

17 December 2004

Online Appendix for:

“REFUGEE OR INTERNALLY DISPLACED PERSON?
To Where Should One Flee?”
Comparative Political Studies, 39, August 2006.

Will H. Moore
Department of Political Science
Florida State University
will.moore at fsu.edu

Stephen M. Shellman
Department of International Affairs
The University of Georgia
Athens GA 30602-1492
smshel at uga.edu

This document discusses a variety of technical issues involved in our study: a brief sketch of the Moore & Shellman (2004) model of forced migration, the statistical model, our research design, the operational indicators we use, and a number of sensitivity analyses that probe the robustness of the inferences we draw.

Information Relevant to P

We hypothesize that the coercive/violent behavior of the state, dissidents, and foreign soldiers are the determinants of peoples' expectations about P , the probability that one will become a victim (Moore & Shellman 2004). We submit that people will monitor the behavior of these actors to evaluate whether a sufficient threat exists that they should abandon their homes. More specifically, when the state engages in or sponsors arbitrary arrests, dispossession of property, counter-insurgent military campaigns, death squad

activity, mass killings, etc., people will raise their expectations of P . Similarly, when dissidents undertake violent protests, bombing campaigns, guerrilla attacks, etc., people will also raise their estimates of P . Moreover, the interaction of states and dissidents engaging in civil wars will raise individuals' estimates of P . Finally, when foreign soldiers undertake military campaigns in a country, the people who live there will revise upward their beliefs about P .

In addition to the primary focus on P we also control for other factors that one might anticipate in a rationalist account. First, we contend that people value institutions that provide for freedom and the rule of law as expressed in modern democracies. Second, we submit that people seek to maximize the economic return on their labor. Finally, we suggest that people strongly value their culture—by which we mean language, religion, families and customs—and are more likely to relocate somewhere when they have a trusted source of information about circumstances in a new location. Stated as hypotheses, we expect people to be more likely to abandon their homes when [1] the institutions in their country of origin are not democratic, [2] expected wages in their country of origin are low, and [3] a Diaspora provides both opportunities to find their culture and family members in locations away from their present home, and information about how to get there and what life will be like once they arrive (Faist 2000).

The Statistical Model

We have hypotheses about the impact of the fear of persecution on the probability that people will abandon their homes, and the impact that fear of persecution in the country of

origin as well as in neighboring countries will have on the proportion of refugees relative to IDPs among those who flee. We also have hypotheses about the impact of other variables on the probability that people will abandon their homes, and about the impact that those variables will have on the proportion of refugees to IDPs. As such, we need a statistical model that takes into account the two stages represented in our argument. We employ a Heckman sample selection model to achieve that purpose. This setup allows us to first model what factors lead countries to produce a forced migrant population in a given year (the selection equation), and then model—for each of those country-years with non-zero forced migrant populations—the proportion of those forced migrants who sought refuge across an international border (outcome equation). More explicitly, we write our selection equation as

$$P^* = \gamma'z + u \tag{1}$$

where P^* is the average threshold at which people will choose to abandon their homes given a reasonable fear of persecution, and z is the vector of variables measuring the characteristics in the origin country in a given year that affect P^* . Note that P^* is not observed. That is we cannot observe the average threshold for fleeing, but we can observe whether a country produces forced migrants in a given year such that

$$P = 1 \text{ if } P^* > 0$$

and

$$P = 0 \text{ if } P^* \leq 0.$$

The second equation models the proportion of refugees to all forced migrants and only applies to those countries that produced forced migrants. It can be written as:

$$\frac{R_{ij}}{FM_{ij}} = \beta' \mathbf{x}_{ij} + \varepsilon_{ij} \quad (2)$$

where R_{ij} is the number of refugees produced in a given country year, FM_{ij} is the total number of forced migrants produced in a given country year and \mathbf{x} is the vector of variables measuring characteristics in both the origin and neighboring countries that influence the proportion of forced migrants who become refugees.

We estimate the model using a Heckman (1979) procedure to account for the selection of countries into the sample of those that produce forced migrants.¹ Doing so is appropriate because “[m]odeling such selection empirically manages sources of bias and allows one to draw truer inferences” (Reed 2000, 84). As Signorino (2002: 97-8) explains, “if the selection mechanism is systematically related to or correlated with the dependent variable of the observed sample, failing to account for that selection mechanism will lead to biased inferences ... even for the given sample.” Moreover, the model controls for the interdependent relationship between forced migration and the proportion of forced migrants that seek refuge abroad, and it allows us to empirically differentiate between some of the independent variables’ effects at each stage in the process. For example, we hypothesize that civil war will positively affect individuals’

¹ A researcher could select several alternative methods to analyze these data. To probe the robustness of the results reported here we estimated three alternative models (regression on the proportion without selection, a random effects Tobit model, and a multinomial logit). The results across these three approaches are extremely similar and are similar to those coefficients produced by the sample selection model. We discuss the similarity of these results and other sensitivity analyses below. We feel the selection model best represents our argument and produces unbiased estimates. Moreover, the selection model allows us to model the differential impacts of the variables on the two dependent variables.

decisions to flee their homes but will negatively affect the proportion of forced migrants who become refugees.

One important issue to consider when using Heckman-type models is the identification of the parameters. In many cases theory suggests that the variables that influence one stage of a model will also affect the second stage. Unfortunately, Heckman-type models are not identified if one uses the same regressors in both the selection equation and the equation of interest (Achen 1986:73-99, Sartori 2003:111-16). Our model appears to have identical regressors, and thus to suffer from this identification problem. However, it does not. To identify the model one must have at least one variable in the selection equation that is not included in the equation of interest. We include the lagged stock of forced migrants (i.e., the sum of refugees and IDPs) in the selection equation, but exclude that variable from the equation of interest. In that equation we use the lag of the proportion of the forced migrant stock that are refugees. These variables have the same components as inputs, but one is a sum of those components and the other is a proportion of those components, and they have different distributions (see Table 1 in the article). As such, our Heckman model is identified. Next, we describe our sample and operationalize our concepts.

Spatial-Temporal Domain

To test our hypotheses we developed a sample using the country-year as our unit of observation. Our temporal domain is 1976-95. We use all countries for which we could obtain data on each of the variables that we used in the study.

The Dependent Variables

To measure refugees we used a data set provided to us by Bela Hovy, the Director of the Statistics Division of the United Nations High Commission for Refugees. Susanne Schmeidl provided our IDP data—it is unpublished data from the Global Refugee Migration Project (Schmeidl and Jenkins, 1999). These are the best available sources of data on refugee and IDP stocks.² However, we are interested in flows, not stocks. That is, we want to know how many people relocated and became refugees or IDPs in a given year, not what the number of people with such status was in a given year. As Schmeidl (1998, 2000) explains, the annual refugee stock data are the most valid, reliable data available and there is virtually no large-N data on refugee or IDP flows. Thus, one needs to work with the stock data to produce a flow measure.

When people speak of flow they generally mean the net flow, which is to say the number of people who have migrated minus those who have returned. We are interested in the former, but not the latter. The best, though imperfect, approach to measurement is to take the first difference of the stock (i.e., subtract the present year's value from the previous year's value), and then recode all negative values (which represent a net return

² The UNHCR data cover the period 1954-1999 and are the product of an extensive, multi-year effort by Hovy's team. The IDP data cover 1970-1997 and were put together by Susanne Schmeidl, working with Craig Jenkins. See Schmeidl, (1998, 2000) and Crisp (1999) for detailed discussions about the strengths and weaknesses of these data.

flow) to zero.³ We used this approach to create both a refugee flow and an IDP flow measure. We then added them together to create a forced migrant flow measure. From these measures we construct two additional variables.

First, we need to distinguish country-years that produced forced migrants from those that did not. To measure this, we create a dependent variable equal to one when there is a nonzero forced migrant flow in a country-year and equal to 0 otherwise. We employ the dummy variable as our dependent variable in the selection equation. Second, we create a proportion variable equal to the number of refugees produced in a given country-year divided by the total number of forced migrants produced in that given country-year.⁴ This variable ranges from 0 to 1 and serves as the dependent variable in

³ The weakness of this operationalization is that it will undercount flow in cases where a non-zero stock exists and a resettlement occurs. For instance, if 1,000 refugees are reported in year 1, but 800 are reported in year two, our measure will produce a value of 0 (the -200 will be recoded to 0). It may well be that 200 people returned, but it might also be the case that all 1,000 people were resettled in a third country and 800 additional refugees crossed the border. Thus, our measure likely undercounts flows and thus the size of the impact of our variables is smaller than they should be.

⁴ Because division by zero is undefined and most countries do not produce any forced migrants (i.e., 0/0), we added 1 to both the numerator and the denominator before dividing them. Then we recoded all values equal to one where the origin country produced zero forced migrants (i.e., (0+1)/(0+1)) back to 0.

our outcome equation.⁵ A positive coefficient in this equation indicates a positive impact on the number of refugees relative to IDPs while a negative coefficient signifies a negative impact on the number of IDPs relative to refugees. We now turn attention towards our independent variables before discussing specification issues.

Measures of Violence

We use two different variables to measure state violence: genocide/politicide and violations of the rights of the integrity of the person. The first variable provides us with an indicator of the state targeting civilian non-combatants (either for their ethnic/religious affiliation or their political opposition affiliation). To measure genocide/politicide we use Barbara Harff's data (Harff 2003, Harff and Gurr 1988, 1996). These data identify events where the governing elites or their agents implicitly or explicitly promoted sustained policies of targeting non-combatants. The variable, genocide/politicide, is a binary indicator that is coded zero for country-years in which no such event occurred and one in country-years when such an event took place.⁶

⁵ We also calculated the proportion of IDPs to Forced Migrants and estimated the model. The same inferences are drawn across both dependent variables. The signs on all the variables flip across dependent variables as expected.

⁶ There are four criteria employed to code a genocide/politicide. First, "authorities' complicity in mass murder must be established" (See State Failure codebook – Marshall, Gurr, and Harff, 2001). Second, the campaign of mass killings must last six months or more. Third, *the victims must be unarmed civilians*. This is the most important criterion for our purposes. Fourth, "In principle, numbers provided in "body counts" do not enter

We use the Political Terror Scale (PTS) data to measure human rights violations (Gibney & Dalton 1996).⁷ The PTS is available for the years 1976-1996 and is a standards-based measure of the extent to which a government violates the physical right to integrity of the person. Violations are coded on a 5-point scale where higher values are associated with greater levels of violation. A score of one indicates a country where rights are respected (political imprisonment, torture, and extra-judicial execution are extremely rare). At the other end of the scale, a score of five indicates a country where the entire population is at risk to political imprisonment, torture, and extra-judicial execution as the state regularly employs all of these tactics as a means of rule. A value of four is assigned to countries where political imprisonment, torture, and extra-judicial killings are routine, but are only employed against those active in politics. A score of three indicates a state that routinely uses political imprisonment and extra-judicial execution occurs, but it is not endemic. Finally, a value of two represents states that employ a limited use of political imprisonment, but largely avoid torture, and rarely resort to political killing. The data are collected via content analysis of two sources: Amnesty International annual reports and the US Department of State's annual reports on

the definition of what constitutes an episode. A "few hundred" killed constitutes as much a genocide or politicide as the deaths of thousands if the victim group is small in number to begin with."

⁷ Apodaca (1998) finds that government coercion—as measured via the Political Terror Scale—has a positive impact on refugee stock. Davenport, et al. (2003) also use the political terror scale to measure government coercion, but fail to find support that it impacts net refugee flows.

human rights. Two different variables are created, one based on the Amnesty International coding and one based on the State Department coding. We report the findings using the Amnesty International indicator.⁸

To measure dissident violence we use an event based measure and calculate the sum of riots and guerrilla attacks using the Cross-National Time-Series Data Archive (Banks nd). Riots are defined as “any violent demonstration or clash of more than 100 citizens involving the use of physical force,” and guerrilla attacks are defined 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.”⁹ This is a revised version of a variable used in Davenport’s (1995) study of government coercion. We dropped the two non-violent components—general strikes and anti-government protests—because these events are unlikely to lead people to conclude that their lives, liberty or person are threatened. The variable is a frequency count of these events in a given country-year.

⁸ Unfortunately, the model would not converge when we used the State Department measure and we have been unable to diagnose the problem. We estimated the outcome equation separately and the coefficients and signs for genocide/politicide, PTS, violent dissent and the interaction of PTS and violent dissent are not different. Those interested in the full results can reproduce them using the replication data set and command file.

⁹ These definitions comes from the online codebook available at the “Variables” link at the Cross-National Time-Series Data Archive website:

<http://www.databanks.sitehosting.net/>.

Both government coercion and dissident violence may be episodic rather than sustained. In fact, this is generally the case. However, when both the government and the dissidents engage in more regularized military clashes, people are likely to have a heightened sense of insecurity, even controlling for government coercion and dissident violence. While we expect sustained state and dissident violence (i.e., civil war) to have positive effects on P , P_O , and P_A , we do not expect the impact on P_O to be greater than that on P and P_A . That is, unlike genocide/politicide events, we anticipate that sustained combat will not have a positive effect on refugees relative to IDPs. To measure this variable we use the Correlates of War (COW) Intrastate and Extra Systemic war lists (Sarkees 2000). “An internal war is classified as a major civil war if (a) military action was involved, (b) the national government at the time was actively involved, (c) effective resistance (as measured by the ratio of fatalities of the weaker to the stronger forces) occurred on both sides and (d) at least 1,000 battle deaths resulted during the civil war.”¹⁰ We used the extra-systemic list to identify cases of colonial war: “international wars in which the adversary was a colony, dependency or protectorate composed of ethnically different people and located at some geographical distance or, at least, peripheral to the center of government of the given system member.”¹¹ We used the lists to create a dichotomous measure scored one for country-years on either list and zero for country-years not listed.

¹⁰ This definition is available at the COW project’s website:

<http://www.umich.edu/~cowproj/dataset.html#CivilWar>.

¹¹ This definition is also available at the COW project website:

<http://www.umich.edu/~cowproj/dataset.html#ExtraStateWar>.

A second measure of the interaction of state and dissident violence is available: the product of PTS and dissident violence. This indicator allows us to test our expectations about the conditional impact of dissident violence on the proportion of refugees.

We use the Correlates of War list of interstate wars to help us code the presence of an interstate war on the territory (IWOT) of the country of origin (Sarkees 2000). The COW project requires more than 1,000 battle deaths per year as well as participation by at least two members of the international system to qualify as an interstate war. To determine whether at least one battle had taken place on the territory of each country involved we examined the descriptions of the wars available in historical digests and an encyclopedia of war. When a battle took place in the territory of a participant country, we assign that country-year observation a value of one. All other countries-years (including those not listed in the interstate war data) are assigned a value of zero.

We identified three other concepts that we expect to affect forced migration and the proportion of refugees: institutions that produce democracy, wages, and Diasporas. To measure institutions that produce democracy we use the Polity IV data. Specifically we subtract the autocracy score from the democracy score to create a measure of institutional democracy that ranges from -10 to 10. One of the challenges of the Polity data is that there are periods of transition in a country where the institutions are not sufficiently embedded for the project to produce a polity code. These cases show up in the data as having values of -66, -77, or -88. Rather than treat these as missing values we assigned them a score of 0 and coded a dichotomous measure of transition polities

(TRANS) which we assign a value of 1 when the democracy score has a value of –66, –77, or –88, and a value of zero when democracy is scored in the range from -10 to 10.

We use gross national product per capita as a proxy measure of average wages. The case for using GNP per capita as a proxy for wages is based on [1] the positive relationship between the two,¹² and [2] the widespread use of that indicator in a number of economic studies of migration (e.g., Borjas 1987). We obtained our GNP measure from both the World Bank’s World Development Indicators data and the Cross-National Time-Series Data Archive (Banks nd). We began with the World Bank data and then filled in missing observations with the Banks data, where available.¹³ Finally, we divided the data by population obtained from Fearon and Laitin’s (2003) study as they have greater coverage than we were able to find in other sources.¹⁴

To measure the extent to which culture, family and friends are present, either in potential asylum countries or within one’s own country, we created a lag stock of the proportion by dividing lag refugee stock by the lagged value of the stock of forced

¹² For an analysis that demonstrates a positive, statistically significant cross-national 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>.

¹³ The correlation between the World Bank’s data and Banks’ data is .918. In addition, when we use either of the series in a regression, the estimated coefficients are essentially the same. Given the strength of the correlations and the extent to which they produced similar estimates, we chose to replace missing data in the World Bank data with observations from the Banks’ data.

¹⁴ Data available at <http://www.stanford.edu/~jfearon/data/apsr03repdata.zip>.

migrants. The stock measures all of the people who had migrated—within their country or across a border respectively—in the past and are still in the country, not simply those who had migrated the year before. This measures the size of the forced migrant population from each country in each country, and thus the extent to which those who abandon their homes will find a community of people who share their culture, language, etc., as well as the ability of people in the country of origin to gain information about the prospects of living abroad and how to get there.¹⁵

Neighborhood Scores

Because we expect the information about benefits and costs of seeking asylum to have a different effect on refugee and IDP flows, we need to measure the characteristics of potential asylum countries. Doing so is not straightforward. The challenge involves drawing comparisons across countries which have different structures of information; that is, different numbers of neighbors. Since some countries have only one bordering neighbor, while others have two, three, four, etc., we cannot compare across a standard set of options.

¹⁵ One might ask whether a measure of cultural similarity might better measure our concept. An ideal measure would include both voluntary and forced migrant stock as it would capture both dimensions of our concept: culture and information. Measures of dyadic cultural similarity, on the other hand, would only measure one dimension. Thus, we submit that our measure has greater content validity. Unfortunately, voluntary migrant stock is unavailable for the vast majority of our sample, so we use forced migrant stock alone.

To get a handle on the asylum versus origin decision, we create a summary measure to assess the status of the ‘asylum neighborhood.’ We create a neighborhood score for the following variables: genocide/politicide, government rights violations, violent dissent, civil war, international war on territory, GNP, democracy, and the lag of the proportion of refugee stock. We calculate the mean of each variable where the number of borders acts as the denominator in the summary statistic. We start with a directed dyadic dataset to calculate the neighborhood scores in each potential asylum country. Then we take the mean of the surrounding countries’ scores and merge the measures with the country-year data. For example, as the mean violence levels increase in surrounding countries we expect less forced migrants in the country of origin to become refugees.

The last indicator that we require is one to measure transaction costs associated with rough terrain. We use a data set that codes the existence of mountains on a border (Shellman 2001). Mountain is a dichotomous measure coded one whenever at least 50 percent of a border has an elevation change of 1,000 feet or more from the surrounding area. To develop an asylum neighborhood measure of both mountain and border we created the proportion of neighboring countries that have mountains on their border. Thus, if a country had four borders and two with mountains, the value is 2 divided by 4, which yields .5.

Validity and Reliability of Refugee and IDP Measures

Crisp (1999) and Schmeidl (1998, 2000) discuss the quality of the data available for refugee and IDP 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 (1998) observes that no international organization has a mandate to serve IDPs and that this has led them to be ignored relative to refugees. In fact, cross-national data are not available on IDPs prior to 1971 and Schmeidl (1998) suggests that the quality of data collection improved circa the mid 1980s.

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, and Schmeidl—who was writing her doctoral dissertation at the time—worked as a volunteer on the project. We use Hovy's refugee data in our study. Schmeidl & Jenkins (1999) cleaned up the data on IDPs, and we use that data in our project.

Despite the fact that the data that are available from UNHCR and Schmeidl & Jenkins 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 and Schmeidl have 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 and Schmeidl & Jenkins. 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. And we must note that the IDP data in particular are hard to collect as only one government has access to IDP populations, and they have an incentive to under-report (indeed, to deny the population's existence entirely). Yet, we do not wish to make too much of this problem: the same issue confronts efforts to collect data on human rights violations, dissident actions, etc. Identifying IDP 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 IDP data then it will be important to re-evaluate the findings reported in our study using that new data.

Predictive Validity

Having briefly reviewed the main issues we wish to take advantage of some empirical analyses that permit us to assess both the validity and reliability of the refugee and IDP data we use. First, one can establish the predictive validity of a measure by determining whether it has a strong empirical relationship with concepts that theory suggests they should be correlated. Studies by Schmeidl (1997) and (Moore & Shellman 2004) show that refugees and the sum of refugees and IDPs, respectively, are strongly influenced by measures of civil war, which theory predicts. When we regress our measures of refugee flow and IDP flow on our measure of civil war the coefficients are statistically significant in both regressions with values of 50,869 for refugees and 111,623 for IDPs. These results suggest that despite the noise in the data, the measures have non-trivial validity.

Reliability and Replication at Lower Levels of Precision

Reliability is a different concern. Much of the discussion in Crisp (1999) and Schmeidl (2000) concerns the reliability of the measures. At the integer level the reliability of these data are certainly suspect. Consider this question: how confident is one that the number of refugees or IDPs in a given year from country X is 43,455 rather than 44,621? These data are, after all, estimates based on census efforts where possible, and estimates based on other information where necessary. As such, it is difficult to imagine making the case that these data are highly reliable as integer values. Yet we use the integer values in our study, and that raises the question: are our inferences compromised by limited reliability?

One useful way to get at that question is to reduce the precision of measurement, thus increasing our confidence in the reliability of the measurement, and see whether the

inferences we draw are affected. We do this in two different ways. First, we recode our integer level data to order of magnitude and re-estimate models 1 and 2. Second, we reduce the precision of measurement to a set of binary outcomes and estimate multichotomous logit models. We describe each approach in turn.

The order of magnitude distinguishes ten from hundreds from thousands from tens of thousands, etc. Thus, we recode values between 1 and 9 to 1, those between 10 and 99 to 2, values between 100 and 999 to 3, through those between 1,000,000 and 999,999,999 to 7. We then used the recoded data to create our proportional measure and re-estimated models 1 and 2.¹⁶ There are very few changes in inferences. Neighborhood democracy in the outcome equation for model 1 retains its unexpected negative sign, but its standard error grows relative to its coefficient such that it is no longer statistically significant at a .10 level if we use a two-tailed test. Each of the other variables in both the outcome and selection equations for model 1 replicate the signs and statistical significance reported in our study, as does the ρ parameter. In model 2 all of the signs and statistical significance levels are replicated with the exception of per capita GNP in the outcome equation: it retains its negative sign, but has a considerably larger standard error and is no longer distinguishable from zero. Thus, enhancing the reliability of our measurement by reducing the level of precision to the order of magnitude does not appreciably affect the inferences we draw in the study.

Our second approach is to reduce the distinctions to three: those country-years that produce zero forced migrants, those that produce more refugees than IDPs, and those

¹⁶ We recognize that addition and division operations on ordinal data can produce misleading values, but submit that this exercise nevertheless has value.

that produce less refugees than IDPs.¹⁷ Doing so produces a trichotomous dependent variable that has even greater reliability than the order of magnitude measure. We can no longer estimate the sample selection model, but note that the information used in both equations is retained, and one can use a multichotomous logit model to estimate parameters. As before, we focus on the sign of effect¹⁸ and its statistical significance.

We more or less draw the same inferences using the multinomial logit approach. For example, in both models the coefficient on genocide is much larger for the Refugees>IDPs category than the IDPs>Refugees category. Moreover, genocide is not significant for the IDP>Refugees category. This finding implies that genocide increases a country's likelihood of producing more refugees than IDPs. Civil war, in both models, is positive and significant in both equations but the magnitude is larger for the IDP>Refugee category than the Refugees>IDPs category. Again, consistent with the reported findings of the selection model, civil war increases the likelihood that a country produces more IDPs than refugees. The interaction term, consistent with the reported results, is negative and significant for the refugees>IDPs category. Neighboring genocide continues to decrease the probability that a country produces more refugees than IDPs. The other results are also consistent such as origin GNP/cap decreasing the probability for both categories. Origin international war on territory is positive across both categories

¹⁷ There are no country-years where the number of refugees equal the number of IDPs.

¹⁸ In a multichotomous logit model one cannot directly interpret the coefficients and we rely on the sign of relative risk ratios (i.e., e^b , where b is the coefficient) to determine the sign of the effect of a variable. A relative risk ratio greater than 1 indicates a positive effect whereas a value less than 1 indicates a negative effect.

but is larger for IDPs>Refugees than Refugees>IDPs. However, neighboring international war is insignificant. Neighboring civil war and international war have a stronger positive effect on IDPs>Refugees than Refugees>IDPs in Model 1 but neighboring international war in Model 2 has no effect on either category. In short, the results confirm the inferences we garner with respect to our results reported in the paper.

Model Selection: Single Equation OLS and Tobit Specifications

As noted in the study, we assess the robustness of our findings using a single equation regression approach and a random effects tobit estimator. To begin the results of the regression and the tobit model are virtually identical. Moreover, the inferences we draw with respect to the sample selection results across the alternative estimators are the same. There are only slight differences with respect to the magnitude of the effects which are not different enough to go into detail here. The results can be replicated using the replication data set and command file.

Sample Selection Bias and Imputed Data

A final concern in any cross-national analysis involves the possibility of sample selection bias introduced via missing data. If data are missing at random, then there is no problem. Yet, if data are more likely to be missing in country-years that produce forced migration, then our estimates are likely to be biased. That missing data are more likely in cases that produce forced migrants is plausible, and therefore it is important for us to evaluate whether the results we report are sensitive to the sample we have used.

One way to do so is to impute the missing values so that we can estimate a full sample. Imputation is a method for using multiple regression to estimate the values that are missing. We used the imputation technique in Stata version 8 to estimate the missing values, thus increasing our sample size from 3,019 in model 1 and 1,972 in model 2 to 4,750 (1971-1995) and 4,143 (1976-1995), respectively. Then we re-estimated models 1 and 2 using the sample that included the imputed values.

The results are largely stable in both models, but there are some differences.¹⁹ In the outcome equation of model 1 origin democracy becomes indistinguishable from zero when we use the imputed data. Interestingly, origin democracy again changes its significance level in model 2, but this time it shifts from non-significant to significant (it retains its negative sign). Thus, we ultimately reproduce our reported findings, though the effects shift across the equations.

In addition, the proportion of borders that are mountainous no longer produces a statistically significant, positively signed parameter in model 1: it remains positive, but cannot be distinguished from zero. Further, in model 1 the neighborhood transition regimes coefficient retains its negative sign, but is statistically significant using the larger sample. In the outcome equations of both models neighborhood genocide retains its negative sign but has a large standard error and cannot be distinguished from zero. Finally, the selection equations produced no changes in sign and significance. To summarize, there are some minor changes, but nothing that undermines the inferences reported in our study.

¹⁹ The coefficients and other results can be obtained using the replication data set and associated command file.

References

- Achen, Christopher. 1986. *The Statistical Analysis of Quasi-Experiments*, Berkeley: University of California Press.
- Apodaca, Clair. 1998. "Human Rights Abuses: Precursors to Flight?" *Journal of Refugee Studies*, 11(1):80-93.
- Banks, Arthur S. nd. *Cross-National Time-Series Data Archive*, <http://www.databanks.sitehosting.net/www/main.htm>, accessed February, 2002.
- Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77:531--553.
- Crisp, Jeff. 1999. "Who Has Counted the Refugees? UNHCR and the Politics of Numbers," New Issues in Refugee Research, Working Paper #12, Geneva: UNHCR, <http://www.unhcr.ch/refworld/pub/wpapers/wpno12.htm>, accessed July, 1999.
- Davenport, Christian. 1995. "Multi-Dimensional Threat Perception and State Repression: An Inquiry into Why States Apply Negative Sanctions," *American Journal of Political Science*, 39:683-713.
- Faist, Thomas. 2000. *The Volume and Dynamics of International Migration and Transnational Social Spaces*, Oxford: Clarendon Press.
- 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.
- 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.
- Harff, Barbara and Ted Robert Gurr. 1988. "Toward Empirical Theory of Genocides and Politicides: Identification and Measurement of Cases Since 1945," *International Studies Quarterly*, September, 32(3): 359-71.
- Harff, Barbara and Ted Robert Gurr. 1996. "Victims of the State: Genocides, Politicides, and Group Repression from 1945 to 1995," in Albert J. Jongman (ed.) *Contemporary Genocides: Causes, Cases, Consequences*, Leiden: PIOOM, pp. 33-58.
- Heckman, James. 1979. "Sample Selection Bias as a Specification Error", *Econometrica* 47(1): 153-161.

Marshall, Monty, Ted Robert Gurr, and Barbara Harff. 2001. "State Failure Problem Set: Internal Wars and Failures of Governance, 1955-2000." Available at <http://www.cidcm.umd.edu/inscr/stfail/SFPScodebook.rtf>, accessed September 2004.

Moore, Will H. & Stephen M. Shellman. 2004. "Fear of Persecution: Forced Migration, 1952-95," *Journal of Conflict Resolution* 48(5):723-745.

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.

Sartori, Anne E. 2003. "An estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions," *Political Analysis*, 11(2):111-138.

Schmeidl, Susanne. 1998. "Comparative Trends in Forced Displacement: IDPs and Refugees, 1964-96," in J. Hampton (ed.) *Internally Displaced People: A Global Survey*, London: Earthscan Publications, pp. 24-33.

Schmeidl, Susanne. 2000. "The Quest for Accuracy in the Estimation of Forced Migration," in S.C. Lubkemann, L. Minear and T.G. Weis (eds.) *Humanitarian Action: Social Science Connections*, Occasional Paper #37, Thomas J. Watson, Jr. Institute for International Studies, pp. 127-57.

Schmeidl, Susanne and J. Craig Jenkins. 1999. *Global Refugee Migration Project*, <http://www.soc.sbs.ohio-state.edu/jcj/research.htm>, accessed March, 2002.

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

Signorino, Curtis S. 2002. "Strategy and Selection in International Relations." *International Interactions* 28(1): 93-115.