
	(0.332)	(0.321)	(0.663)	(0.540)
ln Distance	-6,831.228	-4,878.999	4,091.181	10,073.660
	(28,880.160)	(26,520.040)	(64,761.560)	(50,312.800)
CSL		4,209.419		-1,395.934
		(4,841.090)		(15,085.530)
COL		4,125.678*		16,612.890
		(2,469.825)		(10,138.800)
CNL		-51,972.080***		-112,024.400***
		(8,015.897)		(21,034.870)
LP2		-1,107.970		1,123.161
		(1,047.301)		(2,271.125)
Cultural Distance			-2,049.261	-581.357
			(1,849.690)	(1,469.757)
Common Religion	4,417.932	5,676.654	10,179.610	2,897.630
	(3,590.057)	(3,704.969)	(8,557.947)	(7,719.576)
ln GDP Origin t-1	-2,448.553***	-1,155.145	-5,782.720**	-332.456
	(737.608)	(710.112)	(2,354.932)	(1,824.232)
ln GDP Dest. t-1	-3,327.576	-3,505.866	3,480.209	12,304.900
	(24,993.260)	(22,953.710)	(56,980.640)	(44,333.150)
ln Pop. Origin t-1	-0.726	470.421	194.351	979.626
	(460.573)	(422.969)	(1,642.139)	(1,335.690)
ln Pop. Dest. t-1	-314.872	136.108	-1,485.353	1,187.017
	(567.183)	(519.960)	(1,532.819)	(1,193.496)
Polity Origin t-1	62.355	54.199	374.724	54.853
	(112.970)	(105.816)	(287.663)	(233.810)
Polity Dest. t-1	680.992***	432.065**	1,460.978**	769.878*
	(236.261)	(218.004)	(566.512)	(439.350)
Hostility Origin t-1	8,732.803**	4,707.196	38,508.930***	-6,629.283
	(3,996.944)	(3,656.484)	(13,786.440)	(10,737.490)
Hostility Dest. t-1	-8,407.009**	-6,259.415**	-41,432.040***	-10,803.070
	(3,466.304)	(3,168.767)	(11,443.290)	(8,855.684)
Landlocked Origin	-2,177.693	-2,173.521		
	(2,147.712)	(1,968.732)		
Landlocked Dest.	-697.347	-1,049.298	72.026	-616.089
	(2,468.240)	(2,268.873)	(6,497.656)	(5,028.126)
Island Origin	1,925.496	-337.542	4,265.385	471.951
	(2,199.268)	(2,076.697)	(5,866.555)	(5,488.335)
Island Dest.	2,492.331	796.798	7,326.611	1,787.460
	(1,852.861)	(1,701.019)	(4,746.467)	(3,719.166)

Contiguity	161,239.400***	141,619.500***	176,173.600***	28,966.130
	(13,788.030)	(12,823.620)	(29,251.360)	(24,229.540)
Colony	325,258.700***	461,263.300***	684,192.900**	789,419.500***
	(105,436.500)	(97,259.270)	(284,448.100)	(219,765.600)
70s Decade Dummy	4,749.118*	4,932.846**	11,933.960*	9,037.153*
	(2,600.529)	(2,375.283)	(6,109.822)	(4,685.337)
80s Decade Dummy	45.307	1,765.332	2,429.479	4,033.562
	(2,765.963)	(2,530.842)	(6,763.305)	(5,191.528)
90s Decade Dummy	-793.664	1,083.832	-2,939.936	1,246.368
	(3,007.192)	(2,755.019)	(7,560.977)	(5,818.597)
Colony:ln GDP Origin t-1	-15,038.800***	-8,604.859**	-46,541.940***	-26,573.290***
	(4,273.218)	(4,086.924)	(12,193.190)	(10,303.100)
Colony:ln GDP Dest. t-1	-21,038.190**	-40,672.680***	-32,060.880	-55,925.140***
	(9,757.569)	(9,109.627)	(26,989.820)	(21,159.690)
Colony:Polity Origin t-1	1,554.022***	284.121	3,903.726**	-3,204.275**
	(596.288)	(551.984)	(1,792.543)	(1,394.126)
ln Distance:ln GDP Dest. t-1	305.715	277.168	-542.290	-1,500.514
	(2,842.494)	(2,609.416)	(6,368.482)	(4,949.026)
Diaspora:ln GDP Origin t-1	0.298***	0.124***	0.307***	-0.040*
	(0.014)	(0.015)	(0.024)	(0.024)
Diaspora:ln GDP Dest. t-1	0.078**	0.101***	0.121*	0.031
	(0.037)	(0.036)	(0.069)	(0.056)
Diaspora:Polity Origin t-1	-0.006***	0.001	-0.008***	0.010***
	(0.002)	(0.002)	(0.003)	(0.002)
Diaspora:Polity Dest. t-1	-0.234***	-0.245***	-0.284***	-0.267***
	(0.016)	(0.015)	(0.030)	(0.023)
Diaspora:ln Distance	0.545***	0.388***	0.462***	0.320***
	(0.012)	(0.013)	(0.025)	(0.021)
Diaspora:Colony	-0.739***	-0.264***	-0.589***	-0.585***
	(0.027)	(0.060)	(0.072)	(0.109)
Diaspora:CSL		-0.763***		0.674***
		(0.093)		(0.149)
Diaspora:COL		-0.088		-0.327***
		(0.069)		(0.110)
Diaspora:CNL		2.732***		6.418***
		(0.098)		(0.205)
Diaspora:LP2		0.124***		-0.344***
		(0.027)		(0.041)
Diaspora:Cultural Distance			0.123***	-0.045**
			(0.023)	(0.020)
Constant	77,484.500	55,236.040	10,211.530	-91,151.600
	(253,601.800)	(232,976.400)	(578,804.600)	(449,817.500)
Observations	4,257	4,257	1,568	1,568
R^2	0.805	0.838	0.823	0.896
Adjusted R^2	0.803	0.836	0.819	0.894
Residual Std. Error	42,892.840 (df = 4226)	39,163.520 (df = 4218)	66,740.350 (df = 1536)	51,163.780 (df = 1528)
F Statistic	580.493*** (df = 30; 4226)	572.118*** (df = 38; 4218)	230.156*** (df = 31; 1536)	339.131*** (df = 39; 1528)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6 Regression Results: Standard Errors Clustered by Country of Destination

	DV: Adjusted Migrant Flow			
	migrant_flow_adj			
	(1)	(2)	(3)	(4)
Diaspora	-4.532***	-2.175***	-4.208***	0.645
	(0.332)	(0.321)	(0.663)	(0.540)
ln Distance	-6,831.228	-4,878.999	4,091.181	10,073.660
	(28,880.160)	(26,520.040)	(64,761.560)	(50,312.800)
CSL		4,209.419		-1,395.934
		(4,841.090)		(15,085.530)
COL		4,125.678*		16,612.890
		(2,469.825)		(10,138.800)
CNL		-51,972.080***		-112,024.400***
		(8,015.897)		(21,034.870)
LP2		-1,107.970		1,123.161
		(1,047.301)		(2,271.125)
Cultural Distance			-2,049.261	-581.357
			(1,849.690)	(1,469.757)
Common Religion	4,417.932	5,676.654	10,179.610	2,897.630
	(3,590.057)	(3,704.969)	(8,557.947)	(7,719.576)
ln GDP Origin t-1	-2,448.553***	-1,155.145	-5,782.720**	-332.456
	(737.608)	(710.112)	(2,354.932)	(1,824.232)
ln GDP Dest. t-1	-3,327.576	-3,505.866	3,480.209	12,304.900
	(24,993.260)	(22,953.710)	(56,980.640)	(44,333.150)
ln Pop. Origin t-1	-0.726	470.421	194.351	979.626
	(460.573)	(422.969)	(1,642.139)	(1,335.690)
ln Pop. Dest. t-1	-314.872	136.108	-1,485.353	1,187.017
	(567.183)	(519.960)	(1,532.819)	(1,193.496)
Polity Origin t-1	62.355	54.199	374.724	54.853
	(112.970)	(105.816)	(287.663)	(233.810)
Polity Dest. t-1	680.992***	432.065**	1,460.978**	769.878*
	(236.261)	(218.004)	(566.512)	(439.350)
Hostility Origin t-1	8,732.803**	4,707.196	38,508.930***	-6,629.283
	(3,996.944)	(3,656.484)	(13,786.440)	(10,737.490)
Hostility Dest. t-1	-8,407.009**	-6,259.415**	-41,432.040***	-10,803.070
	(3,466.304)	(3,168.767)	(11,443.290)	(8,855.684)
Landlocked Origin	-2,177.693	-2,173.521		
	(2,147.712)	(1,968.732)		
Landlocked Dest.	-697.347	-1,049.298	72.026	-616.089
	(2,468.240)	(2,268.873)	(6,497.656)	(5,028.126)
Island Origin	1,925.496	-337.542	4,265.385	471.951
	(2,199.268)	(2,076.697)	(5,866.555)	(5,488.335)
Island Dest.	2,492.331	796.798	7,326.611	1,787.460
	(1,852.861)	(1,701.019)	(4,746.467)	(3,719.166)

Contiguity	161,239.400***	141,619.500***	176,173.600***	28,966.130
	(13,788.030)	(12,823.620)	(29,251.360)	(24,229.540)
Colony	325,258.700***	461,263.300***	684,192.900**	789,419.500***
	(105,436.500)	(97,259.270)	(284,448.100)	(219,765.600)
70s Decade Dummy	4,749.118*	4,932.846**	11,933.960*	9,037.153*
	(2,600.529)	(2,375.283)	(6,109.822)	(4,685.337)
80s Decade Dummy	45.307	1,765.332	2,429.479	4,033.562
	(2,765.963)	(2,530.842)	(6,763.305)	(5,191.528)
90s Decade Dummy	-793.664	1,083.832	-2,939.936	1,246.368
	(3,007.192)	(2,755.019)	(7,560.977)	(5,818.597)
Colony:ln GDP Origin t-1	-15,038.800***	-8,604.859**	-46,541.940***	-26,573.290***
	(4,273.218)	(4,086.924)	(12,193.190)	(10,303.100)
Colony:ln GDP Dest. t-1	-21,038.190**	-40,672.680***	-32,060.880	-55,925.140***
	(9,757.569)	(9,109.627)	(26,989.820)	(21,159.690)
Colony:Polity Origin t-1	1,554.022***	284.121	3,903.726**	-3,204.275**
	(596.288)	(551.984)	(1,792.543)	(1,394.126)
ln Distance:ln GDP Dest. t-1	305.715	277.168	-542.290	-1,500.514
	(2,842.494)	(2,609.416)	(6,368.482)	(4,949.026)
Diaspora:ln GDP Origin t-1	0.298***	0.124***	0.307***	-0.040*
	(0.014)	(0.015)	(0.024)	(0.024)
Diaspora:ln GDP Dest. t-1	0.078**	0.101***	0.121*	0.031
	(0.037)	(0.036)	(0.069)	(0.056)
Diaspora:Polity Origin t-1	-0.006***	0.001	-0.008***	0.010***
	(0.002)	(0.002)	(0.003)	(0.002)
Diaspora:Polity Dest. t-1	-0.234***	-0.245***	-0.284***	-0.267***
	(0.016)	(0.015)	(0.030)	(0.023)
Diaspora:ln Distance	0.545***	0.388***	0.462***	0.320***
	(0.012)	(0.013)	(0.025)	(0.021)
Diaspora:Colony	-0.739***	-0.264***	-0.589***	-0.585***
	(0.027)	(0.060)	(0.072)	(0.109)
Diaspora:CSL		-0.763***		0.674***
		(0.093)		(0.149)
Diaspora:COL		-0.088		-0.327***
		(0.069)		(0.110)
Diaspora:CNL		2.732***		6.418***
		(0.098)		(0.205)
Diaspora:LP2		0.124***		-0.344***
		(0.027)		(0.041)
Diaspora:Cultural Distance			0.123***	-0.045**
			(0.023)	(0.020)
Constant	77,484.500	55,236.040	10,211.530	-91,151.600
	(253,601.800)	(232,976.400)	(578,804.600)	(449,817.500)
Observations	4,257	4,257	1,568	1,568
R^2	0.805	0.838	0.823	0.896
Adjusted R^2	0.803	0.836	0.819	0.894
Residual Std. Error	42,892.840 (df = 4226)	39,163.520 (df = 4218)	66,740.350 (df = 1536)	51,163.780 (df = 1528)
F Statistic	580.493*** (df = 30; 4226)	572.118*** (df = 38; 4218)	230.156*** (df = 31; 1536)	339.131*** (df = 39; 1528)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 7 Regression Results: Standard Errors Clustered by Country Pair

	DV: Adjusted Migrant Flow			
	migrant_flow_adj			
	(1)	(2)	(3)	(4)
Diaspora	-4.532***	-2.175***	-4.208***	0.645
	(0.332)	(0.321)	(0.663)	(0.540)
ln Distance	-6,831.228	-4,878.999	4,091.181	10,073.660
	(28,880.160)	(26,520.040)	(64,761.560)	(50,312.800)
CSL		4,209.419		-1,395.934
		(4,841.090)		(15,085.530)
COL		4,125.678*		16,612.890
		(2,469.825)		(10,138.800)
CNL		-51,972.080***		-112,024.400***
		(8,015.897)		(21,034.870)
LP2		-1,107.970		1,123.161
		(1,047.301)		(2,271.125)
Cultural Distance			-2,049.261	-581.357
			(1,849.690)	(1,469.757)
Common Religion	4,417.932	5,676.654	10,179.610	2,897.630
	(3,590.057)	(3,704.969)	(8,557.947)	(7,719.576)
ln GDP Origin t-1	-2,448.553***	-1,155.145	-5,782.720**	-332.456
	(737.608)	(710.112)	(2,354.932)	(1,824.232)
ln GDP Dest. t-1	-3,327.576	-3,505.866	3,480.209	12,304.900
	(24,993.260)	(22,953.710)	(56,980.640)	(44,333.150)
ln Pop. Origin t-1	-0.726	470.421	194.351	979.626
	(460.573)	(422.969)	(1,642.139)	(1,335.690)
ln Pop. Dest. t-1	-314.872	136.108	-1,485.353	1,187.017
	(567.183)	(519.960)	(1,532.819)	(1,193.496)
Polity Origin t-1	62.355	54.199	374.724	54.853
	(112.970)	(105.816)	(287.663)	(233.810)
Polity Dest. t-1	680.992***	432.065**	1,460.978**	769.878*
	(236.261)	(218.004)	(566.512)	(439.350)
Hostility Origin t-1	8,732.803**	4,707.196	38,508.930***	-6,629.283
	(3,996.944)	(3,656.484)	(13,786.440)	(10,737.490)
Hostility Dest. t-1	-8,407.009**	-6,259.415**	-41,432.040***	-10,803.070
	(3,466.304)	(3,168.767)	(11,443.290)	(8,855.684)
Landlocked Origin	-2,177.693	-2,173.521		
	(2,147.712)	(1,968.732)		
Landlocked Dest.	-697.347	-1,049.298	72.026	-616.089
	(2,468.240)	(2,268.873)	(6,497.656)	(5,028.126)
Island Origin	1,925.496	-337.542	4,265.385	471.951
	(2,199.268)	(2,076.697)	(5,866.555)	(5,488.335)
Island Dest.	2,492.331	796.798	7,326.611	1,787.460
	(1,852.861)	(1,701.019)	(4,746.467)	(3,719.166)

Contiguity	161,239.400***	141,619.500***	176,173.600***	28,966.130
	(13,788.030)	(12,823.620)	(29,251.360)	(24,229.540)
Colony	325,258.700***	461,263.300***	684,192.900**	789,419.500***
	(105,436.500)	(97,259.270)	(284,448.100)	(219,765.600)
70s Decade Dummy	4,749.118*	4,932.846**	11,933.960*	9,037.153*
	(2,600.529)	(2,375.283)	(6,109.822)	(4,685.337)
80s Decade Dummy	45.307	1,765.332	2,429.479	4,033.562
	(2,765.963)	(2,530.842)	(6,763.305)	(5,191.528)
90s Decade Dummy	-793.664	1,083.832	-2,939.936	1,246.368
	(3,007.192)	(2,755.019)	(7,560.977)	(5,818.597)
Colony:ln GDP Origin t-1	-15,038.800***	-8,604.859**	-46,541.940***	-26,573.290***
	(4,273.218)	(4,086.924)	(12,193.190)	(10,303.100)
Colony:ln GDP Dest. t-1	-21,038.190**	-40,672.680***	-32,060.880	-55,925.140***
	(9,757.569)	(9,109.627)	(26,989.820)	(21,159.690)
Colony:Polity Origin t-1	1,554.022***	284.121	3,903.726**	-3,204.275**
	(596.288)	(551.984)	(1,792.543)	(1,394.126)
ln Distance:ln GDP Dest. t-1	305.715	277.168	-542.290	-1,500.514
	(2,842.494)	(2,609.416)	(6,368.482)	(4,949.026)
Diaspora:ln GDP Origin t-1	0.298***	0.124***	0.307***	-0.040*
	(0.014)	(0.015)	(0.024)	(0.024)
Diaspora:ln GDP Dest. t-1	0.078**	0.101***	0.121*	0.031
	(0.037)	(0.036)	(0.069)	(0.056)
Diaspora:Polity Origin t-1	-0.006***	0.001	-0.008***	0.010***
	(0.002)	(0.002)	(0.003)	(0.002)
Diaspora:Polity Dest. t-1	-0.234***	-0.245***	-0.284***	-0.267***
	(0.016)	(0.015)	(0.030)	(0.023)
Diaspora:ln Distance	0.545***	0.388***	0.462***	0.320***
	(0.012)	(0.013)	(0.025)	(0.021)
Diaspora:Colony	-0.739***	-0.264***	-0.589***	-0.585***
	(0.027)	(0.060)	(0.072)	(0.109)
Diaspora:CSL		-0.763***		0.674***
		(0.093)		(0.149)
Diaspora:COL		-0.088		-0.327***
		(0.069)		(0.110)
Diaspora:CNL		2.732***		6.418***
		(0.098)		(0.205)
Diaspora:LP2		0.124***		-0.344***
		(0.027)		(0.041)
Diaspora:Cultural Distance			0.123***	-0.045**
			(0.023)	(0.020)
Constant	77,484.500	55,236.040	10,211.530	-91,151.600
	(253,601.800)	(232,976.400)	(578,804.600)	(449,817.500)
Observations	4,257	4,257	1,568	1,568
R^2	0.805	0.838	0.823	0.896
Adjusted R^2	0.803	0.836	0.819	0.894
Residual Std. Error	42,892.840 (df = 4226)	39,163.520 (df = 4218)	66,740.350 (df = 1536)	51,163.780 (df = 1528)
F Statistic	580.493*** (df = 30; 4226)	572.118*** (df = 38; 4218)	230.156*** (df = 31; 1536)	339.131*** (df = 39; 1528)
Note:				*p<0.1; **p<0.05; ***p<0.01

References

Adserà, A. Language and culture as drivers of migration. IZA World of Labor 2015: 164 doi: 10.15185/izawol.164

Aw, B. Y., Lee, Y., & Vandenbussche, H. (2019). The importance of consumer taste in trade.

Banerjee, A. V., & Duflo, E. (2019). Good economics for hard times: Better answers to our biggest problems.

Beine, M., F. Docquier and Ç. Özden, (2011a). Diasporas. Journal of Economic Development 95(1): 30-41.

Beine, M., F. Docquier and Ç. Özden. (2011b). Diaspora Effects in International Migration. Key Questions and Methodological Issues. World Bank Policy Research Working Paper 5721.

Beine, M., S. Bertoli and J Fernandez-Hueras Moraga. 2014. A practitioner's guide to gravity models of international migration. Mimeo.

Blumer, H. 1958. 'Race Prejudice as a Sense of Group Position', Pacific Sociological Review 1, 3–7

Bobo, L. D. & Tuan, M. 2006. Prejudice in Politics: Group Position, Public Opinion and the Wisconsin Treaty Rights Dispute. Cambridge, MA: Harvard University Press.

Bobo, L. D. 1999. 'Prejudice as Group Position: Microfoundations of a Sociological Approach to Racism and Race Relations', Journal of Social Issues 55, 445–72.

Borjas, George J. 2015. "Immigration and Globalization: A Review Essay." Journal of Economic Literature, 53 (4): 961-74. DOI: 10.1257/jel.53.4.961 (more migrant productivity)

Brewer, M. B. & Brown, R. J. 1998. 'Intergroup Relations', in Gilbert, D. T., Fiske, S. T. &

Lindzey, G., eds, Handbook of Social Psychology, 4th edn. New York: Oxford University Press.

Cavalli-Sforza, L.L., P. Menozzi, and A. Piazza. 1994. The History and Geography of Human Genes (Princeton, NJ: Princeton University Press).

Central Intelligence Agency. (2009). The CIA world factbook 2010. Skyhorse Publishing Inc.

Clemens, M. A. (2011). Economics and Emigration: Trillion-Dollar Bills on the Sidewalk? In H. F. Chang (Ed.), Law and Economics of Immigration (pp. 21–44). Elgar Research Collection. Economic Approaches to Law, vol. 46. Cheltenham, UK and Northampton, MA: Elgar

Clemens, M. A., & Pritchett, L. (2019). The New Economic Case for Migration Restrictions: An Assessment. Journal of Development Economics, 138, 153–164.

Collier, Paul. Exodus: How Migration Is Changing Our World. Oxford, Oxford UP, 2013.

Correlates of War Project. Colonial Contiguity Data, 1816–2016. Version 3.1. 2017.

De Sousa, J., Mayer, T., & Zignago, S. (2012). Market access in global and regional trade. Regional Science and Urban Economics, 42(6), 1037-1052.

Ethnologue: Languages of the World, 16th ed. Summer Institute of Linguistics, International Academic Bookstore, Dallas, TX (available online).

Eurobarometer, S. (2006). Europeans and their Languages. European Commission.

Fearon, James D. (2003). Ethnic and Cultural Diversity by Country, Journal of Economic Growth. 8:195–222.

Giles, M. W. & Evans, A. 1986. 'The Power Approach to Intergroup Hostility', Journal of Conflict Resolution 30, 469–85.

Gurevich, T., & Herman, P. (2018). The dynamic gravity dataset: 1948–2016. US International

Trade Commission, Office of Economics Working Paper.

Gurevich, T., Herman, P., Abbyad, N., Demirkaya, M., Drenski, A., Horowitz, J., & Kenneally, G. (2018). The Dynamic Gravity Dataset: Technical Documentation. https://www.usitc.gov/data/gravity/html_documentation/dynamic_gravity_technical_documentation_v1_00.html#x1-63001r2

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. American economic review, 105(10), 3150-82.

Global Bilateral Migration Database, World Bank Group (2011). https://databank.worldbank.org/source/global-bilateral-migration

Hatton, T. J. and Williamson, J. G. (2005). What fundamentals drive world migration? In Borjas, G. and Crisp, J., editors, Poverty, International Migration and Asylum. Palgrave-Macmillan.

Hofstede, Geert (1980) Culture's consequences: International differences in work-related values. Beverly Hills: Sage Publications.

Jacques Melitz, Farid Toubal, Native language, spoken language, translation and trade, Journal of International Economics, Volume 93, Issue 2, 2014, Pages 351-363, ISSN 0022-1996, https://doi.org/10.1016/j.jinteco.2014.04.004.

Karemera, D., Oguledo, V. I., and Davis, B. (2000). A gravity model analysis of international migration to North America. Applied Economics, 32(13):1745—1755.

Kogut, B., & Singh, H. (1988). The Effect of National Culture on the Choice of Entry Mode. Journal of International Business Studies, 19, 411-432.

Konara, P., & Mohr, A. (2019). Why we should stop using the Kogut and Singh index. Management International Review, 59(3), 335-354.

Koopmans, R., I. Michalowski and S. Waibel. 2012. Citizenship Rights for Immigrants: National Political Processes and Cross-National Convergence in Western Europe, 1980-2008. American Journal of Sociology 117(4):1202-1245.

Marshall, M. G., Gurr, T. R., Davenport, C., & Jaggers, K. (2002). Polity IV, 1800-1999: Comments on Munck and Verkuilen. Comparative Political Studies, 35(1), 40-45.

Mayda, Anna Maria. (2007). International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows. Journal of Population Economics. 23. 1249-1274. 10.1007/s00148-009-0251-x.

Ozden, C. Parsons, M. Schiff and T. L. Walmsley (2011) 'Where on Earth is Everybody? The Evolution of Global Bilateral Migration, 1960-2000', World Bank Economic Review 25(1):12-56

Palmer, G., d'Orazio, V., Kenwick, M., & Lane, M. (2015). The MID4 dataset, 2002–2010: Procedures, coding rules and description. Conflict Management and Peace Science, 32(2), 222-242.

Putnam, R.D. (2007), E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture. Scandinavian Political Studies, 30: 137-174. doi:10.1111/j.1467-9477.2007.00176.x

Putnam, R. D., Leonardi, R., & Nanetti, R. (1993). Making democracy work: Civic traditions in modern Italy. Princeton, N.J: Princeton University Press.

Quillian, L. 1995. 'Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-immigrant and Racial Prejudice in Europe', American Sociological Review 60, 586–611.

Quillian, L. 1996. 'Group Threat and Regional Change in Attitudes towards African

Americans', American Journal of Sociology 102, 816–60 Ravenstein, E. (1885). The laws of migration. Proceedings of the Royal Statistical Society, XLVII(2):167—235.

Shenkar, O. (2012). Beyond cultural distance: Switching to a friction lens in the study of cultural differences. Journal of International Business Studies, 43(1), 12-17.

Spolaore, E. and R. Wacziarg. 2009. The Diffusion of Development. Quarterly Journal of Economics 124 (2): 469–529.

Spolaore, E. and R. Wacziarg. 2013. How Deep are the Roots of Human Development? Journal of Economic Literature 51: 1–45.

Taylor, M. 1998. 'Local Racial/Ethnic Proportions and White Attitudes: Numbers Count', American Sociological Review 63, 512–35.

World Bank, (2016). World Development Indicators. https://data.worldbank.org/data-catalog/world-development-indicators.