# Whither Will They Go?
## A Global Analysis of Refugee Flows, 1955-1995

**Will H. Moore**
Department of Political Science
The Florida State University,
Tallahassee, FL, 32306-2230
will.moore@fsu.edu
http://garnet.acns.fsu.edu/~whmoore/
&
**Stephen M. Shellman**
Department of Government
College of William & Mary
Williamsburg, VA 23187-8795
smshel@wm.edu
http://faculty.wm.edu/smshel/

–DRAFT #1–

**Abstract**

What leads people seeking refuge from violence to select one country over another? We propose an expected utility model to answer that question and focus on fear of persecution, wages, diaspora population, and relocation costs. The model informs a statistical specification that we estimate using a pooled cross-sectional time-series data set containing all directed-dyads over the period from 1955 through 1995. The results suggest that diaspora populations, borders, local violent behavior, and relative wages have a systematic impact on refugee flows across country pairs.

To be presented at the 2004 Annual Meeting of the Midwest Political Science Association, Chicago, 15-18 April. We wish to thank T.K. Ahn, Shaun Bowler, Tom Carsey, Jim Cobbe, Mark Lubell, Lanny Martin, Sara Mitchell, Sol Polachek, Rafi Reuveny, and Mark Souva for helpful discussions, as well as the participants at colloquia at Rice University and Florida State University for their comments and suggestions. This research was partially supported by the National Science Foundation (SES-0099313) and the Council on Research and Creativity at Florida State. We also wish to thank Bela Hovy of the United Nations High Commission on Refugees for sharing with us their data. Brad Ellis, Ahmed Khanani, and Jackie Rubin provided efficient and helpful coding and data management support. We are solely responsible for the views and interpretation of the data reported here.

# 1 Introduction

At various times and locations large numbers of people choose to abandon their property and livelihoods to relocate elsewhere. Known as forced migrants, many of them relocate within their country of origin, but others flee across an international border and invoke the protection of international law as refugees. This study focuses on the latter group and asks: 'What leads people to seek refuge in one country versus another?' In doing so, we are able to address a secondary question that is of considerable interest in industrialized, Western democracies: 'Are refugees pushed by violence or pulled by economic opportunity?' There is broad agreement in the literature that violence is the key factor that pushes people to abandon their homes and livelihoods, regardless of whether they relocate abroad or within their country.[1] Given that (1) international law defines refugees as persons who have a reasonable fear of persecution should they be returned to their country of origin and (2) the standard definition of an internally displaced person is one who has relocated due to a fear of persecution, the finding is not controversial. But we do not yet have studies that examine the choice of location among refugees. Where do they go, or, more generally, what accounts for the variance in their destination choices?

Some information is available in the United Nations High Commissioner for Refugees (UNHCR) database on refugees.[2] Table 1 reports the top ten refugee destinations and the top ten directed-dyadic refugee flows over the period from 1955 through 1995. The most common destinations are countries better known for the armed conflict they have endured: Iran, Pakistan, Zaire, Somalia, and Sudan. To explain this pattern consider that these countries share one or more borders with countries that have experienced considerable armed conflict in the post World War II era. The 'Afghanistan to Iran' and 'Iraq to Iran' directed dyads in the second column of Table 1 underline the point.

[Table 1 about here]

Not surprisingly, the United States and (West) Germany make the list, though perhaps farther down than people living in those countries might have guessed. Eight of the top ten destination countries are on the list because they bordered a country that produced refugees, and violent conflict is more common in the global south than it is in the global north. The UNHCR's figures for 2003 indicate that 27% of the world's roughly 20 million refugees have sought refuge in North America, Europe and Oceania: 73% have sought refuge outside of the industrialized capitalist democracies.[3] Schmeidl's (1998) review of global trends from 1964-96 suggest that this distribution has been roughly constant for three decades, and the UNHCR's 50th anniversary edited volume describes a shift from Europe in the years immediately following World War II to Asia, Africa and Central America (UNHCR 2000).

[1]See, for example, Stanley (1987), Zolberg, Suhrke & Aguayo (1989), Morrison (1993), Gibney, Apodaca & McCann (1996), Schmeidl (1997), Apodaca (1998), Davenport, Moore & Poe (2003), and Neumayer (2003). Borjas (1987:551) finds that political instability in the country of origin affects immigration–not refugee–rates to the United States, and Karemera, Oguledo & Davis (2000) report similar findings for immigration to the US and Canada. We restrict our attention to refugees.

[2]We discuss the data and operationalization below.

[3]Available online at: `http://www.unhcr.ch/cgi-bin/texis/vtx/statistics`. Visited January 2004.

We also see that groups of people seeking refuge from a given conflict will often emigrate to *different* countries; for example, some Afghanis will seek refuge in Pakistan while others will seek refuge in Iraq (see also Gibney, Apodaca & McCann, 1996). Each of these observations is further reinforced by looking at the top 10 directed-dyadic refugee flows by decade, as displayed in Table 2.

[Table 2 about here]

For example, Rwanda sent a large number of refugees to Burundi, Tanzania, Uganda, and Zaire in multiple decades, and Ethiopia sent large numbers of refugees to Sudan and Somalia. Further, the top 40 directed dyads across the four decades are overwhelmingly in the global south and involve bordering countries. Nevertheless, several non-bordering directed dyads appear in Table 2.

This brief review reveals a few stylized facts. First, refugees do not appear to abandon their homes in an effort to maximize their income or employment opportunities. Second, borders and violence appear to play a major role, at least in the largest cases. Finally, not all of the refugees fleeing violence select the same destination. Given that [1] most countries have more than one border, and [2] a non-trivial percentage of refugees seek asylum far from home (e.g., in the United States or Germany), it is important to ask what explains the differences in destination choice. We turn to theory to answer that question.

# 2 Theory and Hypotheses

## 2.1 Fear of Persecution

We are interested in aggregate flows of human beings between countries, and thus anticipate that the aggregate characteristics of countries influence those flows. However, as human beings are the actors (i.e., those who make the decision to seek refuge), we begin by grounding our theory in individual human decision making. More specifically, we put forth a rational choice account that is consistent with earlier work (e.g, Davenport, et al. 2003, Moore & Shellman 2002, 2003). The 1951 United Nations Refugee Convention codifies in international law that a country extend asylum to people who have a reasonable fear of persecution. We can thus think of freedom from persecution as a positive right, and focus on freedom from restrictions on liberty, property, the physical integrity of one's person, and life.[4] To be more explicit, we are thinking about freedom from arbitrary arrest, alienation of property, torture, or killing. The central theoretical task, then, is to specify an individual's fear of persecution function, which is to say, the characteristics of countries that will lead people to leave their country of origin and seek asylum abroad.

Before we can develop a function of an individual's fear of persecution we must identify a process by which people develop beliefs. We place information at the heart of our model; to form beliefs about the probability that they will be victimized, people reference a relevant information set. They update those beliefs (in accord with Bayes' theorem) in response to

[4] This view is consistent with those rights enshrined in the Universal Declaration of Human Rights (UDHR; see `http://www.un.org/Overview/rights.html`).

changes in that information set. Put differently, though we do not present a formal model of the process, we assume that individuals form expectations about the likelihood of becoming a victim of persecution. They then update those expectations by referencing an information set composed of aggregate characteristics of their country.

What characteristics *of a country* will influence the probability that a given person will expect to be a victim of persecution? We conjecture that the relevant information set is composed of the coercive activities of three actors: [1] the state, [2] dissidents, and [3] foreign troops.[5] That is, people monitor the behavior of these groups, and when these groups engage in violence, people update their beliefs accordingly about the prospect that they will become victims of persecution: the relationship between violence and fear of persecution is positive.

Our first claim, then, is that the primary determinant of refugee flows is fear of persecution, which we contend is a function of the violent behavior of three actors. We submit that these variables should have a greater collective impact on refugee flows than any other (sets) of variables. This is not to say that other variables will not have an effect–we strongly suspect that they will. And we turn our attention to those other factors.

## 2.2 Other Determinants of Refugee Flows

Freedom from persecution is surely important, but it is not the only good people pursue. We assume that freedom from persecution is simply an extreme manifestation of what are more general preferences that people hold. In other words, people are not simply interested in avoiding arbitrary arrest, seizure of property, torture, or killing, but have preferences for governing institutions that provide for liberty, the protection of property rights, and the provision of public order. Further, we assume that people are interested in the ability to provide for the material needs of themselves and their families (i.e., wages).[6] As such, we must specify the aggregate observables that we claim people monitor to develop beliefs about the extent to which they will be able to meet these interests in their country.

We submit that the institutions of government are important. That is, different institutions provide different levels of liberty, protection of property rights, and provision of public order. Wages are also important as they form the basis for material survival. That said, we anticipate that these factors will play a lesser role in the decision to seek refuge abroad, and have shown in other work that the violence indicators noted above have a considerably stronger affect on the decision to flee one's home than do indicators of institutions (Moore & Shellman 2003).[7]

[5] See Moore & Shellman (2002, 2003) for more detailed discussion. Davenport, et al. (2003) make a similar argument.

[6] We recognize that this is a universalist approach to human interests that is sometimes criticized as being ethnocentric or Western. We cited the UDHR above because it nicely reflects our view. Donnelly (2003) provides a useful defense of the universalist view. In future studies others might wish to revise the theory by specifying a different set of human interests that are not rooted in the universalist tradition. See Lauren (1998:1-13, 50) for an argument that the concept of universal human rights is not Western.

[7] Note that the decision to flee one's home is not the same as the decision to seek refuge abroad. In Moore & Shellman (2003) we conduct a monadic analysis that only examines the characteristics of the country of origin. In this study we conduct a directed-dyadic analysis which examines a comparison of the

Second, we contend that people place a high value on family, friends and their culture (including language, food, religion, etc.). This is especially important with respect to seeking refuge abroad where one is likely to become a minority in a foreign culture (see Massey 1990 and Faist 2000). Economists also argue that communities of emigrants from country A in country B provide information that reduces the costs of relocation for potential emigrants in country A (e.g., Hatton & Williamson 1998: 14, 38). Finally, many countries have policies that favor the re-unification of families, thus making it easier for direct relatives of refugees to gain refugee status. We thus anticipate that people will respond to information about the size of a diaspora population when deciding whether to seek refuge in a given country. We expect this effect to be substantial: it should be the largest (set of) variable(s) in a statistical model.[8]

Finally, costs of relocation are going to play a role. More specifically, the more costly it is to move from one country to another, the less likely people are to seek asylum. Several factors play a role here. First, distance is important. Second, as technology has changed, travel has become less expensive. Third, some countries are more likely to accept asylum applications than others. Each of these cost factors should influence directed-dyadic refugee flows.

## 2.3 Comparing Country O with Country A

To this point we have developed a list of the concepts that we expect to influence refugee flows. We have not, however, specified how we expect comparisons to be made between two countries such that we might observe a positive flow of individuals from one country seeking refuge in another country. A standard approach in the migration literature is to specify a utility function and assume that individuals compare their expected utility in the two countries and then select the country with the higher level of expected utility. Let $f$ indicate fear of persecution, $w$ indicate wages, and $d$ indicate diaspora population. We assume that utility is a power function of these terms, as indicated in the equations below. Let the subscripts $_O$ and $_A$ indicate the country of origin and the (potential) country of asylum, respectively. We can now write the following utility functions for an individual, $i$:

$$E(U_{iO}) = -f_O^{B_1} + w_O^{B_2} + v_O \tag{1}$$

$$E(U_{iA}) = -f_A^{B_3} + w_A^{B_4} + d_{OA}^{B_5} - c_{OA}^{B_6} + v_O \tag{2}$$

where $v$ is a random variable that is distributed normally with zero mean and represents other factors that influence people's expected utility, and $c_{OA}$ is the cost of moving from country O to country A. The $_{OA}$ subscript on $d$ in equation 2 indicates that this term represents the size of the diaspora of country O in country A. The $B$ terms are parameters that represent the weight that people assign to each term when developing their beliefs about their utility.

characteristics of the country of origin with those of a (potential) country of refuge.

[8]In our monadic study of the decision to flee one's home, stocks of previous forced migrants had a large impact, but not as large as the combined effect of our measures of violence. However, we expect our measure of refugee stock to have a larger effect in the directed-dyadic analysis than measures of violence.

They are nonlinear because we expect higher values of fear, wages, and costs to have an increasing impact on utility (i.e., $B_1, B_2, B_3, B_4, B_6 > 1$). We expect the size of the diaspora population to have marginally diminishing returns to utility (i.e., $0 < B_5 < 1$).

Having written utility functions for the individual, the next step is to aggregate them at the country level. To do so we must make some assumptions about the distribution of the values individuals assign to $f$, $w$, $d$ and $c$ as a function of the information they observe. That is, we argued above that people monitor information to develop beliefs. Expected utility is a summary statement of those beliefs, and the $f$, $w$, $d$ and $c$ terms are the components. If all people in country A responded to the information set[9] by assigning the same values to these components then the entire population of country O would make the same decisions (e.g., seek refuge in country A). And this implication is plainly inconsistent with the empirical record: we observe non-uniform distributions over populations' refuge seeking decisions. As such, it is useful to assume that the distribution of assignments to $f$, $w$, $d$ and $c$ is not uniform with respect to the information set. To be concrete we assume that assignment of values in response to the information set is normally distributed across the population in country O.[10] Taking logs of both sides we can now rewrite equations 1 and 2 as aggregate (or country level) equations:[11]

$$\log E(U_O) = -\beta_1 \log f_O + \beta_2 \log w_O + \log \epsilon_O \tag{3}$$

$$\log E(U_A) = -\beta_3 \log f_A + \beta_4 \log w_A + \beta_5 \log d_{OA} - \beta_6 \log c_{OA} + \log \epsilon_A \tag{4}$$

where the $\beta$ and $\epsilon$ parameters are the mean of the $B$ and $\upsilon$ parameters in equations 1 and 2, respectively.[12]

The next step is to observe that when $E(U_A) > E(U_O)$, we anticipate a non-zero refugee flow from country O to country A. Further, the difference between expected utility in A and O (i.e., $E(U_A) - E(U_O)$) will be positively associated with the size of the refugee flow from O to A. This allows us to write:[13]

$$\begin{aligned} \log R_{OA} \quad = \quad & \beta_3 \log \frac{f_O}{f_A} + \beta_4 \log \frac{w_A}{w_O} + \beta_5 \log d_{OA} - \beta_6 \log c_{OA} \\ & + (\log \epsilon_A - \log \epsilon_O) \end{aligned} \tag{5}$$

where $R_{OA}$ is the refugee flow from country O to country A. One can estimate the parameters of this equation as follows:

$$\log R_{OA} = \alpha + \beta_3 \log \frac{f_O}{f_A} + \beta_4 \log \frac{w_A}{w_O} + \beta_5 \log d_{OA} - \beta_6 \log c_{OA} + \epsilon \tag{6}$$

[9] Recall that we argued above that this information set is composed of violence by [1] three actors, [2] government institutions, [3] wages, and [4] diaspora population.

[10] The key is that they are not uniform–we conjecture that many other distributions can also be assumed.

[11] Note that we have dropped the $_i$ subscript. With respect to the log transformation, recall that $x^b = \log bx$.

[12] Equation 4 represents the aggregate expected utility assuming that a representative sample of people in country O sought refuge in country A.

[13] We write down all of the steps to produce equation 5 in an Appendix.

where $\alpha$ is a constant representing the log refugee flow when all the variables take a value of zero, and $\epsilon = \log \epsilon_A - \log \epsilon_O$. To test the hypotheses, then, we must operationalize the concepts. We turn to that task and other issues of research design in the following section.

# 3 Research Design, Estimation & Measurement

There are a number of inferences we wish to draw. The primary ones concern the hypotheses presented in the preceding section. An issue that we discussed in passing above, however, is that (potential) refuge seekers face a number of options when considering where to seek refuge: it is not a two country issue as our discussion of countries O and A suggests. Thus, we want a research design that can give us some leverage on the substitutes available. More specifically, we need to select a data structure and then create some variables that enable us to draw inferences about the variables in equation 6 in the context of alternatives. We have constructed an annual directed-dyadic data set comprised of all country-pairs and then coded data on (1) whether the two countries share a border, (2) the distance between the two countries' capitals, and (3) the number of countries on which the country of origin borders. We discuss operationalization below. The point here is that the first two indicators get at the cost of relocating, but the third gets at the availability of (relatively) low-cost substitutes. Further, and more importantly, all substitutes (both low cost and high cost) are included in the data. Thus, our design enables us to draw inferences about the impact of the variables of interest given alternative (potential) destinations. The temporal domain of the sample is 1955-1995[14] and the spatial domain is all country-pairs for which we were able to obtain data.

We have a large set of cross-sectional time-series data, and thus confront a number of estimation issues. There has been a veritable explosion in standardized routines to estimate model parameters using cross-sectional time-series data. Since definitive diagnostics to help one determine which of the bevy of approaches is optimal in any given situation have not yet been developed, we select the approach that we believe is best, but conduct an exhaustive set of robustness checks to determine whether the estimates we report are fragile to using a different approach. We summarize those findings in an appendix.

The challenges created by a cross-sectional time-series data structure are relatively easy to state: heteroskedastic errors, serially correlated errors (within and/or across directed-dyads), and missing data (which produce unbalanced panels and might lead to sample selection bias). A number of different estimation approaches that have been developed in the literature to address the first two problems are now readily available in standard statistical packages, and we implement the routines available in release 8 of Stata. We also use imputation, interpolation and extrapolation to fill in missing values, and re-estimate the models reported in the main text using those augmented data sets. Details are provided in the appendix, but the primary conclusion of these robustness checks is that our findings are largely stable to a variety of different statistical models and augmented data sets.

To measure our dependent variable we use a data set collected by the United Nations High Commissioner for Refugees (UNHCR) which records the annual estimated stock of

[14]In two models we include variables that reduce the sample size to 1960-1995 and then 1976-1995.

refugees in a given country who originated in a given other country.[15] Thus, we have an annual, directed-dyadic stock measure. Because flow data do not exist,[16] we use the stock data to create a flow measure. We took the first difference in the stock (i.e., the value in year t minus the value in year t-1) and then truncated it at zero (thereby eliminating the negative values).[17] Then we took the natural log of our flow measure.[18] These data are available from 1955 through 2001.

Most of our other variables are measured for both the country of origin and the country of asylum. The discussion that follows assumes that is the case, unless otherwise noted. In the tables we use the capital letter O to signify that a variable measures the country of origin and the capital letter A to denote measures associated with the country of (potential) asylum.

To measure the coercive behavior of dissidents we employ data from Banks' Cross-National Time-Series Data Archive (CNTS) to create an event count of the number of times dissidents used violence in a given year.[19] More specifically, we use the sum of guerrilla warfare attacks[20] and riots[21] to create our event count indicator of violent dissent. These data are available for 1955 through 1995.

We use two indicators to measure the coercive behavior of the state: a dichotomous indicator of mass killing and a multichotomous indicator of state terror. The dummy variable indicator of genocide and politicide events is due to Barbara Harff and her work on the State Failure Project.[22] Harff's measure is available for the entire domain of our study. The measure of state terror is taken from the Political Terror Scale project which uses content analysis of human rights reports to code a five point scale where higher values are associated with greater violations of the right to the integrity of the person (Gibney & Dalton 1996). The political terror scale is available for the years 1980-2000, and Steven Poe was kind enough to share with us his data for the years 1976-79.[23]

We use the Correlates of War project's measure of civil war (Sarkees 2000) to operationalize the interaction of the state and dissidents.[24] It is a dummy variable that is coded

---

[15]The data were provided by Bela Hovy, Head of the Population Data Unit, Division of Operational Support, UNHCR, Geneva.

[16]See Schmeidl (1998, 2000) and Crisp (2000) for useful discussions of the strengths and weaknesses of the data on forced migration.

[17]It is important to eliminate the negative values that occur when a year of non-zero refugees is followed by a year with zero refugees. If we did not recode those scores to zero, then our variable would record large negative values that are simply the inverse (around zero) of the previous year's value in years when the number of refugees was in fact zero.

[18]Throughout we used the `log` function in Stata which calculates the natural log.

[19]The data are available online at: `http://www.databanks.sitehosting.net/`.

[20]The CNTS project defines this as "Any armed activity, sabotage, or bombings carried on by independent bands of citizens or irregular forces and aimed at the overthrow of the present regime" `http://www.databanks.sitehosting.net/www/var_group.htm#Domestic`.

[21]Riots are defined as "Any violent demonstration or clash of more than 100 citizens involving the use of physical force."

[22]See Harff (2003) and Harff & Gurr (1988, 1996) for discussion. The data are available at the State Failure Project's website: `http://www.cidcm.umd.edu/inscr/stfail/`.

[23]The Political Terror Scale is available as an Excel file at: `http://www.unca.edu/politicalscience/faculty-staff/gibney_docs/pts.xls`.

[24]We used both the extra-systemic and civil war data sets as the extra-systemic data set codes wars

1 for country years with a civil war. To measure the coercive behavior of foreign soldiers we also utilize Correlates of War data, though we needed to determine whether battles took place on the territory of each war participant. The Interstate War data have a value of one for country years in which a country was a participant in a war that produced a minimum of 1,000 battle deaths. We used the *Dictionary of War* and a number of other standard references to determine whether battles occurred on the territory of each participant and recoded the value to zero for participants on whose territory battles did not take place. The Correlates of War data are available for our entire sample.

To operationalize freedom as provided by institutions we turned to the Polity IV data on institutions and utilized the difference of the democracy and autocracy scales. They are 11 point scales based on a handful of component measures of the extent to which political participation is regulated, executive recruitment is open, etc.[25] The difference ranges from -10 to 10.

The polity project assigns missing values for the democracy and autocracy indicators for transition regimes which do not have established polity characteristics. Many of these countries are examples of what has come to be known as 'failed states.' Rather than drop cases from the statistical analysis due to missing data we recoded these missing values to the value 0 and coded a dummy variable that we named 'transition regime' (we assigned it a value of 1 when the democracy and autocracy measures had a missing value, and 0 otherwise). In addition to resolving a missing data problem we used the transition variable as an indicator of the provision of public order. That is, we submit that the absence of authority characteristics that could be coded is a useful proxy of an expectation of a lack of order. The polity data are available for our full sample.

The next concept we need to measure is the existence of institutions that protect property rights. Some direct indicators of property rights have been developed and are used as components of indexes in the Heritage Foundation's Economic Freedom Index[26] and Gwartney & Lawson's (2003) Economic Freedom of the World Index. These data (and their components) are only available for a limited number of cross sections, so instead we use an indirect indicator known as contract intensive money, or CIM (Clague, Keefer, Knack & Olson 1999). CIM is "the ratio of non-currency money to the total money supply, or $(M_2 - C)/M_2$, where $M_2$ is a broad definition of the money supply and $C$ is currency held outside banks" (p. 188). When CIM approaches 1 people have high confidence in the protection of property rights and thus are willing to hold non-cash assets, and when CIM approaches zero, people hold cash rather than non-cash assets, which suggests that they have limited confidence in the protection of property rights. The International Monetary Fund provides the data to construct CIM, which is available from 1960 through the present.[27]

We also need a measure of average wages. The standard measure of average wages in large pooled time series studies of migration is per capita gross national product (e.g., Borjas

between a colonial metropole and its colony.

[25]For details, see the "Authority Characteristics (Component Variables)" entry at `http://www.cidcm.umd.edu/inscr/polity/`.

[26]See `http://www.heritage.org/research/features/index/`.

[27]Mark Souva used the IMF's International Financial Statistics data to construct the CIM indicator and was kind enough to share it with us.

1987, Hatton & Williamson 1998, Karemera, Oguledo & Davis, 2000).[28] To maximize valid observations in our sample we merged GNP data from two sources: the World Bank World Development Indicators and Banks' CNTS.[29] We use the World Bank data and supplement missing observations with the CNTS data. We obtained our measure of population from Fearon and Laitin's (2003) data set on civil wars.[30] These data are available for our full temporal sample.

The next concept for which we need a measure is the size of the diaspora of country O in country A. To operationalize this we use the lagged value of the refugee stock indicator from the UNHCR. This is not a lagged dependent variable, though it does have a functional relationship with the dependent variable. As such, it also serves something of an econometric role as Beck (2001) recommends using a lagged dependent variable to address dynamics in pooled cross-section time-series designs (about which we have more to say in the appendix).

The final concept that we need to measure is the cost of relocation. We develop four indicators to capture this concept. First, we measure the presence/absence of a border in the dyad. If one is walking, riding a bike, or travelling by car/bus/train, then the least costly alternative is bordering countries. We use a data set developed by Shellman (2001) and assign any pair of countries that share a land border or a water border of less than 200 miles a value of one. In addition to the dichotomous measure of a border we also count the number of borders for each country of origin. This allows us to get a measure of the number of low cost substitutes, and we expect the number of borders to have a negative impact on the size of the refugee flow in any given directed dyad.

Because there is variance in the cost of travelling to different countries of (potential) asylum with which the country of origin does not share a border, we need a measure that reflects that variance. The most widely used measure in statistical studies of migration is the distance between the countries, which is often weighted by the average wage (e.g., Borjas 1987, Hatton & Williamson 1998, Karemera, et al. 2000).[31] We used the EUGene (v. 3.03) software to produce the Correlates of War project's measure of distance between capitals[32] and we used the data described above for average wages (i.e., gnp per capita). The distance data are available for all the years in our sample as are the average wages data.[33]

Finally, we include a measure of time to capture increases in technology that have reduced the costs of travel over time. We simply included the variable 'year' as our proxy indicator.

[28] For an analysis of the relationship between GNP and wages, see the US government Import Administration's report on the topic at `http://ia.ita.gov/wages/98wages/98wages.htm`.

[29] The correlation between these two sources is .918 for those observations where both sets have valid data.

[30] Those data are available online at: `http://www.stanford.edu/group/ethnic/publicdata/publicdata.html`.

[31] Weighting by the average wage in the country of origin creates a 'cost per earnings per hour' measure which is comparable across all countries of origin as it represents the amount of time the typical individual in that country of origin would need to spend to earn the cost of migrating. These are known as 'time-equivalent' measures in economics (see Polachek & Siebert 1992: 72).

[32] The EUGene software is available at `http://www.eugenesoftware.org/`.

[33] Might the best way to do this be to recode a variable equal to border 0 when a border exists, 1 when it does not, and then multiply the new variable with distance such that distance only takes a non-zero value when a border does not exist? It wouldn't be an interaction term (i.e., we would not include distance, border, and the interaction of distance and border). But it seems like creating distance that way is superior.

Having sketched the operational indicators and their sources we should observe that equation six specifies that many of the variables are ratios and that all of the variables are logged. Some of our measures are dichotomous (e.g., genocide, civil war), and in those cases rather than take the ratio of two dichotomous variables we included the origin country's value and the (potential) asylum country's value separately. When we use those indicators we expect the signs to be opposite. In addition, we used the `log` function in Stata to take the log. Because the log of zero is undefined, we added a value of one to each variable before taking the log. We did not take the log of the dichotomous measures.

# 4 Findings

We report the results obtained using ordinary least squares with panel corrected standard errors.[34] Estimates for three models are reported in Table 3: one that excludes the measure of contract intensive money (property rights) and political terror (government coercion) and covers the years 1955-95; one that includes contract intensive money but excludes political terror and covers the years 1960-95, but fewer directed-dyads than are included in the first model; and a third that includes all of the measures and covers the years 1976-95.

[Table 3 about here]

We begin with the fear of persecution measures, which we hypothesized would have the largest impact. We have five measures of the violent behavior of the state, dissidents, and foreign soldiers, one of which is a ratio of activity in the country of origin relative to the country of asylum, and the others of which are dichotomous indicators measured in each country. Table 3 shows that the dummy variables measuring activity in the country of origin have a statistically significant impact, but neither the ratio measure nor the dummy variables measuring the activity in the country of asylum have such an effect. We will see below that while the country of origin measures do indeed influence directed-dyadic refugee flows, the findings are inconsistent with our expectation about relative impact.

Consider violent dissent, which cannot be statistically distinguished from zero in models 1 and 2, and has a negative impact in model 3. Our hypothesis leads to the expectation of a positive relationship between the ratio of violent dissent in the origin and (potential) asylum countries, so we fail to find support for the hypothesis. The value of -0.002 in model 3 indicates that a 1% increase in the number of armed attacks/riots produces an expected decrease of 0.2% in refugee flow.[35] Put differently, a change from 10 events to 15 events (a 50% increase) produces only a 1% decrease in expected refugee flow, so the impact is small.[36]

[34] We used the `xtpcse, hetonly` command in version 8 of Stata to produce the estimates. Please see the appendix for a discussion and justification of this choice.

[35] The coefficient for a logged independent variable in a regression where the dependent variable is also logged is the percentage change in the dependent variable caused by a percentage change in the independent variable: $\%\Delta Y = \beta\%\Delta X$ (Wooldridge 1999:45).

[36] We should note that the potential for sample selection bias is likely highest in model 3. Further, this is something of an anomalous finding in that violent dissent has a positive and statistically significant coefficient in four of the six additional models reported in Table 4 (see the appendix) and is not statistically significant

The other measures of violence in the country of origin produce estimates that are consistent with our hypotheses: genocide, government terror, civil war, and war are positively signed and statistically significant. The estimates for genocide and civil war are consistent across the three models, and civil war has a slightly larger impact. With a range of 0.024 to 0.027 moving from an absence of civil war to a presence of civil war in the country of origin increases the expected refugee flow by 2.4% to 2.7% while moving from the absence of genocide to the presence of genocide raises the expected refugee flow by 1.6% to 2.1%.[37] The estimates of foreign soldiers prosecuting war in the country of origin vary across the three models. Model 1 implies a modest effect of 1.1% whereas model 3 implies a considerably larger impact of war in the country of origin: 6.9%.

Government terror has a positive and statistically significant coefficient in model 3. The coefficient of .04 indicates that a small percentage increase in government terror is associated with a four tenths percent increase in refugee flow from O to A. To make it more concrete, a 50% increase in government terror (e.g., a change from 2 to 3 on the scale) will produce a 2% increase in refugee flow. So the impact of government terror is small.

Next, consider the coefficients for violent behavior in the country of (potential) refuge. Table 3 records that none of these variables produce a statistically significant parameter estimate.[38] We thus infer that, ceteris paribus, the violent activity of states, dissidents, and foreign soldiers does not play a role in refugees' decisions about where to flee. This is a marked difference from the impact of violence in the country of origin, and surprised us some: we anticipated that people would be sensitive to avoiding jumping from the frying pan into the fire. But the findings in Table 3 suggest quite strongly that violence in the country of origin, but not in the country of (potential) refuge, has an impact on refugees' decisions.

To summarize, the primary inference is that the coercive behavior of the state, the interaction of dissident and state violence, and the coercive behavior of soldiers fighting in the country of origin have an impact on the size of dyadic refugee flows. The substantive impact of these variables, however, is modest: holding all other variables constant, a country experiencing genocide, a civil war and war on its territory will experience between roughly 5% and 12% more refugees to any given country than will a country experiencing none of these events. Figures 1 through 3 report the percentage change in refugee flow as a function of the variables that were statistically significant. And these figures demonstrate that the impact of these variables is relatively small in comparison to a few of the other indicators.

[Figures 1 through 3 about here]

---

in the other two models. Thus, we view the model 3 finding with some suspicion. Nevertheless, we choose to interpret the coefficients rather than explain away unexpected findings in a post hoc fashion. When we use the augmented data sets we fail to find further support for negative coefficient on violent dissent reported in model 3.

[37]When a continuous independent variable is measured in levels and the dependent variable is measured in logs, the coefficient represents 1/100 of the percentage change of Y in X: $\%\Delta Y = (100\beta X)$. However, this does not hold for dummy variables (Halvorsen & Palmquist 1980). Kennedy (1981) makes a correction to Halvorsen & Palmquist, showing that the formula to calculate the impact of a dummy variable on a logged dependent variable is: $\beta^* = exp(\beta - \frac{1}{2}\hat{V}(\beta)) - 1$, where $\hat{V}$ is the variance of $\beta$.

[38]When we use different statistical models or augmented data sets, some of these variables produced a statistically significant coefficient. See the appendix for details.

We argue that in addition to the expected coercive behavior of actors, fear of persecution is also a function of expected protections of freedom. As such, we also estimated parameters for proxy indicators of institutions that promote freedom, protect property rights, and generate order. Neither the institutional democracy nor the contract intensive money indicators produced statistically significant estimates.[39]

The other indicator, a dichotomous measure of the absence/presence of political institutions (i.e., an institutional transition) in the country of origin and the country of (potential) asylum, is positively signed and statistically significant in model 2 for O and in all three models for A. We expected the sign of O to be positive, but were anticipating a negative sign for A. The coefficient for OTrans in model 2 implies that when the country of origin is undergoing institutional transition the expected number of refugees rises by a modest 1.3%. The estimated parameters for ATrans suggest that the presence of transitory institutions in a (potential) country of asylum increases the expected number of refugees between 0.5% and 1.2%. We did not expect refugees to exhibit a preference–albeit a small preference–for countries that do not have established political institutions, but the finding is quite robust.[40]

An examination of the directed-dyads where the (potential) country of refuge had a transition regime suggests that this result is caused by what Weiner (1996) calls bad neighborhoods. A cursory review of the receiving countries in these directed dyads revealed that Bosnia, Sudan, Burundi, and Afghanistan were typical of those with a transition regime. Conditional cross-tabular analyses (not reported) revealed that among directed-dyads with transition regime present in the (potential) receiving country, those that shared a border were considerably more likely to produce a refugee flow than those that did not. We further probed this result by interacting Atrans with border, and the interaction term was statistically significant (as was border, while ATrans no longer produced a statistically significant estimate. Further, the impact is substantial: the direct effect of a bordering country with a transition regime is a 16% increase in the change of refugee flow (the direct effect of border drops from 8.8% to 7.9%).[41]

Weiner's argument about bad neighborhoods suggests that cross border diffusion processes that we do not model play a role in refugee production, and our transition regime finding is indirectly consistent with such a claim. We thus explored the incidence of some of our measures of violence across borders by looking at more conditional contingency tables. Those analyses (not reported) suggest that civil war has a stronger association with transition regimes across borders than either genocide or war on territory: when one considers only

[39]Surprisingly, the Polity project's democracy measure produced a negatively signed statistically significant coefficient in three of the models reported in Table 4 (the other models produced a statistically non-significant coefficient). In one of the models CIM was significant and negatively signed (in the other models it was not significantly different from zero). Because these measures are scaled such that higher values are associated with positively valued institutions, we created them as the ratio of the value in the country of (potential) asylum to that in the country of origin. As such, the expected sign is positive. While we draw the inference that these variables do not have an impact (as Table 3 implies), we do not have an explanation for these unexpected findings. Neither of the negative signs are sustained when we use the augmented data: the coefficients are positive when statistically significant, or they are nonsignificant.

[40]This holds in Table 4 as well: every model we estimated produced a statistically significant, positively signed coefficient.

[41]We also interacted border with the other measures of the country of (potential) asylum and none of those terms produced statistically significant parameter estimates.

directed-dyads that produced a non-zero flow *and* shared a border, the relative incidence of civil war in the country of origin is considerably higher when the neighboring country has a transition regime than when it does not. That relationship does not hold for either genocide or international war on the territory of the country of origin. Are civil wars more likely in countries that have a neighbor with a transition regime? It turns out that there is a slight tendency for this to be so, but it is neither strong nor statistically significant. Thus, a simple cross-border contagion mechanism via state collapse and civil war does not appear to account for what Weiner called "clusters of countries producing refugees" (p. 26).[42] But it is the case that civil wars in the country of origin are more prevalent in directed-dyads that share a border, have a transition regime in the country of refuge, and produce refugee flows. This suggestive pattern is something we cannot explore further in this study, but it does suggest a potentially fruitful avenue for future research.

Let us turn to the second term in our utility model, wages. The coefficients for the ratio of gnp per capita are statistically significant and suggest that a small percentage increase in the log ratio will produce an expected two one thousandths to one one hundredths percent increase in dyadic refugee flow.[43] A 50% increase in the log ratio of asylum to origin gnp per capita produces a very small one to five tenths percent increase in expected refugee flow from the origin to (potential) asylum country.[44] Thus, while average wages have an impact, their impact is small, even relative to the modest effects of the violence indicators. Thus, while these findings demonstrate that the average refugee is not indifferent to wage differentials, they cast substantial doubt on any claims that the average refugee is primarily seeking greater economic opportunity: violence in the country of origin has a substantially larger impact on refugee flows than wage differentials.

Unlike the variables we have considered thus far, the size of the diaspora in the (potential) country of asylum has a large substantive impact, as Figure 3 depicts. The estimated parameters are statistically significant and their values imply that a small percentage increase in the stock of refugees produces an expected three tenths percent increase in refugee flow.[45] A 50% increase in the refugee stock[46] leads to an expected 15% increase in refugee flow, which is considerably larger than the other variables we have considered.

The final four variables measure the transaction costs of relocation. Border, which measures the presence of a land border or a water border of less than 200 miles produces a statistically significant coefficient in all three models. Time, which proxies improvement in travel technology, has a statistically significant coefficient in models 1 and 2, and both the

[42] A logit model using civil war in the origin country as the dependent variable and civil war in the (potential) country of origin, border, and the interaction of the two produces a statistically significant estimate for the interaction term only. This provides some evidence that it might prove useful to explore the contagion of civil war across borders in future work on refugee flows.

[43] GNP per capita is statistically significant in each of the additional models we estimated–see Table 4.

[44] To provide a concrete example, consider Mexico as the country of origin and Chile and Peru as the countries of (potential) asylum. Using 1998 World Bank data, the Brazil ($4,610) Mexico ($4,020) directed dyad had a 0.059 log ratio whereas the Chile ($4,890) Mexico ($4,020) directed dyad had a 0.085 log ratio, which is almost 50% greater.

[45] This variable has a statistically significant coefficient in each of the alternative models, as recorded in Table 4.

[46] FIND AN EXAMPLE OF TWO DIRECTED DYADS, ONE OF WHICH HAS A 50% LARGER VALUE ON O2ASTOCK.

number of borders–which is a proxy measure for the number of substitutes–and cost–our time equivalent measure of distance–have a statistically significant coefficient in model 1 only. The number of borders, cost, and time have minor effects on the expected number of refugees. A 50% increase in the number of borders[47] leads to an expected decrease of 0.02%, and a 50% increase in time equivalent distance[48] produces only a 0.05% drop in expected refugee flow, holding all other variables constant. To interpret the coefficient for time, consider that the difference between 1955 (1975) and 1975 (1995) is 1%. Thus, a difference of 20 years leads to a small ˜0.4% increase in expected refugee flow. Border, on the other hand, has a sizeable effect. A directed dyad with a land or water border has between a 9% and 13% increase in the expected number of refugees.

## 4.1 Comparison with Extant Studies

While our study is the first to use a directed-dyadic design with a global sample of data,[49] there a handful of existing studies that report relevant findings. As such, we briefly compare our findings with those that can already be found in the literature.

There are a handful of large-n statistical studies of refugee flows, but only two that bear directly on our findings. Gibney, et al. (1996) and Apodaca (1998) limit their analysis to countries that produced refugees, and Schmeidl (1997) focuses her analysis on countries from the developing world. More importantly, both studies consider only the characteristics of the countries of origin, and thus their findings are not directly comparable to ours. That said, all three studies report that human rights abuses have an positive impact on refugee production.

Davenport, et al. (2003) examine net refugee flows from all countries over the period 1964-1989. By including both forced emigrants and forced immigrants[50] in their dependent variable, their estimated coefficients represent the extent to which country of origin characteristics push people out *and* pull people in. Neumayer (2003, 2004) examines asylum applications in Western Europe during the 1980s and 1990s.[51] Unlike the other studies, his focuses on the characteristics of the countries of asylum. One of the studies focuses exclusively on the countries of asylum (2004), the other is directed-dyadic and thus includes information on both the sending and receiving country (2003). That study is the most directly relevant to our own, though it focuses on Europe alone.

Davenport, et al. (2003) report that the violent behavior of the state and dissidents have substantial impacts on forced migration, but the relative size of the economy (i.e.,

[47]For example, we might compare Benin, which has four borders, to Chad, which has six.

[48]FIND TWO DIRECTED DYADS TO COMPARE WITH A 50% DIFFERENCE IN THE DISTANCE/GNP/CAP MEASURE

[49]As Davenport, et al. (2003:36-7) explain, a research design that includes information on both the country of origin and the country of (potential) refuge is needed to really get at what they call push and pull factors of forced migration. Put differently, to draw inferences regarding the question asked in this study, a directed-dyadic cross-sectional time-series design is superior to the country-year cross-sectional time-series design used in other studies.

[50]Their measure also includes internally displaced persons, or IDPs.

[51]Schneider & Carey (nd), which is an English language presentation of Schneider, Holzer & Carey (2001), study asylum applications in Germany and Switzerland.

per capita GNP) does not. Neumayer (2003) observes that his studies are the first to examine asylum destination choice across a large number of countries, and infers that the 'economic migrants' or 'refugees fleeing violence' dichotomy is inadequate: asylum applicants to Western European countries both flee violence and are drawn to the promise of high wages.

A handful of findings emerge from these studies. First, refugees flee violence (Gibney, et al., Schmeidl, Apodaca, Davenport, et al., Neumayer). Second, there is mixed support for the impact of economic pull factors on refugee flows (Schmeidl, Davenport et al., Neumayer). Third, the studies that include a measure of diaspora populations find support for that variable (Schmeidl, Davenport, et al., Neumayer). We discuss our results in the context of each of these findings.

Our analysis adds more evidence to the impact of local violence. One interesting contrast with Neumayer (2003) is worth noting. His study reports that genocide in the country of origin does not have an impact on asylum applications in Europe. He uses the same indicator that we use, and we find that genocide has an impact on refugee flows. Taken together these findings suggest that genocide pushes people to seek refuge across the border, not in far off countries. To explore this further we interacted border and OGenocide, and reestimated model 1 in Table 3 with the product term and the two components. Ogenocide is no longer statistically significant, and both border and the product term are. That result further implies that genocides produce local refugee flight.

More generally, we can say more than previous studies: the violent behavior of states, dissidents and foreign soldiers in the country of origin have an impact, but the violent behavior of these actors in the country of (potential) asylum does not have an effect. This finding strengthens the claim that local violence pushes people out. That is, our findings suggest that those who seek refuge are more concerned about getting out of the fire than they are about jumping into the frying pan.[52]

Second, unlike previous studies our research design affords a much more direct analysis of the impact of economic pull factors than previous studies. Neumayer (2004) examines economic conditions in European countries as pull conditions, but does not include the economic conditions in the countries of origin. That is he focuses on pull factors without considering push factors. Neumayer (2003), on the other hand, only examines economic conditions in the country of origin: pull factors are ignored. Our study is the first to examine both push and pull using a directed-dyadic design. Further, Neumayer's studies use a sample of the 30% of refugees who seek asylum in Europe, and we study all country pairs.[53] Neumayer also uses per capita GDP as a proxy for wages, and he finds that it has the expected relationships: country of origin GDP has a negative impact (2003) and country of asylum GDP has a positive impact (2004) on asylum applications to European countries. Because his studies examine origin and destination wages in different equations, we cannot directly

[52]To be sure, we cannot make too much of this claim. For example, a Mayan fleeing Guatemala in the 1980s to seek refuge in Nicaragua might have good reason to believe that she is unlikely to be victimized in Nicaragua, despite the civil violence in that country. A study like ours cannot get at such fine grained distinction, but time-series case studies like Stanley (1987) or Morrison (1993) could do so.

[53]While Neumayer's decision to focus on European countries is defensible given his interest in the debate about asylum seekers in Europe, his estimates are biased vis-à-vis the population of people who seek refuge. Specifically, one would expect that refugees filing for asylum in Europe are more likely to be motivated by economic concerns than those who simply crossed the nearest border.

compare his findings with ours to determine whether relative wages have a larger impact on those who sought refuge in Europe versus those who sought refuge elsewhere. That said, Neumayer reports that a 1% increase in origin GDP per capita produces a 1% decrease in asylum applications (2003), whereas a 1% increase in destination GDP per capita produces a 7% increase in the share of applications across European countries (2004).[54] We re-estimated model 1 in Table 3 replacing our ratio of asylum to origin GNP per cpaita measure with the separate origin and asylum measures. The origin country measure is statistically significant and suggests that a 1% increase in GNP per capita produces a 0.7% decrease in refugee flow. The (potential) asylum country measure is not statistically significant. Thus, when we include all country pairs, the evidence suggests that country of (potential) asylum wages do not have an impact. But if we study only pull factors and focus on Europe as a destination, then wages have an impact. More research that take into account relative wages (as in our study) and examines distinct samples is warranted to definitively sort out these effects.

Finally, diaspora populations play a large role in the destination choice of people seeking refuge. Our study reports a large effect, and previous studies have also reported that similar measures of diaspora populations have the largest substantive effects in their models. Neumayer (2003) is the most relevant of existing studies, and he finds that a 1% increase in the moving average of the stock of asylum seekers over the preceding five years produces a 9% increase in asylum applications. While not directly comparable, our finding is roughly similar.

# 5 Conclusion: What Remains to be Done?

To recapitulate, our study asks why people seek refuge in some countries rather than others. We use an expected utility model that focuses attention on fear of persecution, wages, culture, and the costs of relocation. Our findings suggest all four of these factors play a role. More specifically, while the violent behavior of states, dissident and foreign soldiers in the country of origin push people to seek refuge abroad, refugees do not discriminate among destinations on the basis of the behavior of these actors in (potential) countries of asylum. Further, we found little support for the hypothesis that institutions that protect freedoms play a role in destination choices. Wages, culture, and costs, on the other hand, play a role. Diaspora population has the largest substantive impact. Relative wages, while significant, have a smaller impact than violence in the country of origin. And transaction costs, especially as measured by the presence of a border, have large substantive effects as well. Thus, our study finds that people have a strong tendency to (1) go where others have gone before them, and (2) to the nearest location where they can avoid the violence. Put differently, refugees flee violence and their destination choice is overwhelmingly local and where others like them have gone in the past. Relative wages are not irrelevant, but their impact is small.

That said, we hope that this will be the first of many global studies of the destination choices of forced migrants, not the last word on the subject. As such, we close with a brief

[54]The dependent variable in Neumayer (2004:14) is "the share of asylum seekers coming to the destination country from a country of origin relative to the total number of asylum seekers in Western Europe, normalised by population size."

discussion of ways in which our study can be extended and/or improved. We consider three opportunities.

First, global sample PCSTS studies like this produce average effects. While average effects are important, policy makers understandably have limited interest in such findings. And an understanding of the destination choices of forced migrants has policy implications: contingency planning could be well served by a model capable of producing serviceable out-of-sample forecasts. And a global model such as this one is a poor candidate for such a model. However, it can serve as a foundation upon which to construct such a model. The first step would be to conduct time-series case studies.

Stanley (1987) studies Salvadoran refugee flows to the United States and Morrison (1993) examines IDP movement in Guatemala. While our study illuminates the average effect of these variables, we don't know whether these effects represent a broad spectrum of 'typical cases' or if cases are actually clustered in such a way that few actual cases exhibit this particular constellation of relative effects.[55] Time-series case studies also make it possible to leverage the often considerable descriptive-historical literature on the case. In addition, superior measurement is possible in time-series case studies. And these issues are important if we are to contribute useful tools for contingency planners.

Another set of extensions involve the exploration of more complex models that endogenize the violent behavior that we treat as exogenous here. In our brief digression into Weiner's notion of 'bad neighborhoods' we argue that violent activity may diffuse across borders. This is, of course, something of a conventional wisdom. And while constructing theoretically informed models that endogenize such conflict processes is a far-from-trivial task that implies multiple equation modelling, it is nevertheless a frontier that beckons.

Another interesting possibility to explore in future work concerns whether refugee flows are a step function of violence in the country of origin such that the other variables in the model only come into play once some threshold of violent behavior in the country of origin is crossed.[56] In other words, it would be interesting to treat violence in the country of origin as a necessary condition and the other variables as conditional such that they are only operative when the necessary condition has been met. We are somewhat dubious about the likely success of such a research project–necessary conditions are knotty. But it would certainly be an important project were it to prove successful.

A final potentially interesting direction for future work is to parse the impact of the size of the diaspora in the (potential) asylum country. This variable not only varies considerably across directed dyads, but will grow over time in any given directed dyad *once a nontrivial number of refugees becomes established in the country of refuge*. Because we have focused on the estimated coefficients, which represent the mean effects across a sample of directed dyads, we cannot say much of interest about this process: few directed dyads are likely represented by these coefficients given the temporal and cross-directed dyad variance in the variable. Future work should probe this more, perhaps conducting time series case studies

[55] One potentially fruitful approach would be to use the replication data set to identify outliers and other influential cases on the results reported here, and then conduct time-series case studies on those directed-dyads.

[56] And one need not think only of intensity thresholds. Constellations of actors pursuing specific behaviors might serve as a threshold.

along the lines of those reported in Hatton & Williamson's (1998) study of migration.

These are just some ideas–readers will surely think of others. We look forward to seeing what future work on this topic can teach us.

# 6 Appendix

In this appendix we walk through the algebra required to produce equation 5 and also discuss the statistical robustness checks that we preformed.

## 6.1 The Algebra behind Equation 5

Assume that:

$$R_{OA} = E(U_A) - E(U_O) \tag{7}$$

We can rewrite equation 7 by replacing the expected utility terms with the right hand side of equations 3 and 4, yielding (observe that equations 3 and 4 are written as log expected utility):

$$\begin{aligned} \log R_{OA} &= (-\beta_3 \log f_A + \beta_4 \log w_A + \beta_5 \log d_{OA} - \beta_6 \log c_{OA} + \log \epsilon_A) - \\ &\quad (-\beta_1 \log f_O + \beta_2 \log w_O + \log \epsilon_O) \end{aligned} \tag{8}$$

Collect like terms:

$$\begin{aligned} \log R_{OA} &= (\beta_3 \log f_O - \beta_1 \log f_A) + (\beta_4 \log w_A - \beta_2 \log w_O) \\ &\quad + \beta_5 \log d_{OA} - \beta_6 \log c_{OA} + (\log \epsilon_A - \log \epsilon_O) \end{aligned} \tag{9}$$

Assume that $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$[57] which yields:

$$\begin{aligned} \log R_{OA} &= \beta_3(\log f_O - \log f_A) + \beta_4(\log w_A - \log w_O) \\ &\quad + \beta_5 \log d_{OA} - \beta_6 \log c_{OA} + (\log \epsilon_A - \log \epsilon_O) \end{aligned} \tag{10}$$

Since the difference of logs is the log of the ratio (i.e., $\log x - \log y = \log \frac{x}{y}$), we can rewrite equation 10 as:

$$\begin{aligned} \log R_{OA} &= \beta_3 \log \frac{f_O}{f_A} + \beta_4 \log \frac{w_A}{w_O} + \beta_5 \log d_{OA} \\ &\quad - \beta_6 \log c_{OA} + (\log \epsilon_A - \log \epsilon_O) \end{aligned} \tag{11}$$

This is equation 5.

[57] This assumption states that when developing beliefs the average individual will assign the same weights to each term in both countries. In other words, we are assuming that people in country A will **not** weigh fear, wages, or diaspora population differently when calculating their expected utility of living in country A and living in country B. This is a very natural assumption: it would be odd to assume that the typical person will assume that s/he will weight values differently if s/he lives in one country versus another.

## 6.2 Statistical Robustness Inquiries

As discussed above, to determine whether our findings were fragile to the model we selected we estimated the parameters of our variables using a wide array of the routines available in version 8 of Stata. We also augmented the data to evaluate the extent to which the data suffer from plausible sample selection bias. We begin with a brief discussion of model selection and a summary of the results obtained using different models. Then we describe the three augmented data sets we created and summarize the results we obtained when using the augmented data.

### 6.2.1 Alternative TSCS Models

An issue that arises when estimating pooled time-series cross-section (TSCS) data is whether to (1) assume that there are no unit effects (i.e., assume directed-dyadic homogeneity in our study), or (2) model unit effects using either a fixed effects or a random effects model. Homogeneity is the assumption that the process being modelled operates the same in all cases. In our study this assumption states that the process that produces refugee flows from Viet Nam to the USA is precisely the same as the process that produces refugee flows from El Salvador to Nicaragua, which is the same as that for Ghana to Pakistan, etc. That is, given the specification of our statistical model, the only differences across directed-dyads are random differences that are adequately modelled by the error term. As King (2001) observes, if we fail to include other relevant variables, then our estimates will suffer from omitted variable bias. Green, Kim & Yoon (2001) suggest that the homogeneity assumption is especially strong in TSCS in comparative politics and international relations data sets such as the one we employ in this study. They suggest that scholars employ fixed effects to remedy the situation, but Beck & Katz (2001) respond by noting that fixed effects only make sense in rather specific situations, and argue that the fixed effects remedy will often do more damage than the problem it is intended to solve. In his review of the debate King (2001:504) summarizes the situation this way: "we have a problem and no solution." We certainly cannot settle the issue in an applied paper on refugee flows. What we can do, however, is explain why we chose to report the results in Table 3 and, in so doing, describe why we eschewed other options. Having done so, we will briefly summarize the extent to which the results reported are robust to using different approaches.

We reported panel corrected standard errors in Table 3 largely due to the case presented in Beck and Katz's work (1995, 2001; Beck 2001). They show that OLS with panel corrected standard errors produces more conservative estimates of the standard errors than the feasible GLS approach which was popular prior to the publication of their 1995 paper. Beck (2001) councils the use of a lagged dependent variable (or a single equation error correction model) to model temporal dynamics. Because we include the lag of the stock of refugees as a substantive variable of theoretical interest, we do not include the lag of refugee flow.

We should observe that the TSCS data are useful (relative to cross-sectional or time-series only) to the extent that the data are informative (i.e., have variance both across cases and over time within cases). To the extent that this is not so (e.g., that the data are constant or sticky within cases over time or across cases), the TSCS design provides little information (see King 2001: 502, Beck & Katz 2001). Further, as Beck (2001:274) explains, "TSCS methods are justified by asymptotics in T and typically require a large T to be useful," which is to say that it is important to have a minimum number of temporal observations for each case. As a rule of thumb, Beck suggests that one should be suspicious of estimates that exploit TSCS data and methods and have fewer than 10 temporal observations per case. Our data has a heterogeneous T: for some directed-dyads T=45 (i.e., we have all but one observation for each year from 1955-95), but for some T=1. Put differently, our data are temporally unbalanced, and this should not be surprising given that some countries disappear while others come into being. If we were to select cases that fit a balanced panel, we would exclude countries that cease to exist (e.g., South Vietnam) as well as those that come into being (e.g., Rwanda). Given that (1) many refugee episodes emerge in countries that were formerly colonies or no longer exist, and (2) refugee episodes are relatively rare, excluding cases to create a balanced panel would almost surely introduce selection bias (and at a minimum would throw out information). As such, we proceeded with an unbalanced panel.

The unbalanced panel is suboptimal, however, with respect to estimating panel corrected standard errors. The standard assumption about the errors is that they are (1) heteroskedastic across panels (i.e., each directed-dyad has its own variance) and (2) contemporaneously correlated across panels (i.e., the errors from one directed-dyad are correlated with the errors from other directed-dyads). Unfortunately, one must have balanced panels (i.e., the same number of temporal observations for each directed-dyad) to relax both assumptions and correct the standard errors accordingly. Rather than throw away observations to balance the panels we assume that the error variances are constant within each directed-dyad, but heteroskedastic across directed-dyads. The `xtpcse` command in version 8 of Stata allows one to impose this assumption by specifying the `hetonly` option, and this model does not require balanced panels. These are the results reported in Table 3.

Alternatively, one can specify the `independent` option with `xtpcse` and thereby assume that the errors are homoskedastic across panels, and allow the error variances to vary within each directed-dyad. We report the sign and statistical significance of the parameters using that model in Table 4. The results do not change when we use the `independent` option rather than the `hetonly` option.

[Table 4 about here]

Table 4 indicates that there are some changes in the estimates when we use a different model. For example, the ratio of violent dissent produces a statistically significant, positively signed parameter in four of the six additional models. That finding is consistent with our expectations. The negative sign on the ratio of democracy in three of the models is inconsistent with our expectations: it implies that people are less likely to leave a country with a relatively low democracy score for one with a relatively high score. The predominant inference is a null finding, but if a negative sign is the proper inference, then it might imply an unmodelled endogeneity: democracies control their borders better than non-democracies, they are clustered such that they are considerable less likely to border non-democracies, etc. Outside of the three democracy estimates, however, the major implication of Table 4 is the similarity of the estimates. While there are a few changes, there is considerable stability.

Further, with the exception of the democracy variable, the other changes are consistent with our expectations.

### 6.2.2 Sample Selection Bias and Augmented Data

When cases are excluded from analysis because they are missing data then the sample suffers from a biased selection mechanism. This will produce inferential problems if the process that produces missing data is related to one or more of the independent variables. The ideal procedure for investigating this issue is to produce a model of the missing data process. While doing so might be feasible, it is a nontrivial task, and we have opted instead for a second best approach: augmenting the data to eliminate missing cases. That is, we estimate the values of the missing data, and then re-estimate the models with the augmented data. To the extent that our parameter and standard error estimates do not change when we use the augmented data we increase our confidence in our inferences vis-à-vis sample selection bias.

We use three different approaches to augment the data: interpolation, extrapolation, and imputation.[58] Each of these approaches rely on rather strong assumptions to estimate the missing values. To describe them it is useful to consider a time series, $x_{t1}, x_{t2}, \ldots x_{tn}$. Interpolation fills in missing values for all observations missing data between non-missing observations. That is, if we have non-missing observations for $x_{t2} \ldots x_{t8}$ and $x_{t11} \ldots x_{t15}$, then interpolation will produce estimates for $x_{t9}$ and $x_{t10}$, but not for $x_{t1}$. Extrapolation, on the other hand, fills in missing values both between and beyond non-missing observations. Thus, if the series ran from $x_{t1}$ to $x_{t20}$, then extrapolation will produce estimates for $x_{t9}$ and $x_{t10}$, but for $x_{t1}$ and $x_{t16} ... x_{t20}$ as well. Both interpolation and extrapolation assume a linear trend.

Rather than assume a linear trend imputation relies on regression analysis to produce estimates. That is, the researcher specifies a regression, which is estimated using the cases for which there are non-missing observations.[59] Then those parameter estimates, and the values of the non-missing independent variables, are used to produce an estimate of the missing value for the variable being modelled. Thus, interpolation/extrapolation imposes a strong assumption of linearity whereas imputation is more data driven (though theory presumably plays some role in specification). And the augmented data are only as good as these assumptions.

That said, we confine our exploration of robustness of parameter estimates to two issues: the sign and statistical significance of the parameter estimate.[60] To the extent that using a different estimation approach does not change the sign and/or statistical significance of the

[58] We used the `ipolate` command in Stata v. 8 to create the interpolated and extrapolated data, and the `impute` command to create the imputed data.

[59] The regressions we used to impute the missing data are available as part of the replication data set that we will deposit with the ICPSR.

[60] Rather than simply consider the sign of the parameter it would be preferable to consider the relative size of the estimate. For example, one might evaluate whether the variables have the same rank order with respect to their impact on the dependent variable. Unfortunately, given the number of alternative models we estimated, space considerations prohibit such an exercise. However, the replication data set will make it possible for people to reproduce all of the parameters and examine relative size.

estimate, we determine that our findings are robust.

The interpolated data set adds the fewest observations: sample sizes increase from 575,300 to 577,892 (model 1), 317,176 to 328,893 (model 2), and 157,149 to 171,028. As such, one would expect the results to be strongly robust, and indeed they are: not a single parameter estimate changes with respect to statistical significance or sign.

Extrapolation adds more cases than interpolation: model 1 increases from 575,330 to 666,838, model 2 from 317,176 to 501,096, and model 3 from 157,149 to 478,926. These are substantial increase, especially for models 2 and 3 where 37% and 67% of the cases, respectively, include estimated data for one or more variables. Given the large percentage of cases that contain data estimated via strong assumptions, one does not want to put too much stock in these findings (e.g., report them in lieu of the results obtained in Table 3).

Are the results markedly different when we use the data augmented via extrapolation? In all three models war on the territory of the country of origin is no longer statistically significant. Genocide in the (potential) country of origin and, a transition regime in the country of origin become statistically significant (with the expected sign) in models 1 and 3. The ratio of democracy becomes significant (with the expected sign) in model 2, and the ratio of GNP per capita becomes significant (with the expected sign) in model 3. The number of borders is no longer statistically significant (model 1), and the ratio of violent dissent is non-significant in model 3. As such, there were no sign changes and none of the other variables changed significance levels. Thus, there are some minor changes from the results reported in Table 3, but the majority of results are unchanged. Further, those changes that did occur were consistent with our hypotheses.

Imputation increases our sample size to 860,783 for all three models. In these data sets most variables contain substantial numbers of estimated observations.[61] Those variables that drive down sample size in Table 3 have considerably larger estimated values than real values.[62] Thus, while this data set clearly does not suffer from sample selection bias, it is composed of a large number of estimated values.

Do the results change substantially when we use the imputed data? With the exception of war on the territory of the country of origin (OWar), all of the variables that were statistically significant remain so (no signs changed). However, a handful of variable that were not statistically significant become so: civil war and war on territory in the country of (potential) asylum; the ratio of democracy; and the presence of a transition regime in the country of origin (OTrans) become statistically significant in all three models.[63] Each of the variables has a positive sign, though we expected a negative sign for the civil war and war on territory indicators.[64] In addition, in models 2 and 3 both the number of borders and our measure of relocation costs produced statistically significant parameter estimates (and the expected sign). Finally, the time measure is statistically significant in model 3 (and has the expected sign).

While one variable's estimate becomes non-significant and a number of other variables

[61] The exceptions are the dummy event variables for genocide, civil war, etc.

[62] Those variables are 'government terror' and 'CIM.'

[63] The difference among the three models is minor given that the sample size is the same.

[64] The result suggests that refugees are drawn toward countries experiencing civil war or foreign soldiers fighting on their territory. Democracy and CIM have the expected sign.

become statistically significant, we submit that the results using the imputed demonstrate that most of our findings are quite robust. Despite changing the sample size rather dramatically, the majority of the findings reported in Table 3 hold. The major challenge to the findings in table 3 are the results for civil war and war on territory in the (potential) country of asylum. That is, with the execption of those two parameter estimates, the results obtained using the imputed data are more consistent with the hypotheses drawn from our model than those reported in Table 3.

What commonalities exist across the data sets? The most consistent finding is that foreign soldiers in the country of origin does not have a significant impact on refugee flows. In addition,

Taken as a whole the results from the augmented data sets do not raise major concerns about sample selection bias. To be sure, we have not demonstrated that sample selection bias does not play a role in any of our inferences. But that is an unrealistic standard to meet. Rather, we have met a lower, but reasonable, standard: we have shown that when we augment our data with estimated values for missing observations that the majority of our inferences are unaffected.

Table 1: Top 10 Refugee Flows, 1955-95

| Rank | Destination | Directed Dyad |
|---|---|---|
| 1 | Iran | Afghanistan → Pakistan |
| 2 | Pakistan | Afghanistan → Iran |
| 3 | Zaire | Ethiopia → Somalia |
| 4 | Somalia | Europe[a] → (W) Germany |
| 5 | Sudan | Rwanda → Zaire |
| 6 | United States | Iraq → Iran |
| 7 | (W) Germany | China → Hong Kong |
| 8 | Tanzania | Mozambique → Malawi |
| 9 | Hong Kong | Ethiopia → Sudan |
| 10 | Ethiopia | Indochina[a] → United States |

*a. Some entries in the UNHCR statistics list a region rather than a specific country of origin* .

Table 2: Top 10 Refugee Flows By Decade

| Rank | 1955-1964 | 1965-1974 | 1975-1984 | 1985-1994 |
|---|---|---|---|---|
| 1 | China → Hong Kong | Indochina[a]→ Indochina[a] | Afghanistan → Pakistan | Afghanistan → Iran |
| 2 | Angola → Zaire | Angola → Zaire | Ethiopia → Somalia | Europe → (W) Germany |
| 3 | Unknown[b]→ France | China → Hong Kong | Afghanistan → Iran | Rwanda → Zaire |
| 4 | China → Macau | Burundi → Tanzania | Indochina[a]→ United States | Iraq → Iran |
| 5 | Rwanda → Uganda | Unknown[b]→ Brazil | Ethiopia → Sudan | Mozambique → Malawi |
| 6 | Rwanda → Burundi | Guinea Bissau → Senegal | Guinea → Ivory Coast | Afghanistan → Pakistan |
| 7 | Rwanda → Zaire | Sudan → Uganda | Indochina[a]→ Australia | Rwanda → Tanzania |
| 8 | Tibet → India | Mozambique → Tanzania | North Vietnam → China | Somalia → Ethiopia |
| 9 | Zaire → Uganda | Ethiopia → Sudan | Angola → Zaire | Liberia → Guinea |
| 10 | Unknown[b]→ Canada | Sudan → Zaire | Zaire → Angola | Liberia → Ivory Coast |

*a. Some entries in the UNHCR statistics list a region rather than a specific country of origin.*

*b. Some entries in the UNHCR statistics list the country of origin as 'Unknown'.*

**Table 3: Refugee Flows, 1955-95**

| | ViolDiss | OGenoc | Gov Terr | OCivWar | OWar | AGenoc | ACivWar | AWar |
|---|---|---|---|---|---|---|---|---|
| Model 1 | 0.007 | 0.021* | – | 0.024* | 0.011* | -0.004 | 0.001 | 0.004 |
| | (0.005) | (0.004) | | (0.003) | (0.006) | (0.003) | (0.002) | (0.004) |
| Model 2 | -0.0002 | 0.017* | – | 0.025* | 0.032* | -0.002 | -0.000 | -0.002 |
| | (0.001) | (0.006) | | (0.004) | (0.011) | (0.004) | (0.002) | (0.007) |
| Model 3 | -0.0024* | 0.016* | 0.004* | 0.027* | 0.067* | -0.004 | 0.001 | -0.009 |
| | (0.001) | (0.008) | (0.002) | (0.005) | (0.022) | (0.005) | (0.003) | (0.013) |
| | Democ | CIM | OTrans | ATrans | GNP/cap | $\text{Stock}_{t-1}$ | Border | # Borders |
| Model 1 | -0.0004 | – | 0.004 | 0.005* | 0.001* | 0.310* | 0.084* | -0.0004* |
| | (0.0003) | | (0.004) | (0.002) | (0.0002) | (0.008) | (0.009) | (0.0001) |
| Model 2 | 0.0001 | -0.002 | 0.013* | 0.006* | 0.001* | 0.320* | 0.087* | 0.0000 |
| | (0.0003) | (0.002) | (0.005) | (0.003) | (0.003) | (0.013) | (0.013) | (0.003) |
| Model 3 | 0.0007 | -0.003 | 0.011 | 0.012* | 0.0002 | 0.311* | 0.122* | -0.0003 |
| | (0.0006) | (0.003) | (0.008) | (0.005) | (0.0006) | (0.016) | (0.020) | (0.0004) |
| | Cost | Time | $\alpha$ | N | Groups | $\mu_T$ | | |
| Model 1 | -0.001* | 0.374* | -2.480* | 575,330 | 24,441 | 23.5 | | |
| | (0.0004) | (0.078) | (0.592) | | | | | |
| Model 2 | -0.0006 | 0.448* | -3.397* | 317,176 | 17,880 | 17.7 | | |
| | (0.0005) | (0.103) | (0.781) | | | | | |
| Model 3 | -0.0004 | -0.285 | 2.167 | 157,149 | 15,878 | 9.9 | | |
| | (0.0009) | (0.276) | (2.094) | | | | | |

Note: Standard errors are reported in parentheses.

Table 4: Summary of Alternative Approach Estimates

| | Table 1 | `xtpcse, i` | `reg, r/c` | `xtreg, f` | `xtreg, r` | `xtregar, f` | `xtregar, r` |
|---|---|---|---|---|---|---|---|
| ViolDiss | 0 | 0 | 0 | + | + | + | + |
| OGenoc | + | + | + | + | + | + | + |
| OCiv War | + | + | + | + | + | + | + |
| OWar | + | + | + | 0 | 0 | 0 | 0 |
| AGenoc | 0 | - | - | - | - | - | 0 |
| ACivWar | 0 | 0 | 0 | 0 | 0 | + | 0 |
| AWar | 0 | 0 | 0 | + | 0 | 0 | 0 |
| Democ | 0 | 0 | 0 | - | - | - | 0 |
| OTrans | 0 | + | 0 | 0 | + | 0 | + |
| ATrans | + | + | + | + | + | + | + |
| GNP/cap | + | + | + | 0 | + | + | + |
| $\text{Stock}_{t-1}$ | + | + | + | + | + | + | + |
| Border | + | + | + | na | + | na | + |
| # Borders | - | - | - | na | - | na | - |
| Cost | - | - | - | 0 | - | - | - |
| Time | + | + | + | + | + | 0 | + |
| F all $u_i$=0 | $-$ | $-$ | $-$ | 2.05$*$ | $-$ | 1.82$*$ | $-$ |
| $\rho$ | $-$ | $-$ | $-$ | $-$ | $-$ | 0.24 | 0.24 |

Note: Entries summarize the sign of the parameter, when statistically significant at the .10 level. `xtpcse, i` uses panel corrected standard errors under the assumption of hetroskedastic errors across panels, and common error variances within panels; `reg, r/c` is OLS with robust standard errors that are corrected under the assumption of variable error variances within panels; `xtreg, f` is a fixed effects model; `xtreg, r` is a random effects (aka Bayesian hierarchical) model; `xtregar, f` is a fixed effects model with an AR(1) term; `xtregar, r` is a random effects model with an AR(1) term. na indicates that the variable was excluded because it is colinear with the fixed effects dummy variables.

# References


Apodaca, Clair. 1998. Human rights abuses: Precursor to refugee flight? *Journal of Refugee Studies* 11:80–93.

Beck, Nathaniel. 2001. Time-series–cross-section data: What have we learned in the past few years? *Annual Review of Political Science* 4:271–293.

Beck, Nathaniel and Jonathan N. Katz. 1995. What to do (and not to do) with time-series cross-section data. *American Political Science Review* 89:634–647.

Beck, Nathaniel and Jonathan N. Katz. 2001. Throwing out the baby with the bathwater: A comment on green, kim, and yoon. *International Organization* 55:487–495.

Borjas, George J. 1987. Self-selection and the earnings of immigrants. *American Economic Review* 77:531–553.

Clague, Christopher, Philip Keefer, Stephen Knack, and Mancur Olson. 1999. Contract-intensive money: Contract enforcement, property rights, and economic performance. *Journal of Economic Growth* 4:185–211.

Crisp, Jeff. 2000. Who has counted the refugees? unhcr and the politics of numbers. In *Humanitarian Action: Social Science Connections*, edited by S.C. Lubkemann, L. Minear, and T.G. Weiss, Occasional Paper # 37, Providence: Thomas J. Watson, Jr. Institute for International Studies, pp. 33–62.

Davenport, Christian A., Will H. Moore, and Steven C. Poe. 2003. Sometimes you just have to leave: Threat and refugee movements, 1964-1989. *International Interactions* 29:27–55.

Donnelly, Jack. 2003. *Universal Human Rights in Theory & Practice*. Ithaca: Cornell University Press, second edition.

Faist, Thomas. 2000. *The Volume and Dynamics of International Migration and Transnational Spaces*. Oxford: Clarendon Press.

Fearon, James D. and David A. Laitin. 2003. Ethnicity, insurgency and civil war. *American Political Science Review* 97:75–90.

Gibney, Mark, Clair Apodaca, and James McCann. 1996. Refugee flows, the internally displaced and political violence (1980-1993): An exploratory analysis. In *Whither Refugee? The Refugee Crisis: Problems & Solutions*, edited by D.L. Cingranelli, Leiden: PIOOM, pp. 45–66.

Gibney, Mark and Matthew Dalton. 1996. The political terror scale. *Policy Studies and Developing Nations* 4:73–84.

Green, Donald P., Soo Yeon Kim, and David H. Yoon. 2001. Dirty pool. *International Organization* 55:441–468.

Gwartney, James A. and Robert Lawson. 2003. *Economic Freedom of the World: 2003 Annual Report*. Washington, DC: Cato Institute.

Halvorsen, Robert and Raymond Palmquist. 1980. The interpretation of dummy variables in semilogarithmic equations. *American Economic Review* 70:474–475.

Harff, Barbara. 2003. No lessons learned from the holocaust? assessing risks of genocide and political mass murder since 1955. *American Political Science Review* 91:57–74.

Harff, Barbara and Ted Robert Gurr. 1988. Toward empirical theory of genocides and politicides: Identification and measurement of cases since 1945. *International Studies Quarterly* 32:359–371.

Harff, Barbara and Ted Robert Gurr. 1996. Victims of the state: Genocides, politicides, and group repression from 1945 to 1995. In *Contemporary Genocides: Causes, Cases, Consequences*, edited by A.J. Jongman, Leiden: PIOOM, pp. 33–58.

Hatton, Timothy J. and Jeffrey G. Williamson. 1998. *The Age of Mass Migration: Causes and Impact*. New York: Oxford University Press.

Karemera, David, Victor Iwuagwu Oguledo, and Bobby Davis. 2000. A gravity model analysis of international migration to north america. *Applied Economics* 32:1745–1755.

Kennedy, Peter E. 1981. Estimation with correctly interpreted dummy variables in semilogarithmic models. *American Economic Review* 71:81.

King, Gary. 2001. Proper nouns and methodological propriety: Pooling dyads in international relations data. *International Organization* 55:497–507.

Lauren, Paul Gordon. 1998. *The Evolution of Human Rights: Visions Seen*. Philadelphia: University of Pennsylvania Press.

Massey, Douglas S. 1990. The social and economic origins of immigration. *Annals of the American Academy of Political and Social Science* 510:60–70.

Moore, Will H. and Stephen M. Shellman. 2002. Refugee or internally displaced person? to where should one flee?, unpublished ms., available online at: http://garnet.acns.fsu.edu/~whmoore/researchpapers.html.

Moore, Will H. and Stephen M. Shellman. 2003. Fear of persecution: A global study of forced migration, 1952-1995, unpublished ms., available online at: http://garnet.acns.fsu.edu/~whmoore/researchpapers.html.

Morrison, Andrew R. 1993. Violence or economics? what drives internal migration in guatemala? *Economic Development and Social Change* 4:817–831.

Neumayer, Eric. 2003. Bogus refugees? the determinants of asylum migration to western europe, unpublished ms., available online at: http://netec.mcc.ac.uk/WoPEc/data/Papers/wpawuwpla0311002.html.

Neumayer, Eric. 2004. Asylum destination choice: What makes some western european countries more attractive than others? *European Union Politics* Forthcoming.

Polachek, Solomon W. and Stanley W. Siebert. 1993. *The Economics of Earnings*. New York: Cambridge University Press.

Sarkees, Meredith Reid. 2000. The correlates of war data on war: An update to 1997. *Conflict Management and Peace Science* 18:123–144.

Schmeidl, Susanne. 1997. Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990. *Social Science Quarterly* 78:284–308.

Schmeidl, Susanne. 1998. Comparative trends in forced displacement: Ipds and refugees, 1964-96. In *Internally Displaced People: A Global Survey*, edited by J. Hampton, London: Earthscan Publications, pp. 24–33.

Schmeidl, Susanne. 2000. The quest for accuracy in the estimation of forced migration. In *Humanitarian Action: Social Science Connections*, edited by S.C. Lubkemann, L. Minear, and T.G. Weiss, Occasional Paper # 37, Providence: Thomas J. Watson, Jr. Institute for International Studies, pp. 127–157.

Schneider, Gerald and Sabine Carey. nd. Hirschman revisited: Asylum seeking in germany and switzerland between 'voice' and 'exit', unpublished manuscript.

Schneider, Gerald, Thomas Holzer, and Sabine Carey. 2001. Hirschmann auf dem prufstand: 'voice' und 'exit' als widerstreitende fluchtursachen. In *Asylpolitik auf Abwegen: Nationalstaatliche und europaische Reaktionen auf die Globalisierung der Fluchtlingsstrome*, edited by Gerald Schneider and Thomas Holzer, Opladen: Leske und Budrich, pp. 114–139.

Shellman, Stephen M. 2001. Global directed dyad geographic borders database, 1964-1999, available online at: http://faculty.wm.edu/smshel/Data/Data.html.

Stanley, William Deane. 1987. Economic migrants or refugees from violence? a time-series analysis of salvadoran migration to the united states. *Latin American Research Review* 22:132–155.

UNHCR. 2000. *The State of the World's Refugee's: Fifty Years of Humanitarian Action*. New York: Oxford University Press.

Weiner, Myron. 1996. Bad neighbors, bad neighborhoods: An inquiry into the causes of refugee flows. *International Security* 21:5–42.

Wooldridge, Jeffrey. 1999. *Introductory Econometrics*. New York: Chapman.

Zolberg, Aristide R., Astri Suhrke, and Sergio Aguayo. 1989. *Escape From Violence: Conflict and the Refugee Crisis in the Developing World*. New York: Oxford University Press.