Online Appendix to
Moore & Shellman's "Whither Will They Go? A Global Study of Refugees' Destination, 1965-1995" *International Studies Quarterly*, 2007, December, Volume 51, Number 4, pp. 811-834.

22 September 2007

This document discusses three issues: the operational indicators used (including the validity and reliability of the refugee data), the statistical models employed, and a more detailed discussion of our findings than in the body of the article.

Operational Indicators

Monadic Indicators

Our monadic variables are measured for both the country of origin and the country of asylum (our 'pull' variables). In the tables we label each variable Asylum or Origin. We argue that characteristics in the origin countries will affect whether or not a refugee flow is produced in a given country, while the characteristics of potential asylum countries will affect where refugees end up. Below we sketch the monadic indicators.

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 from 1965-1995.[1] More specifically, we use the sum of guerrilla warfare attacks[2] and riots[3] to create our event count indicator of violent dissent in both the origin and asylum countries.

We use two indicators to measure the coercive behavior of the state in both countries: a dichotomous indicator of mass killing and a multichotomous indicator of state terror. The dummy variable indicator of genocide and politicide events comes from Barbara Harff and her work on the State Failure Project.[4] 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.[5]

We use the Correlates of War project's measure of civil war (Sarkees 2000) to operationalize the interaction of the state and dissidents in both origin and asylum

---

[1] The complete set of data is available for purchase at: http://www.databanks.sitehosting.net/; the variables we have used will be made available as part of the replication data set.

[2] 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.

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

[4] 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/.

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

countries.[6] It is a dummy variable that is coded 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 Wars* (Kohn 1999) 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 in both asylum and origin countries we turn to the Polity IV data on institutions and utilize 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.[7] 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 "transition value" (e.g. -88), 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 could be coded as a useful proxy of an expectation of a lack of order. The polity data are available for our full sample.

We also need a measure of average wages for both countries in the directed dyad. The standard measure of average wages in large pooled time series studies of migration is per capita gross national product (e.g., Borjas 1987, Hatton & Williamson 1998, Karemera, Oguledo & Davis, 2000).[8] To maximize valid observations in our sample we merged GNP data from two sources: the World Bank World Development Indicators and Banks' CNTS.[9] 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.[10] These data are available for our full temporal sample.

Finally, we measure whether or not a potential asylum country signed the 1951 UN Refugee Convention or the 1967 Protocol. Our dummy variable indicator is constructed using the data published in Annex 1 of *The State of the World's Refugees* UNHCR (2000).

---

[6] We use both the extra-systemic and civil war data sets as the extra-systemic data set codes wars between a colonial metropole and its colony.

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

[8] 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.

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

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

Dyadic Indicators

Our study requires measures for six dyadic-level concepts: the presence of a contiguous border; when the two countries are at war with one another; when they share a border and both have a civil war; the number of borders the country of origin has; whether there is a common language spoken in each country; and whether the countries have a military alliance.

The presence of a border is one of three indicators we use to capture the cost of relocation (we describe the others below). If one is walking, riding a bike, or traveling 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. Data are available from 1964-1999.

We employ the number of borders for each country of origin as a second indicator of the costs of relocation. 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.

Our indicator for the presence of a common language in both countries comes from the Centre D'etudes Prospectives et D'informations Internationales' (CEPII) Distance Measures database.[11] We use the common language variable from their bilateral file which is a dichotomous indicator "set to one if a language is spoken by at least 9% of the population in both countries" (Mayer & Zignago 2006:4).[12] The data are available for every year in our study.

We turn to Leeds' (2005) Alliance Treaties, Obligations and Provisions (ATOP; http://atop.rice.edu/) project for our measure of the presence of a military alliance between the country of origin and the country of (potential) asylum. We use the dichotomous *atopally* variable, but recoded the cases where there was a nonaggression obligation to a value of zero, thus eliminating the nonaggression pacts. These data are available for the entire temporal domain of our study.

We argue that border wars will positively influence refugee flows (probability and size) and thus need a measure of border wars. We used our 'war on territory' variable (described above) and our contiguous border measure to identify the candidate cases and then consulted the COW project's list of wars to identify the border wars (Sarkees 2000: 134-38). We also rely on the COW project's measure of civil war and our border data to measure the presence of a civil war in each of two bordering countries.

Directed-Dyadic Indicators

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

---

[11] These data are available online at: http://www.cepii.fr/anglaisgraph/bdd/distances.htm. In addition to these data we were interested in using the Gleditsch and Ward (2001) data in this study, but one must be able to program with Perl to do so, and though we downloaded the file, we were unable to access the data.

[12] The CEPII distance project used the web site http://www.ethnologue.com/, the CIA World Factbook web site (http://www.cia.gov/cia/publications/factbook/), and Jacques Leclerc's "Lamégement Lingustique dans le Monde" website (http://www.tlfq.ulaval.ca/axl/) as sources for their data. Mayer & Zignago (2005) use the data in their study of international trade.

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.

We also require an additional measure of the cost of relocation: because there is variance in the cost of traveling 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). We weight distance by the average wage in the country of origin and create 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.[13] These are known as 'time-equivalent' measures in economics (see Polachek & Siebert 1993, 72). We used the EUGene (v. 3.03) software to produce the Correlates of War project's measure of distance between capitals[14] and we used the data described above for average wages (i.e. GNP per capita).

The final concept that we need to measure is the presence of a colonial tie between the country of origin and the country of potential asylum. We used the International Correlates of War (ICOW) Colonial History data set (Hensel & Mitchell 2006) to measure the presence of a colonial tie between the country of origin and the country of (potential) asylum.[15] The ICOW data has a `Type1-2` variable with four values to describe the colonial relationship: (1) possessor controlled all or most of entity, (2) possessor controlled part of entity, (3) occupation, and (4) League of Nations Mandate/UN Trusteeship. Our hypothesis is unlikely to be operative in cases of occupation, so we code a dummy variable set to one when `Type1-2` equals one, two or four, and zero otherwise.

Validity and Reliability of Refugee Measures

Crisp (2000) and Schmeidl (2000) discuss the quality of the data available for refugee populations. Crisp focuses primarily on the political motivations that undermine accurate reporting of data: countries of origin have incentives to under report the number of refugees who have left their country and host countries have incentives to over report refugee populations. Schmeidl, on the other hand, offers a comprehensive overview and comparison on the data collected on refugees (and internally displaced persons) by different sources. While several organizations collect such data, the major organization that collects these data, the United Nations High Commissioner for Refugees, did not hire a trained statistician whose primary responsibility was data collection until the early 1990s when they put Bela Hovy in such a post. Hovy conducted a multi-year overhaul of the UNHCR's data on refugees. We use Hovy's refugee data in our study.

---

[13] The indicator is the distance between countries divided by the per capita GNP in the country of origin. Weighting distance in this way takes into account the fact that people earning higher wages can travel a given distance more readily than people earning lower wages (e.g., the average South African is more likely to be able to travel to Europe than the average Mozambican).

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

[15] These data are available online at: http://garnet.acns.fsu.edu/~phensel/icowdata.html#colonies.

Despite the fact that the data that are available from UNHCR have been vetted by trained social scientists and are the best available, they remain, without question, noisy: they are estimates to begin with, and as Crisp has documented, these estimates are subject to political manipulation and for many years were collected using less than professional standards. That said, it is important to consider these issues in the context of similar cross-national efforts. Estimates of national accounts (e.g., gross national product) and population are therefore of interest. These data are also subject to political manipulation and, early on, less than professional data collection standards. A major effort to reduce the noise in national accounts data conducted by the Penn World Tables (PWT) project is of interest as it is similar to the efforts of Hovy. The PWT documents problems with cross-national national accounts statistics, and even goes as far as producing a quality rating for each estimate. Interestingly, few if any results have been overturned as a consequence of the improvements that the PWT project has brought to national accounts estimates. That is, despite the fact that national accounts data were even more noisy prior to the PWT project, studies that used the improved data did not overturn extant findings. In other words, the signal present in the national accounts data came through despite the noise.

We do not wish to make a case for the cavalier use of any cross-national data set simply because it is available. The validity and reliability of data are important and the results we report in our study hinge on a strong signal being present in the data. Of course, the same issues confront efforts to collect data on human rights violations, dissident actions, etc. Identifying refugee populations is inherently no more difficult than identifying many of the other independent variables we use in our study. Yet, to return to the main point, if future data collection leads to revisions of refugee data then it will be important to re-evaluate the findings reported in our study using that new data. We now address some alternative model choices and how the results from those models fit or do not fit with the results reported in the paper.

Heckman Selection and Zero-Inflated Negative Binomial Models

We estimate a hybrid of Heckman's (1979) two-step sample-selection model and a zero-inflated negative binomial (ZINB) regression model. To execute the first step of the Heckman we estimate a probit regression to determine the probability that an origin country-year produced a refugee flow. "Modeling such selection empirically manages sources of bias and allows one to draw truer inferences" (Reed 2000, 84). We include only origin (i.e., monadic 'push') variables in the selection (probit) equation. These variables affect the probability of any given origin country producing refugees (e.g., Moore & Shellman 2006).

The Heckman two-step model includes an OLS regression in its second step, but we replace that with a ZINB regression. We do so because, as mentioned briefly in the main manuscript, our sample has two sub-populations: directed-dyad-years in which the country of origin produced a refugee flow and those directed-dyad-years in which the origin country did not. Further, our dependent variable is an event count. The ZINB is a useful statistical model for studying situations with a sample with two groups with distinct risks of experiencing a non-zero count and an integer count for the sub-group with greater risk (we offer more detail below). Our approach is a hybrid because we include the inverse Mills ratio from the first step of the Heckman model selection

equation as a regressor in both of the ZINB model's equations. In the second step of the Heckman selection model the inverse Mills ratio serves as a "correction term" for the bias that arises from selectivity. Yet, rather than estimate an OLS regression which includes the inverse Mills ratio as our second step of the Heckman, we estimate a ZINB model and include the inverse Mills ratio as a regressor in both of the ZINB equations.[16]

As implied above, the ZINB is a two equation model in which the first equation is a probit model with a dependent variable coded 1 when refugees left the country of origin and sought refuge in the country of potential asylum, and assigned a zero when the country of origin produced refugees, but none of them sought refuge in the country of potential asylum. The probit equation samples is comprised of all the directed-dyad-years in which the country of origin produced a refugee flow to at least one country (i.e., the cases that we used the first step of the Heckman model to identify as directed-dyad-years where the origin country produced a refugee flow).

The second equation of the ZINB model uses as its dependent variable a count of the number of refugees that sought asylum in the country of origin. We use the negative binomial model because our dependent variable in this stage is an event count. Typically, in such cases, scholars choose a model from the Poisson/negative binomial family. We argue that peoples' decisions are linked via a common information set so the probability of any given person in a given year in a given country choosing to abandon her home is not independent of any other person in that same country-year. Thus, the Poisson model is inappropriate (King 1989b; Long 1997, chap 8). In contrast, the negative binomial model allows one to estimate a parameter that represents the extent to which the events influence one another within each observation.[17] Our argument implies that the parameter will be positively signed and statistically significant. As such, we assume that the refugee data were produced by a negative binomial-like process.

Although, our sample includes only origin countries that produced refugee flows in a given directed-dyad-year, less than 1.5% percent of those directed-dyad-year observations contain positive refugee counts. Again, this is consistent with the conjecture that refugee exodus is an uncommon event. In addition, within the sample, there are two distinct populations: those directed-dyads which have zero probability of producing cross-border refugee flows in a given year and those with a nonzero probability of producing cross-border refugee flows in a given year. For example, while 626,200 Rwandans sought refuge in Tanzania in 1994, none sought refuge in India, Turkey, or South Africa. As such, in this stage we need a statistical model capable of distinguishing between two populations, given covariates.

A class of zero-inflated models is capable of doing split population analyses, and we use the zero-inflated negative binomial (ZINB) model to estimate our parameters.[18]

---

[16] See Heckman (1979). We estimated the Heckman and ZINB models separately (thus making it a two-step Heckman) and used Stata's (ver 9) heckman and zinb commands to do so. The command (.do) file will be made available as part of the replication dataset.

[17] See King (1989a:764-9) for a detailed explanation of why the negative binomial model is useful for this sort of argument.

[18] The hurdle Poisson model (King 1989b) is another option, but we selected the zero-inflated negative binomial model because of the assumption it makes about the possible count values in the population with a non-zero probability of producing forced migrants. More specifically, the zero-inflated negative binomial model assumes that some of the cases at risk to producing a positive count may produce a count of 0 whereas the hurdle Poisson model assumes a truncated count such that all cases at risk to

The model allows us to estimate the effects of the covariates on (1) the probability a directed-dyad-year produces zero refugees ("inflate" probit equation) and (2) the magnitude of the flows which are produced (negative binomial). We include only the asylum and relocation costs variables in the ZINB model. We know at this stage that the country of origin produced refugees. In the second stage, we want to identify and assess the factors that influence where refugees go. The asylum and relocation cost variables enter both the inflate and negative binomial equations because we suspect that the factors which influence positive flows also influence the probability of producing zero flows. Furthermore, we suspect that the signs on the variables across equations will be opposite one another.[19] Since we are analyzing a pooled-cross-section data matrix we are concerned with heteroskedasticity and autocorrelation and their threats to efficiency. To address the issues, we calculate robust standard errors for all three equations.

Full Discussion of Results

In the article that this Appendix supplements we focused on a key set of findings. Here we present a more complete discussion. We begin by discussing the origin factors associated with keeping in or forcing out refugees and move onto discussing the monadic, dyadic, and directed dyadic factors that are associated with attracting and dissuading refugees.

Origin Push Effects

Because of substantive interest, space considerations, and the fact that other studies report on the impact of 'push' factors (e.g., Schmeidl 1997, Moore & Shellman 2004), we limit our discussion of the results to the substantive effects of the estimates for the ZINB model (i.e., the probability of observing a nonzero flow and the size of the flow conditional on the presence of one) and only briefly describe the results for the first step of the Heckman model. All of the variables in the selection equation are statistically significant. As expected, origin dissident violence, genocide, civil war, and foreign troops increase the probability that an origin country produces a refugee flow. Moreover, as origin democracy and origin GNP per capita increase, the probability that an origin country produces refugees decreases. Finally, we find that transitioning regimes increase the probability that the origin country produces a refugee flow.

Selection Effects

The primary purpose of estimating the selection model was to produce estimates of the inverse Mills ratios (IMR). Table 3 demonstrates that this was a useful decision: the IMR variable produces statistically significant coefficient estimates for both the probit and NB equations of the ZINB model in Table 3. Both estimates indicate that we should be using

---

producing a count will produce a non-zero count (Zorn 1998). We expect that the covariates that are established by our hypotheses will both distinguish those dyads at risk to producing flows from those that are not and affect the number of refugees produced. However, we do not anticipate that all dyads at risk to producing refugees in a given year will produce a non-zero forced migrant flow (e.g. Rwanda to Brazil). As such, the zero-inflated negative binomial model is a better choice than the hurdle Poisson for our study.

[19] As we explain below, preliminary estimates led us to recognize spatial heteroskedasticity in our data that changes this expectation for some indicators. In general, however, when they do not have the same sign, one coefficient should be statistically indistinguishable from zero.

a sample selection model for these data.[20] When the IMR is significant, techniques employed without the IMR parameter yield biased results. We now turn toward a discussion of the ZINB estimates beginning with the monadic effects.

Monadic Effects & Interactions

We begin with the measures of violence in the country of (potential) asylum. Violent dissent in the asylum country does not affect the probability of a refugee flow nor the conditional expected number of refugees. However, genocide/politicide in the country of (potential) asylum both reduces the probability of observing a flow and, if there is a flow, reduces the expected number of refugees who will seek asylum there. Specifically, the IRR shows that asylum genocide/politicide reduces the expected count of refugees by 54%. The substantive effects of genocide are also depicted in Figure 1A. The presence of a genocide taking place in the asylum country reduces the expected count of directed dyadic refugee flows by almost 2,100.

The presence of foreign troops at war in the country of asylum also reduces the conditional expected number of refugees as well as the probability of observing a flow. More specifically, foreign troops reduce the expected number of refugees (assuming a flow) by 76%. Figure 1A shows that for noncontiguous countries, Asylum War on Territory reduces the expected count of refugees by almost 3,000 individuals.

The presence of a civil war in the country of potential asylum increases the probability of not observing a refugee flow but does not significantly influence the conditional expected number of refugees. We explore this finding in combination with the presence of civil wars in both the origin and asylum country below.

We now turn attention towards examining the effects of the potential asylum country's regime type, economy, and UNHCR treaty signatory status. To our surprise regime type fails to influence either the probability of observing a flow or the conditional expected number of refugees. Neither the component term (asylum democracy) nor the interaction term produces statistically significant coefficient estimates.

Non contiguous transitioning regimes unexpectedly attract refugees. A regime that has no polity in place (i.e., is undergoing turmoil) will have expected flow 344% larger than one that is not undergoing a transition. Moreover, asylum countries experiencing transitions that border the origin country do not alter the relationship.

Average wages (measured as GNP/capita) also involves an interaction with border. The Asylum GNP/capita coefficient is statistically significant in the probit and NB model. Figure 1B plots the relationship between Asylum GNP and the expected number of directed dyadic refugees for non-contiguous countries. As Asylum GNP increases by ½ standard deviation, the expected change in refugees decreases by about 640. The graph in Figure 3B shows how border interacts with asylum GNP/capita. When the directed-dyad is not contiguous, rising Asylum GNP/capita slightly decreases the expected count of refugees. However, when the directed-dyad is contiguous, rising Asylum GNP/capita increases the expected count of refugees. For example, when the asylum country's GNP/capita is equal to $15,000 the difference between the expected number of directed dyadic refugees across contiguous and noncontiguous countries is

[20] Leung and Yu (1996) show that the t-test can be biased when the independent variables are highly correlated with the IMR. The highest correlation is .20 between relocation costs and the IMR. The other inter-IMR correlations are all less than .05.

more than 10,000 refugees and over 20,000 refugees for countries with GNP/capita values of 40,000.

The final monadic level characteristic is our measure of whether the country of (potential) asylum is a signatory to one or both of the UNHCR nonrefoulement treaties. This indicator is also interacted with border. The component term—Asylum Treaty Signatory—represents the impact of this variable for non-contiguous directed-dyad years. The coefficient estimate is not statistically significant in the probit equation but is statistically significant in the NB model. Non-contiguous treaty signatories do not have a greater probability of observing a refugee in-flow than non-contiguous treaty signatories of potential asylum. However, noncontiguous treaty signatories do decrease the expected directed dyadic flow of refugees if there is a directed dyadic flow. This effect is demonstrated in Figure 1A. When border is equal to zero, the presence of a treaty signatory regime decreases the number of refugees flowing across a given dyad by more than 2,000. When border is equal to one, the probability of observing a flow across the dyad is reduced, but if a flow does occur across that contiguous dyad, being a treaty signatory has minimal if any impact (the coefficient on the interaction term is insignificant and the IRR is only 0.87).

Can we account for this finding? We submit that the answer is in the introduction to this study. There we observed that that vast majority (generally greater than 90%) of refugees walk across a border, but that both the United States and Germany are among the top-10 asylum destinations over the period of study. Even once we correct for spatial autocorrelation, then, we should know that fewer refugees travel great distances than small ones. Thus, the expected number of refugees traveling to countries that sign treaties—which are generally further away than those that do not sign treaties—will be lower *given that we have conditioned that expectation on the fact that some refugees will make the trip*. This last part is crucial: the NB model is conditional. The result states that signatories will receive fewer refugees than non-signatories, and we know that signatories are—on average—further away from the typical non-signatory. Therefore, we need to examine the interaction of treaty signatory and relocation costs. Figure 3A shows that as relocation costs grow larger, the expected count of refugees increases for treaty signatory countries but decreases for non-treaty signatory countries. Thus, as we expected, treaty signatory countries far away are more attractive that far away non-treaty signatory countries.

Frances Stewart suggests an alternative possibility.[21] Our dataset demonstrates that the probability of being a signatory to one of the UN non-refoulement treaties is positively correlated with being a democracy. Stewart speculates that xenophobia might more readily find expression in democracies—where politicians must respond to public sentiment—than in autocracies. If we assume that xenophobia is equally distributed across democracies and autocracies—and this strikes Stewart and the authors as plausible—then it follows that while democracies must be (somewhat) responsive to xenophobic outcries, autocracies need not do so.

Dyadic & Interactive Effects

We have six dyadic level variables, and five of them produce a statistically significant coefficient in one or more of the two models. First, whether or not the (potential) asylum

[21] Comment Stewart made at the Economics of Forced Migration Workshop, MIT, December 2005.

country borders the origin refugee producing country influences both the probability of observing a flow and the number who seek asylum (assuming any do). We must remember that border is interacted with several other variables in the model and a proper interpretation must take into account all of the other variables' values. So we set all dummy variables equal to zero except border. We then set the dummy variables interacted with border equal to zero. We set the interval variables at their means and the interval variables interacted with border equal to the mean of the interval variable (1 * mean = mean). The result is an expected increase of more than 4000 directed dyadic refugees when the origin and asylum country border one another.

The number of contiguous borders that a country of origin has both reduces the probability of observing a flow across any given directed-dyadic border and has a negative impact on the expected number of refugees who cross that directed-dyadic border (given that a flow occurs). Each additional border decreases the expected number of refugees by 11%. Figure 1B shows that as we move across origin countries with five borders to eight borders, the number of directed dyadic refugees decreases by about 1,000 individuals. There is another 1,000 refugee decrease associated with a shift from a country with eight borders to a country with eleven borders. These results are consistent with our expectations.

The coefficient for the Common Language variable in the NB model indicates the estimated impact of a shared common language for non-contiguous origin and asylum states. The probit model produced a negatively signed coefficient which is consistent with our expectation that a common language would increase the probability of observing a directed dyadic refugee flow given that refugees left the origin country. The IRR value of 0.348 in the NB model indicates that a common language spoken by at least 9% of the population in each non-contiguous country decreases the expected refugee flow, given that one occurs, by 65.2%. Like our finding for treaty signatory this result is contrary to our expectations.

To explore this possibility, we need to consider our interaction terms containing common language and border and common language and relocation costs. The best way to examine the effects of these interactions is to plot the expected counts given specific values of the different independent variables and interaction terms. Figure 3C shows that the difference in expected refugee counts across dyads that share and do not share common language, do not change considerably for countries that border one another. However, for noncontiguous countries, fewer refugees are expected to end up in asylum countries that share a common language with the origin country. This suggests that contiguity matters more than common language. We also examined the interaction of common language and relocation costs, but the relationship was not meaningful given that the interaction term was insignificant.

Table 3 indicates that while military alliances increase the probability of observing a refugee flow between two countries, they do not influence the expected size of the flow when one does occur. This is consistent with the idea that alliances are a proxy for positive relations among countries (and, presumably, their populations) and inconsistent with the expectation that countries will not accept refugees from their allies.

The presence of an international border war with battles fought in both countries does not affect the probability of observing a flow, but it dramatically increases the expected size of a flow if one does occur. Figure 2 shows that border war produces an

expected count of almost 60,000 refugees when we set all interval level variables to their means, dichotomous variables equal to zero, the variable border equal to one, the variable asylum international war on territory equal to 1, and the variable Border War equal to one. This is in contrast to when we set Border War equal to zero, border equal to zero, asylum international war on territory equal to zero. These values only produce a positive count of 8,300. Furthermore, setting border equal to one only raises that value to 9,100.

Finally, the presence of bordering civil wars decreases flows but increases flows when both countries experiencing a civil war border on another. Specifically, Bordering Civil War in the inflate equation is negative but not statistically insignificant. However, the coefficient for Bordering Civil War in the NB equation is positive and statistically significant. Figure 2 reveals that the expected count for directed dyadic refugee flows across bordering countries that are both experiencing civil war is almost 20,000. The expected count is significantly reduced (8,300) for non-bordering countries not experiencing civil war. We further discuss the bordering violence findings in the discussion section below.

Directed-Dyadic Effects

The stock of a refugee population in the country of (potential) asylum, the average cost of relocating from one to the other country, and the presence of a colonial tie are our three directed-dyadic level variables. To begin, the coefficient on relocation costs is positive and significant in the inflate equation and negative and significant in the NB equation. Thus, it decreases the probability of a flow, but if it does occur, it decreases the expected count of directed dyadic refugees. Figure 1B shows the effects of relocation costs for non-contiguous, non-treaty signatory countries. As relocation costs increase by a ½ standard deviation and then another ½ standard deviation, directed dyadic refugees decrease dramatically relative to the changes we observe for other independent variables in Figure 1B. We already described the interactive effects of relocation costs illustrated in Figure 3A for treaty signatories.

Second, we argued that people place a high value on their family, friends, and culture, and that Diasporas provide valuable information to potential migrants. We find that indeed, our hypothesis is supported in that the lag of the stock of refugees in potential asylum countries is signed as expected and statistically significant in both the probit and the NB model. Because of the large range of this variable (from 0 to over 1 million) the IRR is not helpful, so we turn to alternative means of assessing the substantive effect. Figure 1B illustrates that as we increase the lag of the stock of refugees from the mean (968) by a half standard deviation (10,000), the expected change in refugees is about 4,100.

Our final variable is a colonial tie between the country of origin and the country of (potential) asylum. The estimates in Table 3 indicate that the presence of a colonial tie does not change the probability of observing a refugee flow, but it does increase the expected size of a flow if one occurs. More specifically, the IRR indicates that a country of asylum that was the metropole for a country of origin can expect, on average, 415% more refugees than a country of asylum that was not the metropole of the country of origin. The substantive impact is also depicted in Figure 1A, as the presence of a colonial tie produces an average expected increase of 16,000 directed dyadic refugees. This is a sizable effect and is consistent with our expectations.

References

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.

Crisp, Jeff. 2000. "Who has Counted the Refugees? UNHCR and the Politics ofNumbers. In *Humanitarian Action: Social Science Connections*, edited byS.C. Lubkemann, L. Minear, and T.G. Weiss. Occasional Paper # 37,Providence: Thomas J. Watson, Jr. Institute for International Studies, pp. 33-62.

Fearon, James D. & David A. Laitin. 2003. Ethnicity, Insurgency and Civil War. *American Political Science Review*, 97(1):75-90.

Gibney, Mark & Matthew Dalton. 1996. The Political Terror Scale. *Policy Studies and Developing Nations* 4: 73-84.

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

Harff, Barbara. 2003. No Lessons Learned from the Holocaust? Assessing the Risks of Genocide and Political Mass Murder since 1955. *American Political Science Review*, 97(1):57-74.

Heckman, James. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47(1): 153-161.

Hensel, Paul and Sara Mitchell. 2006. ICOW Colonial History Codebook. Typescript. Florida State University. <http://garnet.acns.fsu.edu/~phensel/icowdata.html#colonies> (November 2006).

Hoff , Peter D. and Michael D. Ward. 2004. Modeling Dependencies in International Relations Networks. *Political Analysis*, 12(2):160–175.

King, Gary. 1989a. Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator. *American Journal of Political Science* 33(3): 762-84.

King, Gary. 1989b. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.

Kohn, George C. 1999. *Dictionary of Wars*. New York: Facts on File.

Leeds, Brett Ashley. 2005. *Alliance Treaty Obligations and Provisions (ATOP) Codebook*, version 3.0, available online at: http://atop.rice.edu/download/ATOPcdbk.pdf.

Leung, SF and Yu S. On the choice between sample selection and two-part models. *Journal of Econometrics* 1996; 72: 197-229.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

Mayer, Thierry & Soledad Zignago. 2006. Notes on CEPII's Distances Measures. available online at: http://www.cepii.fr/distance/noticedist_en.pdf.

Moore, Will. H. and Stephen M. Shellman. 2004. Fear of Persecution. *Journal of Conflict Resolution* 48(5): 723-45.

Moore, Will. H. and Stephen M. Shellman. 2006. Refugee or Internally Displaced Person? To Where Should One Flee? *Comparative Political Studies*, 39(5):599-622.

Moore, Will. H. and Stephen M. Shellman. 2008. Whither Will They Go? A Global Study of Refugees' Destinations, 1965-1995. *International Studies Quarterly*, forthcoming (pending editorial approval of revisions).

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

Reed. William. 2000. A Unified Statistical Model of Conflict Onset and Escalation. *American Journal of Political Science* 44(1): 84-93.

Sarkees, Meredith Reid. 2000. The Correlates of War Data on War: An Update to 1997. *Conflict Management and Peace Science*, 18/1: 123-144.

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. Weis, Occasional Paper #37, Thomas J. Watson, Jr. Institute for International Studies, pp. 127-57.

Zorn, Christopher J.W. 1998. An Analytic and Empirical Examination of Zero-Inflated and Hurdle Poisson Specifications. *Sociological Methods and Research* 26: 368-400.