Chapter 4
When Civilians Flee:
The Logic of Population Control during Civil War

Cyrus Mohammadian

University of Southern California, School of International Relations

Spring 2016

Abstract

Why do some civil war torn countries produce more refugees relative to their internally displaced population and others displace more of their population internally than across borders? Surprisingly, the relationship between internally displaced persons and conflict has been woefully underexplored. The aim of this chapter is to fill this gap in the literature. Using a panel dataset of civil conflicts by country-year from 1993-2010 and a two-step Heckman selection model, I show that civil wars fought along ethnic lines produce greater refugee flows relative to IDP flows than non-ethnic civil wars. I account for this finding by relying on insights drawn from the previous chapters. Specifically, I argue that in conflicts where combatants are recruited along ethnic lines, ethnic markers allow for less costly and more discriminate targeting of rival civilian populations, which in turn increases the share of forced migrants who seek refugee across borders relative to those displaced internally.

Keywords: Ethnic Conflict; Refugees; IDP; Civil War; Forced Migration

1 Introduction

Much of the literature on refugees and conflict contagion has focused on the conditions under which refugees are more or less likely to spread violence across borders. There is some evidence, for example, that the effects of refugees on the spread of violence is conditioned by a number of factors, such as shared ethnic ties with the host population, resource scarcity, and political instability in the host state. Specifically, this dissertation has argued that refugees contribute to violence in two primary ways. First, refugees alter the existing balance of power between competing ethnoreligious groups at the substate level, which can result in increased levels of violence in the host state if such relations are polarized to begin with. Second, coethnic refugees also increase the likelihood of one-sided violence, especially at the hands of non-state actors, suggesting that their presence is a potential threat to certain rebel groups. The logic in both these circumstances is uniform; refugees represent both a threat and a boon to rebel groups. As much as coethnic refugees are an attractive target for recruitment and resource extraction for one rebel group, their presence is viewed as a threat by competing rebel groups.

What is less clear in the literature, however, are the dynamics that contribute to the emergence of refugees in the first place. In many ways, we've put the proverbial cart before the horse. We know how conflict refugees contribute to the spread of violence yet we do not know which conflicts are most likely to result in a large exodus of refugees in the first place. From the perspective of the conflict contagion literature, we know which countries are at greatest risk of contagion but we do not know which countries are the most *contagious*.

While the forced movement of people is a feature endemic to most conflict-ridden states,[1] the proportion of civil war-torn states with active refugee *flows* represent only 51% of all conflict years between 1993-2011.[2]

[1] For example, between 1993 and 2011, only one country enduring conflict featured no refugees living outside its borders, Papua New Guinea between 1993-1994.

[2] Active refugee flows refer to the number of refugees forced to flee in a given year from a given country, as opposed to refugee stock, which is a count of the cumulative total of refugees.

The range in the number of forced migrants fleeing conflicts is also non-trivial. For example, the 1994 civil war and genocide in Rwanda resulted in the flight of a little over 1.8 million refugees. Similarly, in 2006, the intercommunal violence and civil war that plagued war-torn Iraq resulted in nearly 1.2 million people fleeing their country. Compare those annual figures, with the number of refugees that fled Tajikistan's civil war in the 1990s –only 5,000 refugees were displaced over the course of five years. Of course not all victims of forced movement in conflict become refugees –a good portion of them are displaced internally. In fact, in many cases the forced displacement of victims of war within boarders exceeds the displacement of victims of war across borders. Take for example Pakistan's civil war against the Taliban insurgency, which resulted in the internal displacement of over 1.7 million people in the year 2009 alone. This stands in great contrast to the 2,279 individuals who sought and found refuge *outside* Pakistan's borders that same year.

Why do some civil war torn countries produce more refugees relative to their internally displaced population and others displace more of their population internally than across borders? Surprisingly, the relationship between internally displaced persons and conflict has been woefully underexplored. The aim of this chapter is to fill this gap in the literature. Using a panel dataset of civil conflicts by country-year from 1993-2010 and a *two*-step Heckman selection model, I show that civil wars fought along ethnic lines produce greater refugee flows relative to IDP flows than non-ethnic civil wars. I account for this finding by relying on insights drawn from the previous chapters. Specifically, I argue that in conflicts where combatants are recruited along ethnic lines, ethnic markers allow for less costly and more discriminate targeting of rival civilian populations, which in turn increases the share of forced migrants who seek refugee across borders relative to those displaced internally.

In what remains of this chapter, I introduce a theory, which I label the "the logic of population control", that accounts for variation in levels and patterns of forced migration resulting from civil wars. This logic of population control is grounded in theoretical contributions from three disparate research programs –ethnic conflict, civilian targeting, and

forced migration. In the sections that follow, I briefly review the literature in each of these research agendas in that order, I provide a stylized narrative of the theory, and I present my methodological approach to testing the observable implications of the theory. Finally, in the last two sections I provide an interpretation of the results and conclude with a discussion that ties these findings to the conflict contagion research agenda.

2 Literature Review

2.1 Ethnicity and Ethnic Civil Wars

Although the literature disaggregates civil wars in a number of different ways, such as "old vs new wars" (Kalyvas 2001), "territorial vs governmental" (Gleditsch et al. (2002), "symmetric,non-symmetric, conventional, and irregular" (Kalyvas 2005), etc., perhaps the most widely employed typology of civil war contrasts ethnic with non-ethnic violence. This study embraces this latter approach. In general, justifications for analytically separating ethnic and nonethnic conflict focus on motivations, structural conditions, and opportunities that differentiate one form from the other.[3]

Justifications based on motivation typically highlight the unique set of ethnic grievances that result from differential treatment of groups. For example, Sambanis (2001) finds evidence that ethnic war is waged in response to political grievances, while nonethnic war is associated with lack of economic opportunity. In other words, ethnic discrimination is a uniquely motivating factor in rebellion. Others argue that ethnic and religious discrimination produce grievances that result in conflicts that are more difficult to manage. For instance, Hassner (2009) suggests that religious grievances that result in territorial conflict are more intractable and difficult, if not impossible, to resolve due to the motivating beliefs of those involved over the sacredness of the territory being fought over. Similarly, Fearon

[3] This distinction has not been accepted wholesale in the field. See for example, Meueller (2000) who argues ethnic conflicts resemble other forms of violence in so far as they are waged by small groups of combatants purported to "fight and kill in the name of some larger entity."

(1995; 1998; 2004) argues that ethnic wars produce commitment problems that result in extended durations of warfare relative to nonethnic conflicts. Furthermore, Chaim Kaufmann (1996) contrasts "ethnic civil wars" with "ideological civil wars". The key difference between these two forms of war he suggests is the level of flexibility of individual loyalties. Where loyalty in ideological conflicts is quite fluid, in ethnic conflicts it is far more rigid. According to Kaufmann, the rigidity of loyalty is what makes ethnic conflicts particularly difficult to resolve peacefully.

In contrast to motivation-based explanations, justifications based on the different structural conditions of ethnic and nonethnic warfare rely on the ethnic composition of conflict prone societies. Researchers who proffer these justifications identify variation in levels of ethnic polarization and/or fractionalization as factors that condition warfare in unique ways that require tailored approaches (Reynal-Querol 2002; Garcia-Montalvo and Reynal-Querol 2004; Bhavnani and Miodownik 2009).

Finally, opportunity justifications typically focus on the conditions unique to mobilization when ethnicity is salient and highly politicized. These sorts of explanations look to the ways that ethnic markers shape the ability of groups to mobilize for warfare. Caselli and Coleman (2013) argue ethnic markers help enforce group membership by reducing "free-riding". In homogenous societies members of the losing group can easily pass themselves off as members of the winning coalition but in ethnically heterogeneous societies ethnic markers make this task far more difficult. The fact that leaders in homogenous societies understand this *ex post* dilemma reduces their incentives to mobilize *ex ante* in none-ethnically salient states. These ethnic markers not only reduce the free-riding problem, they also allow groups to easily identify and target loyal populations for their recruitment efforts. Thus, ethnicity plays a role in reducing the coordination costs associated with mobilization. These reduced barriers to mobilization, of course, result in greater risks of conflict escalation (Eck 2009). While ethnic markers provide groups with more efficient means to mobilize, they also provide combatants with more effective means of targeting pools of populations loyal to rivals. In

effect, the opportunity structures of ethnic civil wars increase incentives to target civilians (Valentino, Huth, and Black-Lindsay 2004).

The distinction between ethnic and nonethnic conflict made in this study is motivated by this very phenomenon. Thus, I embrace the opportunity-based justification for analytically separating ethnic from nonethnic conflict. If ethnic markers provide combatants with better opportunities to target the loyal population base of rival groups, then it stands to reason that civilians are at a greater risk of victimization in ethnic civil wars. The resulting atmosphere of violence should increase the incentives of forced migrants to seek refugee across international borders (as opposed to hiding amongst the general population). In the following section, I discuss the extant literature on the targeting of civilians during combat and tie it to the literature on forced migration and ethnic conflict.

2.2 Civilian Victimization

Civil war literature has identified a number of factors that contribute to the victimization of civilians during conflict, including autocratic regime types (Englehardt 1992; Harff 2003; Valentino, Huth, and Black-Lindsay 2004), use of guerrilla tactics by rebels (Valentino, Huth, and Black-Lindsay 2004), and desperation to win (Downes 2006). Scholars studying this topic have also embraced the ethnic-nonethnic distinction in examining civil wars and their analyses suggest ethnic conflict increases the chances of civilian targeting relative to nonethnic conflict (Downes 2006; Fjelde and Hultman 2010). Valentino, Huth, and Black-Lindsay (2004) argue that because "it is more difficult for individuals to disguise their ethnicity than their political affiliation" combatants are better able to discriminate between friendly and hostile civilian populations. In the absence of ethnic cleavages, combatants find it difficult to distinguish friend from foe; as a consequence, nonethnic civil wars actually result in less civilian targeting than ethnic ones (*ibid*). This has important implications for the patterns of forced migration that result from domestic warfare. The following section identifies these observable implications after briefly reviewing recent quantitative literature

on forced migration.

2.3 Forced Migration

A number of significant factors have been shown to increase the risk of forced migration in a given a country. Typically, these fall into one of two categories –push or pull factors. Push factors refer to characteristics and conditions of countries that force populations to flee their homes to safer destinations abroad. Examples of push factors include natural disasters (Drabo and Mbaye 2011) and various forms of violence (Schmeidl 2001; Moore and Shellman 2004; Davenport, Moore, and Poe 2003). Pull factors refer to attractive neighborhood characteristics that make movement across borders less costly than seeking refuge within borders (or staying put all together). These include the regime type of destination countries (Moore and Shellman 2007), shared ethnic affiliation in countries of destination (Rüegger and Bohnet 2015), and hospitable neighboring geography (Moore and Shellman 2007). An analysis of the factors that contribute to one form of forced migration (international refugees) or another form (internally displaced persons) must incorporate both of these elements. Surprisingly, the relationship between external and internal displacement has only received scant attention in the field thus far.

The only work to date that I am aware of that quantitatively compares refugee flight and IDP movement to one another is Moore and Shellman (2006). They employ a two-step Heckman model on a global panel analysis of country-years between 1976-1995. Their analysis suggests that levels of violence in neighboring states increases the proportion of IDPs flows relative to refugee flows. In their formulae, victims of displacement weigh the dangers they perceive at home against those they see in their potential points of destination. When conditions in their potential points of destination are more favorable than the conditions they face at home, they are more likely to migrate (and *vice versa*). The current analysis embraces both the methodology and theoretical foundations of Moore and Shellman's article. However, it departs from their approach in two ways. First, where Moore and Shellman's

narrative emphasizes the agency migrants enjoy in determining their own choice of destination, the present analysis examines the role that combatants play in limiting that agency and influencing that choice. Second, where Moore and Shellman examine "characteristics of countries" that affect patterns of forced migration, the present analysis also examines the *characteristics of conflicts* that affect that phenomenon. This latter point is not trivial and, in fact, represents a major departure from the methodological approach of Moore and Shellman and much of the existing literature on forced migration.

Instead of examining patterns of forced migration for all countries as others have done, I am specifically interested in the patterns of forced migration under the strategic environment that victims and combatants of civil war find themselves. Therefore, I restrict my analysis to country-years experiencing at least one civil war. I argue this exclusion criteria is justified because the choice to stay, flee to other regions, or to seek refugee in other countries is inherently different for those in an environment of organized warfare than those fleeing economic hardships, natural disasters, or other forms of political violence.[4] That difference, I contend, arises from the unique incentives combatants (i.e. rebels and governments) have to control the flow of population within and between their territories and regions controlled by their rivals. Furthermore, I suggest the type of war being waged shapes these very incentives, perhaps to a great deal. The next section explains why. In what follows, I introduce the "logic of population control", I establish the assumptions of the theory, I introduce the relevant actors and their interests, I formalize the opportunity structures that govern their behaviors, and I provide a stylized account of the theory at work.

[4] This exclusion criterion can potentially introduce bias into the model's estimates if civil war-torn and peaceful states differ in unobservable ways either related to the likelihood of forced migration or to the proportions of forced migrants that are refugees (or IDPs). Please see the Data and Methods section for a more in-depth discussion of the selection problem and the tools I use to manage the issue.

3 The Logic of Population Control

The logic of population control is based on the premise that combatants have an incentive to control populations loyal to them and to undermine the control their rivals enjoy over their own loyal populations. In the previous chapters, I examined the role of ethnicity in fomenting conflict in regions of Lebanon heavily populated by Syrian refugees. Specifically, I argued that refugees alter the balance of power between rival ethnic groups, which can result in increased mobilization efforts towards organized violence. My analysis showed that 1) higher refugee numbers increase conflict and 2) ethnically polarized regions are particularly susceptible to this threat. Refugees, like the local population, are a vital resource for the groups that they share ethnic affiliations with. Therefore, rival groups view refugees of rival ethnic groups as a threat. In Lebanon, this manifested itself in one-sided attacks on refugees, organized attacks between rival groups, and increased intercommunal violence amongst the local population.

The analyses from the previous chapters suggest that control over populations is an important goal for combatants. Asserting control over loyal populations and undercutting the support their rivals enjoy provides combatants with a comparative advantage in mobilization efforts and in resource extraction. In the same way that groups have an interest in targeting refugees of rival ethnic groups, combatants also have an incentive to target the local populations of rival ethnic groups. Thus, for the very same reason that the presence of *coethnic* refugees is more destabilizing in a host country than the presence of *non*-coethnic refugees, civil wars fought along ethnoreligious divisions are also more destabilizing than those fought along other societal divisions (i.e. class, ideology, etc.). They are destabilizing in ways that are particularly destructive towards civilians. In the previous section, I discussed the literature on civilian targeting that identified ethnic civil wars as the most violent form of civil war that civilians endure. The theory introduced here examines how this logic of population control influences the patterns of forced migration that result from ethnic and nonethnic conflict alike. Next, I turn my attention to the assumptions of the theory.

3.1 Assumptions

I begin with a number of important assumptions. First, although civilians make a 'decision' to flee (either within or across borders), ultimately "forced migration" invariably entails a level of coercion that really places more agency in the hands of perpetuators of the violence (combatants) than in the hands of its victims (refugees and IDPs). Therefore, I assume that combatants maintain some degree of influence over the decision of civilians to stay, to flee to other regions of the country, or to seek refugee in other states.

My second assumption is that population control is an important aim of combatants in civil war. Control over populations allows parties to a conflict to extract human and material resources in the form of soldiers, field doctors/nurses, taxes, and general economic production. Population control also allows combatants to homogenize their populations in support of their war aims by expelling, detaining, or killing sympathizers of the opposing group. For the very same reason that a controlled but robust and supportive population is critical for success in conflict, warring parties have an interest in undercutting the population support their rivals enjoy. One common way of doing so in the midst of warfare is directly targeting civilians.

My next assumption rests on the difference between war-torn states where ethnicity plays a salient role in the society's divisions and where it does not. The costs of and barriers to population control in ethnically salient conflicts are, as already discussed, less than in nonethnic conflicts because ethnic markers allow groups to better discriminate between supporters and those sympathetic to opposing groups. Therefore, I contend that different *types* of civil wars produce different logics of population control.

Furthermore, many civil wars increase the threat of bodily harm to civilians and they react to this threat by fleeing their homes. They can either flee to other regions of their own country or attempt to seek refugee across the border in neighboring states. I assume, all else equal, movement across international borders is more costly than internal displacement. Therefore, civilians should prefer to relocate as close to their original location of residency as

possible without exposing themselves to high risks of victimization. Ultimately, the decision to relocate internally as opposed to seeking asylum elsewhere, wrests on whether the civilian thinks his/her probability of victimization is higher outside the country than inside. That calculation, I argue, is influenced by the likelihood that combatants will target civilians, which is itself influenced by the type of civil war waged.

My final assumption is that rightly or wrongly, civilians are viewed by rebels (or governments) representing rival groups as a potentially threatening resource at the disposal of their enemies. Under conditions of non-ethnic conflict the tools of targeted repression become blunted because governments (or rebel groups) find it difficult to distinguish between loyal and disloyal pools of civilians. Because mass repression can undercut support among previously loyal populations, rival victims of war are better able to seek refugee undetected within the borders of the state by hiding among populations that warring groups are hesitant to target or find it difficult to target effectively.

In contrast, the conditions of ethnic conflict produce a different logic of repression and flight. Ethnic markers provide warring groups with the ability to not only identify supportive populations but hostile ones as well, a point Kaufmann (1996) makes when he suggests that combatants "can treat all members of the other ethnic group as enemies without risk of losing a recruit" (21). In effect, conditions of ethnic conflict limit the domestic destination options available to fleeing victims of war.

3.2 Actors and Interests

The notion that combatants have incentives to control the type of forced migration that results from their participation in ongoing violence I label the theory of population control. The narrative of this theory focuses on three actors –rebels, governments, and civilians. Each of these groups has their own interests as well. Rebels want to limit the resources of the government and one way to do so is to target populations loyal to them. Similarly, governments want to undercut the support base of opposing rebel groups, so they too benefit

from targeting civilian populations loyal to their rivals. At the same time, both governments and rebels have an incentive to safeguard loyal populations. Therefore, overzealous targeting of civilian groups that puts their own potential supporters at risk is costly. Finally, civilians too are self-interested actors whose primary goal is to reduce their personal likelihood of persecution at the hand of combatants.

Let T_l denote the probability of targeting loyal civilians and let T_d denote the probability of targeting disloyal civilians. Line 1 identifies the incentives versus the constraints combatants face under conditions of nonethnic civil war,

$$T_l = T_d \tag{1}$$

That is, when groups are mobilized along nonethnic lines, the probability of targeting loyal civilians should be roughly equal to the probability of targeting populations loyal to rivals because distinguishing between friend and foe is more difficult without observable markers that aid in differentiating one from the other. In contrast, under conditions of *ethnic* conflict, the likelihood of targeting disloyal civilians is greater than the likelihood of targeting one's own population base,

$$T_l > T_d \tag{2}$$

The balance between T_d and T_l is influenced by the ability of combatants to target rivals' civilian support discriminately. When groups are mobilized along ethnic lines, the dividends, from attacks on civilians is higher than the risks associated with civilian victimization because there are fewer costs to targeting civilians discriminately. However, when fighting takes place along the lines of non-ascriptive identities, such as class, fewer observable markers exists that reveal the loyalties of the civilian population. Under these conditions, combatants find the chances of targeting loyal civilians, T_l and disloyal ones, T_d, roughly the same. Let C denote the costs associated with targeting civilians and B the benefits. When $T_d > T_l$ then,

$$C > B \tag{3}$$

In other words, when the probability of targeting disloyal civilians (T_d) is higher than the probability of targeting loyal ones (T_l) then the costs (C) of targeting civilians is less than the benefits (B) accrued from the tactic. If, V_e, represents the level of violence against civilians during ethnic civil war and V_n the level of violence during nonethnic civil war, then,

$$V_e > V_n \tag{4}$$

Line 4 indicates that under conditions of ethnic conflict violence against civilians is greater than under conditions of nonethnic conflict. Civilians also make a cost-benefit analysis. P_o refers to the probability of perceived victimization in a civilian's country of origin and P_a refers to the probability of perceived victimization in a civilian's potential destination of asylum. When P_o are equal P_a,

$$P_o = P_a \tag{5}$$

civilians will choose to relocate to domestic destinations. That is, when the perceived chances of being targeted is the same at home that it is abroad (or in the process of traveling abroad) then civilians will elect to stay within the borders of their country. Likewise, if the perceived probability of persecution abroad is higher than at home,

$$P_o < P_a \tag{6}$$

then civilians will seek refugee in regions of the country they find safer than the battle grounds from which they escaped. On the other hand, if the perceived probability of persecution at home is higher than the perceived likelihood of victimization abroad (and in the process of traveling abroad),

$$P_o > P_a \tag{7}$$

then civilians will seek refugee across international borders. Let R represents the number of refugees fleeing civil war and I represent the number of IDPs fleeing civil war. F_e and F_n represent the share of forced migrants that are refugees for ethnic and nonethnic civil wars respectively.

$$F_e = \frac{R}{R+I} \tag{8}$$

$$F_n = \frac{R}{R+I} \tag{9}$$

Holding all else equal, combatants find $C < B$, when $T_d > T_l$ is true. And when $T_d > T_l$ is true, civilians will calculate $P_o > P_a$. It is the contention of this chapter that $T_d > T_l$ is true during ethnic conflicts more so than during nonethnic conflicts. If $P_o > P_a$ then, all else equal,

$$F_e > F_n \tag{10}$$

In other words, during conditions of nonethnic conflict the share of forced migrant flows that are composed of refugees is fewer than during conditions of ethnic conflict. Thus compared to nonethnic civil wars, civil wars characterized by ethnic cleavages are more likely to produce conflict environments that push civilians to seek asylum among foreign populations than to find shelter among their own.

3.3 Stylized Narrative

The logic of population control, thus, identifies two patterns of flight. When the costs of targeting loyal populations are higher than the benefits accrued from employing this tactic,

$C > B$, civilians become bystanders in danger of crossfire. As such, their pattern of flight can be characterized as one that avoids regions of high intensity warfare between rival groups. If$C < B$, however, civilians are not mere bystanders in war. In effect, they become targets of war and their flight patterns will reflect their intent to not only avoid regions of high intensity conflict but also peaceful places either controlled by rival adversaries or at risk of control (or targeting) by such groups.

From these two sets of equivalencies we are left with a stylized narrative that describes the conditions under which civil wars produce more or less refugees as a share of total forced migrants. Imagine for a moment a head of a household caught in the crossfire between rebel and government forces. She faces the choice to stay with her family, flee with them to more peaceful regions of the country, or make the potentially dangerous trek across international borders to safety. Therefore, she makes two choices. First, whether to stay or flee and, second, once on the move, whether to relocate to other regions or to seek refugees across the border. Her first decision is simple; she will relocate with her family in attempt to evade the threat of crossfire. Her second choice, however, depends on the perceived likelihood that the violence will follow her and her family to their choice of destination.

The perception of this likelihood is influenced by the deliberate actions of the warring parties because at the very moment she is deciding whether to simply flee or seek actual asylum, group decision makers are faced with a choice whether to target her and her family as they flee. On the one hand, if ethnic markers reveal the direction of her loyalty then the choice to target is simpler to make. If the civilian's ethnic markers reveal her to be a of a rival ethnic group then the risks associated with targeting her are lower and decision makers will likely make the choice to target. If the civilian markers identify her as supportive, then rival groups may target her instead.

On the other hand, if ethnic markers do not coincide with the cleavages of warfare, then choosing to target such a civilian is costly –you may or may not have just targeted a supporter. Thus, under conditions of nonethnic conflict the propensity to target civilians

is reduced and civilians like that head of the household use this information when deciding where to flee. If they experience deliberate targeting by combatants, they will take this as a cue that the violence will follow them to where they flee. Therefore, civilians will be more likely to flee to regions outside the reach of warring parties (i.e. outside the country). But if they are not targeted and, as such, view the threat they face in the conflict as incidental, then they will be more inclined to seek refugee in places they consider safe from crossfire. The following section posits a set of hypotheses derived from the discussion above.

4 Hypotheses

In the previous two sections I introduced the logic of population control. In this section I discuss a number of observable implications of this theory and formally present a set of hypotheses, which I empirically test in subsequent sections. The most basic observable implication of the theory is that in the midst of civil war, refugees respond to direct threats to their personal safety by relocating either to other regions within their own countries or to safe regions in other countries. From this I derive my first hypothesis:

H_1: *Controlling for all other factors, one-sided violence against civilians increases the probability of forced migration among civil war torn states.*

Moreover, for the two-step Heckman correction to work properly at least one variable must act as an "instrument" for the effect of forced migration; that is, it must predict forced migration without affecting the composition of forced migrants. In this case, one-sided violence acts as the instrument. If one-sided violence has a significant *independent marginal* effect on the composition of forced migration then its role as an instrument is compromised because it affects the probability of forced migration to begin with. But if, as I suspect, one-sided violence increases the probability of forced migration (H_1) without affecting the composition of forced migration (i.e. the share of forced migrants that are refugees), then introducing an interaction of one-sided violence and civil war type (as is suggested by H_3 and

H_4 below) should pose no problems for estimation even if the interaction itself is significant. Thus I hypothesize the following null effect:

H_2: *One-sided violence has no independent effect on the proportion of forced migrants that are refugees.*

The key proposition presented in the section discussing actors and their interests is that combatants have incentives to target civilians and when they do, refugees have a choice to flee to other regions of their home countries or to flee to safety across international borders. If the war is fought along ethnic lines, then combatants can more effectively target these civilians based on observable ethnic markets. Realizing this, civilians in the midst of ethnic conflicts will find it less attractive to seek refugee in their own countries, which brings me to my next hypothesis:

H_3 *Controlling for all other factors, compared to nonethnic civil wars, civil wars characterized by ethnic cleavages produce more refugees as a share of total forced migrants.*

As discussed earlier, ethnic markers increase the incentives combatants have to target civilians because it allows for more discriminate targeting and previous research has verified this claim. Therefore, ethnic markers act as a sort of intervening variable where the effect of one-sided violence on migrant patterns is heightened if the conflict is waged along ethnic lines. In fact, I suspect that one-sided violence against civilians increases the share of refugees relative to IDPs but only under conditions of ethnic conflict. Thus, the effect of civil war type on composition of forced migrants is conditional on the presence or absence of one-sided violence:

H_4: *Controlling for all other factors, the positive effect of ethnic civil wars on the share of forced migrants that are refugees is conditional on the presence of one-sided violence.*

If the civilians being targeted find themselves in nonethnic conflict, however, then I expect they will be more inclined to relocate to regions within their own country than to neighboring states primarily because of the difficulty in traversing international borders. If it is easier to hide among the local population, which it is under conditions of nonethnic conflict, civilians

will do so. From this I derive my next hypothesis.

H_5: *Controlling for all other factors, one-sided violence against civilians under conditions of nonethnic conflict should decrease the share of forced migrants that are refugees.*

H_6: *Controlling for all other factors, civil war type should not affect the share of forced migrants that are refugees absent one-sided violence.*

Similarly, I do not expect ethnic conflict to increase the share of forced migrants that are refugees when one-sided violence is not a feature of warfare. In fact, if the only mechanism by which ethnic conflict affects patterns of forced migration is the level of one-sided violence it produces, then it stands to reason that ethnic civil wars free of one-sided violence should have no significant effect. H_6 reflects this expectation that the effect ethnicity on the share of migrants that are refugees is conditional on the presence of one-sided violence against civilians. The next section introduces the methodological approach, estimation techniques, and data.

5 Data and Methods

5.1 Methodological Approach

The empirical strategy adopted in this chapter is primarily driven by the methodological challenge posed by selection bias in at least one stage of the analysis. The problem of sample selection is a form of omitted variable bias that arises from a nonrandom selection of data. When a subset of the data is systematically excluded due to a particular factor then exclusion of the subset can bias estimates.

This problem can emerge as an artifact of the research design or when subjects self-select into certain groups. As an example of the latter, if a researcher is interested in the effect of drug use on mental illness, simply regressing mental illness on past drug use will yield bias results if individuals use drugs to self medicate. The researcher may identify more drug use with greater mental health problems but the relationship may very well be overstated if

we think the reason for that higher drug use had to do with greater levels of mental illness to begin with. This is also known as endogenaity bias because the selection into treatment groups is endogenous to the outcome. In regards to the former, if a researcher is interested in examining the relationship between education and wage offers but only has access to wage offers of individuals currently employed then the factors that predict participation in the labor force may bias the relationship between levels of education and wage offers, specially if labor force participation and wages are related.

Both these forms of bias may exist in the current study and their threats to inference are to be taken seriously. I begin with a discussion of subject-self selection (also known as endogeneity or treatment selection) referred to above in the anecdote of the study of drug use and mental illness. I have already established that ethnic conflicts better facilitate combatant recruitment and mobilization. Not only do ethnic markers allow leaders to better commit to the rank and file, they provide leaders with assurances that their spoils of war will not be diluted by free riders ex post. Furthermore, these ethnic markers allow combatants to more effectively target civilians loyal to rival groups. Therefore, there exists at least some nominal benefit to mobilizing along ethnic lines. However, leaders may face countervailing factors that push them to mobilize along alternative cleavages instead. If leaders are capable of determining the societal fissures along which war is waged and for some unobservable reason(s) leaders that choose to mobilize their rank and file along nonethnic lines also happen to be leaders that are less willing to target civilians during warfare, then the effect of ethnic civil wars (in contrast to nonethnic wars) on the proportion of forced migrants that are refugees will be overstated.

One way to address this threat to inference is to first model the likelihood that civil war is fought along ethnic lines and then move on to examining the outcome of interest. This is may be unnecessary, however, because 1) I contend that while leaders certainly have an interest in mobilizing their rank and file according to the ways they see most beneficial to their cause, their actually ability to do so in meaningful ways is very limited. In fact, the

very politicization of ethnicity is a process that takes shape over many years and under the influence of many factors and 2) the research design faces other antecedent biases in the chain of selection that are more important to address. I turn to those next.

Although, I begin with a total sample of all country years between 1993-2010 (based on data availability), I am only interested in the pressures that forced migrants face in the midst of civil war. Thus, I subset the data to country-years experiencing an ongoing civil war. An argument could be made that excluding all none civil war states will introduce selection bias in the final estimates. But this misses the point; exclusion of observations based on a particular selection of a population of interest does not induce bias on its own. In fact, Wooldridge (2002) asserts "sample selection can only be an issue once the population of interest has been carefully specified" (551). He suggests that if the researcher is interested in a subset of a larger population then the appropriate approach is to specify a model for the part of the population based on randomly selected data from that subset. In this study, I am only interested in examining the strategic environment victims and combatants of civil wars find themselves. Therefore, selection based on my "population of interest" –civil wars –should not influence my results.[5]

Next, I need to examine whether some civil war torn states produce more refugees relative to IDPs than others and accounts for that difference. However, not all civil war-torn states experience forced migration. In fact, between 1993-2010 49% did not. Therefore, I need to further subset the population of civil war-torn states to include only those with active

[5] Indeed, the canonical heckit correction method (Heckman 1979) for identifying and addressing selection effects fails to uncover a selection bias produced by exclusion of non-civil war states; the inverse Mill's ratio of the bivariate probit estimation does not attain significance. Had selection bias been revealed at that stage as it has been in a subsequent stage, then a bivariate probit model would be used to identify both selection effects (selection into civil war states and, given a set of civil war-torn states, selection into those that produce forced migration) and the inverse Mill's ratio obtained from both selection equations would be included as additional regressors in the outcome equation that estimates the effect of civil war type on pattern of forced migration. Such an approach, a multi-stage selection model, would be warranted if the inverse Mill's ratios (identified below) of both selection equations attain significance. However, this method identifies bias only in the second stage of selection, selection into civil war-torn states that produce forced migration. In other words, excluding civil war free country-years from the sample does not bias the results, while exclusion of states that do not produce forced migrants does. Because selection at the first stage poses no threats of inference, I only present results of the standard two-stage heckit model. Please see the appendix for the results of the multi-stage selection model.

refugee and/or IDP flows. This stage of selection may bias estimates. Here I employ the two-stage heckit correction method developed by Heckman (1979). The heckit method entails identifying a "selection equation" –a probit model that estimates the likelihood that a given civil war-torn state experiences forced migration. From the selection equation I obtain the inverse of the Mill's ratio, which is the ratio of the probability density function to the cumulative distribution function of a distribution, and use it as a regressor in the outcome equation –a linear model estimated in ordinary least squares.

Given the nested nature of the panel data, I also employ robust standard errors clustered by country. Moreover, I include region fixes effects to control for unobserved region-to-region heterogeneity in estimating the first stage of the equation (selection into civil war-torn states). Finally, to control for temporal effects, I include a lag of the dependent variable in both stages of estimation. What I end up with is a two-stage heckit selection model with 361 civil war-torn country-years between 1993-2011. The following section formally introduces the model and estimation technique.

5.2 Estimation Technique

As mentioned in the previous section, my analysis involves a two-step estimation process. Given a population of war-torn states, I must first estimate the likelihood that such states produce refugee and/or IDP flows. In the subsequent step, among the remaining pool of states (civil war-torn states that produce forced migrant flows), I must identify factors that influence the share of forced migrants that are refugees. The first stage of analysis involves a basic probit regression written as follows,

$$P(D_i t = 1 | Z_i t) = \phi(Z_i t) \tag{11}$$

where D_{it} indicates forced migration ($D_{it} = 1$ if country i experienced forced migration in time t and $D_{it} = 0$ otherwise), $Z_{it} = 1$ is a vector of explanatory variables, $\lambda$is a vector of

unknown parameters, and ϕ is the cumulative distribution function of the standard normal distribution. In the second stage I include a transformation of the predicted individual probabilities (inverse Mill's ratio) as an additional regressor in a model I estimate using OLS, which is notated as,

$$f_{it}* = X_{it}\beta + u_{it} \tag{12}$$

where $f_{it}*$ denotes the ratio of refugee flows to total migrant flows in country i in time t, which is not observed if the country does produce any forced migrants in a given year. Based on equations 1 and 2, the conditional expectation of the proportion of refugees to all migrant flows given the country experienced forced migration is written as follows,

$$E[f_{it}|X_{it}, D_{it} = 1] = X_{it}\beta + E[u_{it}|X_{it}, D_{it} = 1] \tag{13}$$

If we assume the error terms are jointly normal (i.e. multivariate normal distribution) then we obtain the following,

$$E[f_{it}|X_{it}, D_{it} = 1] = X_{it}\beta + \rho\sigma_u\lambda(Z_{it}\gamma) \tag{14}$$

where ρ is the correlation between unobserved determinants of propensity to produce forced migrants and unobserved determinants of the ratio of refugee flows to total forced migrant flows u, σ_u is the standard deviation of u and λ is the inverse Mill's ratio evaluated at $Z_{it}\gamma$. Hall (2002) argues that the standard two-step estimator results in inconsistent standard error estimates. This can be overcome using a variety of robust methods.

Alternatively, rather than estimating the equations using the standard two-step process, a maximum likelihood estimator (MLE) approach can be used. [6] While the two-step method *controls* for the effect of variables in the selection equation on the outcome equation by

[6] This analysis was conducted in R's "sampleSelection" package, which uses the Newton-Raphson algorithm by default to maximize the log-likelihood function of the estimator. Alternative algorithms produced near identical results.

including the inverse Mill's ratio, the MLE approach removes the effect of the variables in the selection equation from the outcome equation altogether. [7] I present the findings of both in the results section.

A final note in regards to the estimation technique must be made before moving on to the data. For the heckman correction to successfully remove bias three key assumptions must be met. First, the standard estimation assumptions of both the outcome and selection equation equations must not be violated. Second, the selection equation must be specified well. Third, and perhaps most difficult to meet, at least one significant variable must affect the selection equation but have no independent significant effect on the outcome equation. In other words, one or more variables in the selection equation must act as an *instrument* that affects the probability of a country experiencing forced migration in a given year but not the composition of that migration (i.e. ratio of refugees to total forced migrants).

I posit that the effect of ethnic civil wars on the composition of forced migrants is conditional on the presence of one-sided violence. One-sided violence during nonethnic civil wars should either have no effect on the composition of forced migrants or should reduce the share of refugees relative to total forced migrants (i.e. increase the share of IDPs). Similarly, ethnic conflict, absent, one-sided violence, should have no effect on the composition of forced migrants. However, I also have argued that civilians respond to concerns over their personal safety by relocating to safer regions either in or out of their countries. If one-sided violence not only predicts forced migration but also influences the composition of forced migration then it cannot be used as an instrument. On the other hand, if there exists no independent marginal effect of one-sided violence on the composition of forced migrants then the instrument is valid.

[7] See Hall (2002) for more.

5.3 Data

First it must be noted that the population of cases under examination only include country-years in which at least one civil war occurs as defined by Henrikas (2016) newly released dataset on Categorically Disaggregated Civil Wars (CDC).[8] Given certain data availability issues discussed below, the population of cases is restricted to between 1993-2009, which results in 365 country-year observations. Figure 1 is a choropleth map of the 45 states ravaged by civil war during the sample time period.

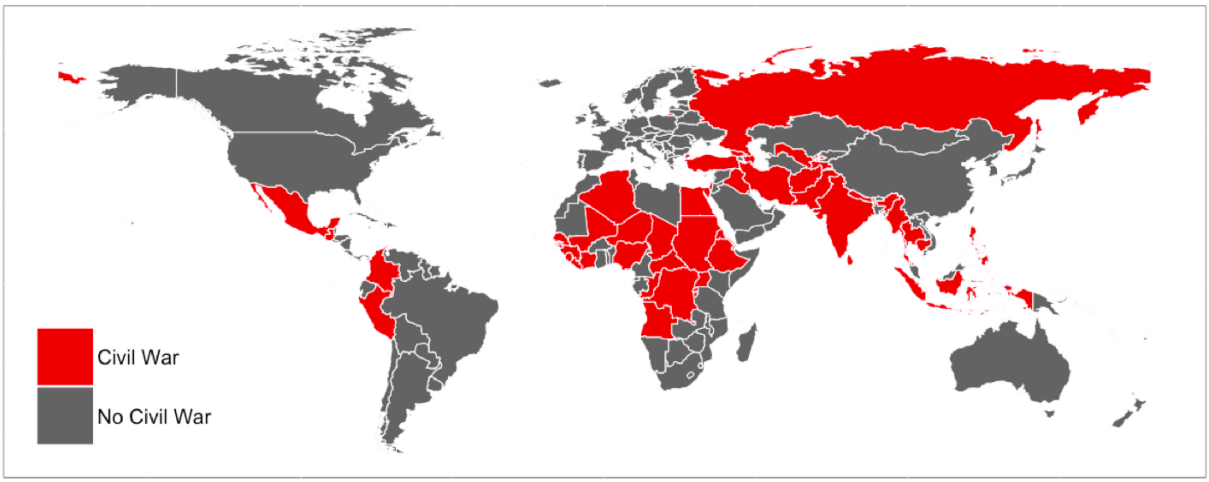

Figure 1: Conflict States 1993-2009

The dependent variable in both the selection and outcome equations is drawn from the UNHCR's database of persons of concern (UNHCR 2014). This database includes complete records of the countries of origin and asylum of refugees between 1951-2013 (incomplete data for 2014-2015). It also includes data on the number of internally displaced persons (and "persons in IDP-like situations") between 1993-2013 (incomplete data for 2014-2015). Data for both refugees and IDPs are total stocks of migrants; however, the logic of population control is concerned with active migrant flows. To obtain flows, I take the first difference between the stock of migrants in time t and the stock of migrants in time $t-1$ and then I truncate all negative values to zero.[9]

[8] See below for a more thorough discussion.

[9] This may underestimate total migrant flows because some individuals may repatriate back to their

For the selection equation, where I estimate the probability of forced migration among civil war states, I simply I add up the total number of refugee flows, IDP flows, and flows of persons in IDP-like situations and then convert this total into a binary measure; 0 for states with no active forms of forced migration and 1 for states with at least some form of forced migration (active refugee and/or IDP flows). For the outcome equation I am interested in the share of migrants that are fleeing refugees. To obtain this information I simply take the ratio of refugee flows to the total number of migrant flows. If Y_{it} is the dependent variable at time t in country i, then let R_{it} represent the number of refugee flows at time t in country i and I_{it} the number of IDP flows at time t in country i,

$$Y_{it} = \frac{R_{it}}{R_{it} + I_{it}} \tag{15}$$

Prior to introducing the data of the regressors for each equation, a note must be made in regards to the operationalization of some of these covariates. As a check for robustness, whenever a data source is a count, such as the number of individuals displaced by natural disasters, I operationalize it in two ways, either as a natural log of the original count variable or a binary variable. To obtain binary scores, I simply choose zero as a threshold. For example, to operationalize the number of individuals displaced by natural disasters as a binary variable I code all country-years with at least one individual affected by natural disasters as a 1 and all remaining country-years (i.e. those with no individuals displaced) as a 0. For the logged counts, I just add one to the base and take the natural log. Table 1 shows the raw values for each of these variables, Table 2 shows their logged values, and Table 3 shows their binary values.

The primary independent variable of interest (i.e. the instrument) in the selection equation is one-sided violence, which I draw from the Uppsala Conflict Data Program (Sundberg

homes, thereby reducing the total stock of forced migrants, which would reduce my measure of total migrant flows even if the total number of forced migrants fleeing conflict did not decrease. In effect, this measure conflates repatriation with decreases in refugee outflows. Because my population of cases is drawn from civil war-torn states I suspect the number of forced migrants returning home is likely very minuscule.

Table 1: Raw Values of Selection Equation Variables

Variable	N	Mean	St. Dev.	Min	Max
One-sided Violence	361	3,184.878	37,262.760	0	501,069
Forced Migrant Flows	365	70,796.720	248,107.600	0	2,107,111
Intercommunal Violence	361	55.413	199.246	0	2,127
Natural Disasters	361	373,037.400	733,338.500	32	4,695,110

Table 2: Binary Values of Selection Equation Variables

Variable	N	Mean	St. Dev.	Min	Max
One-sided Violence	365	0.660	0.474	0	1
Forced Migration	365	0.638	0.481	0	1
Intercommunal Violence	365	0.195	0.396	0	1
Natural Disasters	365	0.321	0.467	0	1

Table 3: Log Count Values of Selection Equation Variables

Variable	N	Mean	St. Dev.	Min	Max
log(One-sided Deaths)	361	3.525	2.832	0.000	13.125
log(Intercommunal Deaths)	361	0.974	2.042	0.000	7.663
log(Natural Disasters)	361	10.833	2.548	3.497	15.362

2009). One-sided violence is defined as "the use of armed force by the government of a state or by a formally organized group against civilians, which results in at least 25 deaths" in a single event (see UCDP one-sided violence codebook). I operationalize this as the natural log of the number of deaths after having added 1 to the base and as a binary variable where 0 indicates less than 25 civilians deaths and 1 more than 25 (see Table 2 and 3). I also include a set of control variables. These include the natural log of number of deaths associated with intercommunal violence (and its binary operationalization, see Table 2 and 3), and the natural log of the number of individuals affected (killed and displaced) by natural disasters derived from The International Emergency Disasters Database (EMDAT 2014) (and its binary operationalization, see Table 2 and 3).

I also include a set of variables that need no transformation. I control for involvement in international war by including a binary indicator of international conflict drawn from the Correlates of War (COW) dataset (Sarkees et al. 2010). I also suspect that forced migration is a feature more common to earlier stages of conflict than later ones as; thus, I control for conflict duration operationalized as the number of years since the civil war began. I also control for civil war intensity with an indicator coded as 1 if the civil war resulted in more than 1000 deaths in a given year and 0 if between 25-1000 casualties were recorded (obtained from the CDC dataset itself). Finally, I include a first order temporal lag of the DV (forced migration in time $t-1$). Table 4 shows the descriptive statics for each of these remaining variables.

Table 4: Untransformed Variables in Selection Equation

Variable	N	Mean	St. Dev.	Min	Max
Interstate War	365	0.055	0.228	0	1
Conflict Duration	365	5.638	4.640	0	17
Conflict Intensity	365	0.159	0.366	0	1

My primary independent variable of interest in this study, the *type* of civil war fought, appears in my outcome equation. Specifically, I am interested in whether countries experiencing

ethnic civil wars produce different patterns of forced migration than countries experiencing nonethnic civil wars. I rely on Henrikas (2016) newly released dataset on Categorically Disaggregated Civil Wars (CDC). I opt for the CDC data over the Ethnic Armed Conflict (EAC) dataset and the Armed Conflict Data to Ethnic Power Relations (ACD2EPR) dataset because the EAC and ACD2EPR definitions of ethnic conflict are more restrictive than the theory tested in this study necessitates. EAC and ACD2EPR code conflicts as ethnic if two conditions are met; one, combatants "explicitly pursue ethno nationalist aims, motivations and interests" and, two, the combatants "recruit fighters and forge alliances on the basis of ethnic affiliations" (Cederman, Wimmer, and Min 2010).

The CDC, in contrasts, codes a conflict as ethnic if and only if its participating groups recruit members along ethnic lines. Because we cannot observe the aims and goals of ethnic groups (at best we can take their public announcements at face value) and because the logic of population control is concerned with observable markers of ethnicity and not with war aims, the CDC dataset is better suited to this study. If combatants recruit based on ethnic affiliation then it makes more sense that they will also target civilians based on ethnic affiliation. This is a simple binary indicator, 1 if a country-year experiences ethnic conflict and 0 if it experiences nonethnic conflict. However, some states experience both ethnic and nonethnic conflicts in the same year so to manage this overlap I code any country-year with at least one ethnic conflict as a 1 (i.e. as ethnic conflict).

An additional observable implication of the theory of the logic of population control is one-sided violence is an intervening variable between civil war type and composition of forced migration. Therefore, an interaction of ethnic conflict with one-sided violence should show an increase in the share of forced migrants that are refugees relative to the interplay between nonethnic conflict and one-sided violence. In other words, not only do I suspect that one-sided violence against civilians increases the likelihood of forced migration (stage 1, selection equation), I also suspect it increases the share of refugees relative to IDPs but *only* under conditions of ethnic conflict. I operationalize this conditional effect as an interaction

term between one-sided violence and ethnic conflict in the outcome equation. Specifically, I interact the binary variable for ethnic conflict with a binary variable indicating the presence of at least one event of civilian targeting.[10] All control variables are lagged one year.

In addition to the primary explanatory variables of conflict type and one-sided violence, I introduce a host of controls in the outcome equation as well, which can be classified into one of two groups – *neighborhood* and *domestic* factors. Neighborhood factors, on the one hand, refer to attributes of a country's region that render seeking refugee across international borders more or less attractive. I expect civilians we be disinclined to seek refugee across international borders if they suspect the likelihood of being targeted is as high or higher in neighboring states as it is in his/her own country. Thus, I account for characteristics of neighboring states, which include controls for whether any neighbors of a country are experiencing civil wars (CDC dataset), one-sided violence (one-sided conflict UCDP dataset), and intercommunal conflict (non-state conflict UCDP dataset, see Sundberg 2012). I also control for borders. I expect island countries to produce fewer refugees on average so I control for whether a country is an island or not. I also suspect fewer borders increases the burden neighbors face in accommodating fleeing refugees and, therefore, should decrease their willingness to accept large influxes of refugees. Thus, I control for the number of international borders a country shares with its neighbors.

Domestic factors, on the other hand, refer to characteristics of a country that make internal displacement more or less attractive for forced migrants. For example, I control for population density in case higher density countries make internal relocation more difficult and external migration more feasible. I also include a control for regime type using the Polity IV data (Marshall and Gurr 2013). All control variables are lagged one year. In addition to these controls, I also include temporal controls in both the selection and outcome equations. In both cases, I employ a first order temporal lag of the dependent variable to control for any autocorrelation in the errors. Table 5 shows the descriptive statistics for each of these

[10] I also test this with a continuous operationalization of one-sided violence as discussed earlier.

variables. The next section presents the findings.

Table 5: Variables for Outcome Equation

Variable	N	Mean	St. Dev.	Min	Max
Density	361	114.322	115.298	5.379	408.377
Island	365	0.085	0.279	0	1
Neighboring Civil War	365	0.597	0.491	0	1
Neighboring Intercommunal Violence	365	0.405	0.492	0	1
Neighboring One-sided Violence	365	0.699	0.459	0	1
Number of Borders	365	4.984	2.888	0	14
Refugee:Total	365	0.493	0.492	0.000	1.000
Ethnic Civil War	365	0.647	0.479	0	1

6 Analysis

6.1 Organization of Results

I have estimated a number of Heckman style models using two different estimation techniques and two different operationalizations of the control variables; the results of only some of these models are presented in this section. One set of models is estimated using the traditional two-step method and the other set is estimated using MLE (Newton-Raphson algorithm). Furthermore, one set of models includes binary operationalizations of the control variables and the second uses the natural log of raw counts. What I end up with is 1) a model with dummy controls estimated using the two-step method, 2) model with dummy controls estimated using MLE, 3) a model with log counts estimated using the two-step method, and 4) another model with log counts estimated using MLE. Furthermore, each of these models contains two stages of estimation –a selection and outcome equation. I also estimate models with and without the interaction terms.

Including the two separate equations for each model, a total of sixteen equations are estimated. Instead of reporting the results for all sixteen models, I present the findings for the binary response variable models only because the results are robust across the two

different operationalizations. The results for the model using log count explanatory variables can be found in Appendix B. The results are also largely robust to the two different estimation techniques; however, the findings of the two-step method and MLE did diverge in a couple of noticeable ways. Although the literature on selection models suggests MLE estimates are more consistent and robust (Hall 2002), I present results for both.

Table 6 lays out the components of each model and identifies whether it is findings are located in the results section or the Appendix. The models shown in Table 6 include the full two-stage model (i.e. not its selection and outcome components). Counting the outcome and selection models for each model in Table 6 results in 19 separate equations (the three-stage model has three equations). Next I turn my attention to the results of the eight equations of interests –selection and outcome stage models with and without interaction terms estimated using MLE and 2step estimation and a set a binary control variables (the first four models in bold found in Table 6).

Table 6: Organization of Model Results

Model	Variable	Estimation	Effects	Section
Model 1	**Binary**	**2step**	**Marginal**	**Results**
Model 2	**Binary**	**2step**	**Conditional**	**Results**
Model 3	**Binary**	**MLE**	**Marginal**	**Results**
Model 4	**Binary**	**MLE**	**Conditional**	**Results**
Model 5	logCount	2step	Marginal	Appendix
Model 6	logCount	2step	Conditional	Appendix
Model 7	logCount	MLE	Marginal	Appendix
Model 8	logCount	MLE	Conditional	Appendix
Three-stage	Binary	2step	Marginal	Appendix

7 Results

I begin by examining the results of the marginal effects selection model (i.e. without interaction effects) in the first stage of estimation. In this stage of estimation, I have regressed

the probability of forced migration on the presence of international conflict, intercommunal conflict, one-sided violence, natural disasters, and civil war intensity. Model 1_{MLE} and Model 2_{2step} of Table 7 shows the selection results of the ML and 2step estimation respectively for the model without interaction effects in its outcome stage. The positive sign and statistical significance of the coefficients of one-sided violence suggest that civilians respond to the risk of persecution by fleeing, either as refugees or IDPs, which confirms H_1. Episodes of of forced migration in the previous year also increase the chances of forced migration, suggesting that 1) similar factors likely persist over the course of a conflict that push civilians to flee and 2) civilians learn from the past to inform their decisions of the future (if they see others persecuted and fleeing in time $t-1$ they will be more like to flee themselves in time t). Forced migration is no more or less likely to occur earlier in conflicts, while occurrence of natural disasters, intercommunal violence, international conflict, and war intensity also all fail to reach statistical significance. It appears as though the number one factor that influences forced migration is the deliberate and organized targeting of civilians.

Now I turn to the outcome equation. ρ and σ in Model 1 are both significant, suggesting that selection bias is present absent a well-specified selection equation in the first stage. In other words, Model 1's outcome equation confirms that using a two-stage Heckman like selection model is the appropriate approach –a finding that is also confirmed by Model 2's outcome equation, whose inverse Mill's ratio is also significant.

A number of factors affect the composition of forced migrants. For example, island countries produce significantly more IDPs as a share of total forced migrants than countries with territorial borders. More importantly, the nature of conflict divisions –whether a civil war is fought along ethnic lines or not –also affects the ratio of refugees to IDPs. The positive sign and statistical significance of the coefficients of the ethnic civil war variable suggests that states embroiled in ethnic civil war produce more refugees relative to total forced migrants than states embroiled in nonethnic civil war, which confirms H_3. This is true of the models estimated using both techniques (MLE and 2step). H_2 posits that although

Table 7: Marginal Effects Models

	Model 1_{MLE}		Model 2_{2step}	
	Selection	Outcome	Selection	Outcome
Civil War Intensity	−0.10		−0.10	
	(0.19)		(0.20)	
Forced Migration$_{t\text{-}1}$	0.55**		0.54**	
	(0.15)		(0.15)	
Interstate War	0.24		0.23	
	(0.29)		(0.29)	
Conflict Duration	−0.02		−0.02	
	(0.02)		(0.02)	
Intercommunal Violence	0.03		0.05	
	(0.17)		(0.18)	
Natural Disasters	−0.19		−0.09	
	(0.15)		(0.15)	
Civilian Targeting	0.68**	−0.12	0.69**	−0.12
	(0.15)	(0.07)	(0.15)	(0.07)
Ethnic War		0.18**		0.15**
		(0.07)		(0.07)
Population Density		0.00		0.00
		(0.00)		(0.00)
Island		−0.39**		−0.37
		(0.13)		(0.14)
Civil War$_{Neighborhood}$		0.04		0.04
		(0.07)		(0.08)
Intcom.Viol$_{Neighborhood}$		−0.05		−0.05
		(0.06)		(0.07)
Civil.Targ$_{Neighborhood}$		−0.05		−0.04
		(0.08)		(0.08)
Borders		−0.02		−0.02
		(0.01)		(0.01)
Ref:Forced Migrants$_{t\text{-}1}$		0.35**		0.35**
		(0.08)		(0.08)
(Intercept)	−0.29	0.67**	−0.22	0.02
	(0.16)	(0.12)	(0.16)	(0.35)
σ		0.39**		
ρ		0.47**		
invMillsRatio				0.88**
Adj. R^2		0.17		0.14
Num. obs.	227	365	227	365

Note: *p<0.05; **p<0.01

one-sided violence increases the chances of forced migration (H_1), it has no independent marginal effect on the composition of forced migrants (i.e. the share that is composed of either IDPs or refugees). Both models fail to uncover any statistically significant relationship between one-sided violence the dependent variable of the outcome equation, which confirms H_2. This is an extremely important finding because had such an effect existed, one-sided violence could not be used as an instrument in the selection equation, while the estimates of the outcome equation would remain biased (i.e. the selection bias would persist).

Surprisingly, there is no evidence to suggest any of the remaining neighborhood effects influence the share of forced migrants that are refugees. The number of international borders, the occurrence of civil war, intercommunal violence, or one-sided violence in neighboring states has no statistically significant effect on the composition of forced migrants. Although surprising, this may be accounted for by the fact civilians likely have incomplete information in regards to the likelihood of persecution abroad. Their impression of where the safest place to flee is likely formed by information they have about conditions at home and not those abroad. Therefore, we would civilians react to *domestic* factors rather than *neighborhood* ones when information is scarce -a typical feature of warfare. Regime type and population density also fail to reach significance.

I now turn to the results of the conditional effects model. Because the interaction effects are not introduced until the second stage, the results of the conditional effects selection equation are nearly identical to the results of the marginal effects selection equation. Here, I am only interested in the *conditional* effects of ethnic conflict and one-sided violence on the share of forced migrants that are refugees. To do this, I can examine the interaction between civil war type and one-sided violence. Table 8 shows the results of the MLE and 2step models, both of which identify a statistically significant relationship between the shared of forced migrants that are refugees and the interaction between civil war type and one-sided violence, which confirm H_4 and H_5.

Moreover, the marginal effect of ethnic civil war when one-sided violence does *not* occur

Table 8: Conditional Effects Models

	Model 3_{MLE}		Model 4_{2step}	
	Selection	Outcome	Selection	Outcome
Civil War Intensity	−0.10		−0.10	
	(0.19)		(0.20)	
Forced Migration$_{t\text{-}1}$	0.53**		0.41**	
	(0.15)		(0.15)	
Interstate War	0.24		0.15	
	(0.29)		(0.30)	
Conflict Duration	−0.02		−0.02	
	(0.15)		(0.02)	
Intercommunal Violence	0.03		0.05	
	(0.18)		(0.18)	
Natural Disasters	−0.20		−0.09	
	(0.15)		(0.15)	
Civilian Targeting	0.68**	−0.27**	0.69**	0.04
	(0.15)	(0.09)	(0.15)	(0.18)
Civilian Target X Ethnic War		0.26**		0.25**
		(0.12)		(0.08)
Ethnic War		0.18**		−0.03
		(0.07)		(0.05)
Population Density		0.00		0.00
		(0.00)		(0.00)
Island		−0.41**		−0.39
		(0.13)		(0.14)
Civil War$_{Neighborhood}$		0.05		0.04
		(0.07)		(0.14)
Intcom.Viol$_{Neighborhood}$		−0.05		−0.04
		(0.06)		(0.06)
Civil.Targ$_{Neighborhood}$		−0.07		−0.06
		(0.08)		(0.08)
Borders		−0.02		−0.02
		(0.01)		(0.02)
Ref:Forced Migrants		0.25**		0.34**
		(0.05)		(0.08)
(Intercept)	−0.29	0.79**	−0.22	0.15
	(0.16)	(0.13)	(0.16)	(0.35)
σ		0.39**		
ρ		0.50**		
invMillsRatio				0.89**
Adj. R^2		0.17		0.14
Num. obs.	227	365	227	365

Note: *p<0.05; **p<0.01

is insignificant, which also confirms H_6. The one difference that arises between the two estimation techniques, is that the marginal effect of one-sided violence on forced migrant composition is insignificant for nonethnic civil wars when estimated using the 2step method but significant and negative when estimated using the MLE approach. The MLE results suggest one-sided violence during nonethnic civil war increases the share of IDPs rather than the share of refugees. Although not shown in the coefficient tables, the marginal effect of one-sided violence on the composition of forced migrants during ethnic civil wars is negative and statistically significant. These findings suggest that civilians fleeing ethnic civil wars have more incentives to seek refugees across borders than in "safer" regions of their home countries if the violence they are fleeing is specifically directed towards them. This has important implications for the spread of conflict, which I turn to in my concluding section.

8 Conclusion

In this chapter I have introduced a theory, which I label the *logic of population control*, that accounts for some of the variation in patterns of forced migration resulting from on going civil wars. The logic of population control draws on the findings of the previous chapters as well as the extant literature on ethnic conflict and civilian targeting.

In the previous chapters I argued that coethnic refugees contribute to increased levels of violence because they can alter the delicate *ethnic balance of power* in host states with volatile ethnic relations. Because refugees are a potential source of recruitment for coethnic groups in the host state, rival groups, rightly or wrongly, view them as a threat. It is for this very same reason that combatants also view local civilians perceived as loyal or sympathetic to their rivals as a threat. However, whether combatants target civilians loyal to their rivals largely depends on if they can determine whether the civilian is a supporter or not. The literature on civilian targeting suggests ethnic markers aid in this process of identification.

If ethnic markers increase violence against civilians, then civilians fleeing ethnic conflict

are more constrained in their choice of destination. Simply relocating to regions away from the crossfire may not be enough; combatants may target the regions they flee to as well. Therefore, I have hypothesized that states suffering ethnic civil wars on average should produce more refugees relative to IDPs than states afflicted with nonethnic civil war.

To test this proposition I applied a two-step Heckit style selection model to a population of all country-years experiencing civil war between 1993-2010. The results of my analysis confirm my expectations that a statistically significant difference exists between patterns of forced migration resulting from ethnic and nonethnic civil wars.

The finding that ethnic conflicts contribute to refugee flight across borders has serious implications for the study of conflict contagion more generally. Previous literature has examined the conditions under which states produce more or less refugees, but these studies, with one or two notable exceptions, have failed to appreciate the counter factual to refugee flight, which is internal displacement.

Civilians face two primary choices during conflict that can impact their very survival; 1) to stay put (either as a civilian or to take up arms) or to flee, and 2) whether to flee to other regions of their own countries or to seek asylum in foreign countries. The current literature on forced migration either aggregates refugees and IDPs into a single category or focuses exclusively on refugee flows. Aggregating all forms of forced migration into a single category allows us to identify the factors that contribute to civilian flight, the first choice civilians in conflict make, but it cannot speak to the second choice they face -their destination of refugee. Moreover, examining only refugee flows leads researchers to under appreciate how IDP flows can also contribute to conflict contagion. For example, opposition groups in neighboring states may find safe zones for IDPs to be useful as safe havens for their own activities.[11] Thus, examining refugees and IDPs together but as analytically distinct elements can help match different mechanisms of conflict contagion with different patterns of forced migration.

[11] The Kurdish PKK in Turkey has found support in the safe zones along Syria's Turkish borders where many Kurdish IDPs have amassed. For more information on how the PKK has used these regions in Syria see http://www.al-monitor.com/pulse/security/2016/04/turkey-pkk-clashes-last-stronghold.html For more on IDPs and the spread of conflict see Bohnet, Cottier, and Hug (2012)

References

Bartusevičius, Henrikas. 2016. "Introducing the Categorically Disaggregated Conflict (CDC) dataset." *Conflict Management and Peace Science* 33(1):89–110.

Bhavnani, Ravi and Dan Miodownik. 2009. "Ethnic polarization, ethnic salience, and civil war." *Journal of Conflict Resolution* 53(1):30–49.

Bohnet, Heidrun, Fabien Cottier and Simon Hug. 2013. "Conflict-induced IDPs and the spread of conflict." *presentation at the European Political Science Association (EPSA) in Barcelona* .

Caselli, Francesco and Wilbur John Coleman. 2013. "On the theory of ethnic conflict." *Journal of the European Economic Association* 11(s1):161–192.

Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. "Why do ethnic groups rebel? New data and analysis." *World Politics* 62(01):87–119.

Database, The International Emergency Disasters. 2016. "EM-DAT Database.". **URL:** *http://www.emdat.be/database*

Davenport, Christina, Will Moore and Steven Poe. 2003. "Sometimes you just have to leave: Domestic threats and forced migration, 1964-1989." *International Interactions* 29(1):27–55.

Downes, Alexander B. 2006. "Desperate times, desperate measures: The causes of civilian victimization in war." *International Security* 30(4):152–195.

Drabo, Alassane and Linguère Mbaye. 2011. "Climate change, natural disasters and migration: An empirical analysis in developing countries.".

Eck, Kristine. 2009. "From Armed Conflict to War: Ethnic Mobilization and Conflict Intensification*." *International Studies Quarterly* 53(2):369–388.

Engelhardt, Michael J. 1992. "Democracies, Dictatorships and Counterinsurgency: Does Regime Type Really Matter?" *Journal of Conflict Studies* 12(3).

Fearon, James D. 1995. "Rationalist explanations for war." *International organization* 49(03):379–414.

Fearon, James D. 1998. "Commitment Problems and the Spread of Ethnic Conflict. The International Spread of Ethnic Conflict. DA Lake and D. Rothchild." *Princeton, NJ, Princeton University Press* 52:269–305.

Fearon, James D. 2004. "Ethnic mobilization and ethnic violence." *Barry R. Weingast and* .

Fjelde, Hanne and Lisa Hultman. 2010. Weakening Your Enemy: Constituencies and the location of violence against civilians in Africa, 1989-2006. In *SGIR 7th Pan-European International Relations Conference*.

Garcia-Montalvo, Jose and Marta Reynal-Querol. 2004. "Ethnic polarization, potential conflict, and civil wars." *Potential Conflict, and Civil Wars* .

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. "Armed conflict 1946-2001: A new dataset." *Journal of peace research* 39(5):615–637.

Hall, Bronwyn H. 2002. "Notes on Sample Selection Models.".
URL: *http://eml.berkeley.edu// bhhall/e244/sampsel.pdf*

Harff, Barbara. 2003. "No lessons learned from the Holocaust? Assessing risks of genocide and political mass murder since 1955." *American Political Science Review* 97(01):57–73.

Hassner, Ron Eduard. 2009. *War on sacred grounds*. Cornell University Press.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica* 47(1):153–161.

Kalyvas, Stathis. 2001. "New and old civil wars." *World politics* 54(1):99–118.

Kalyvas, Stathis N. 2005. "Warfare in civil wars." *Rethinking the nature of war* pp. 88–108.

Kaufmann, Chaim. 1996. "Possible and impossible solutions to ethnic civil wars." *International security* 20(4):136–175.

Marshall, Monty G. and Ted R. Gurr. 2013. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2013.".
URL: *http://www.systemicpeace.org/polity/polity4.htm*

Moore, Will H and Stephen M Shellman. 2004. "Fear of Persecution Forced Migration, 1952-1995." *Journal of Conflict Resolution* 48(5):723–745.

Moore, Will H and Stephen M Shellman. 2007. "Whither will they go? A global study of refugees' destinations, 1965–1995." *International Studies Quarterly* 51(4):811–834.

Mueller, John. 2000. "The banality of "ethnic war"." *International Security* 25(1):42–70.

Reynal-Querol, Marta. 2002. "Ethnicity, political systems, and civil wars." *Journal of Conflict Resolution* 46(1):29–54.

Rüegger, Seraina and Heidrun Bohnet. 2015. "The Ethnicity of Refugees (ER): A new dataset for understanding flight patterns." *Conflict Management and Peace Science* .

Sambanis, Nicholas. 2001. "Do ethnic and nonethnic civil wars have the same causes? A theoretical and empirical inquiry (Part 1)." *Journal of Conflict Resolution* 45(3):259–282.

Sarkees, Meredith Reid and Frank Wayman. 2010. "Resort to War: 1816 - 2007." *Conflict Management and Peace Science* .

Schmeidl, Susanne. 2001. "Conflict and forced migration: a quantitative review, 1964-1995." *Global migrants, global refugees: Problems and solutions* pp. 62–94.

Sundberg, Ralph. 2009. Revisiting one-sided violence–a global and regional analysis. In *States in Armed Conflict*, ed. Lotta Harbom and Ralph Sundberg. Department of Peace and Conflict Research.

Sundberg, Ralph, Kristine Eck and Joakim Kreutz. 2012. "Introducing the UCDP non-state conflict dataset." *Journal of Peace Research* 49(2):351–362.

UNHCR. 2014. "Refugee Statistics and Operational Data.".
URL: *http://www.unhcr.org/pages/4a013eb06.html*

Valentino, Benjamin, Paul Huth and Dylan Balch-Lindsay. 2004. ""Draining the sea": Mass killing and guerrilla warfare." *International Organization* 58(02):375–407.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Appendix

Appendix A: Marginal Effects Models with Log Counts

	Model 5_{MLE}		Model 6_{2step}	
	Selection	Outcome	Selection	Outcome
Civil War Intensity	−0.24		−0.11	
	(0.20)		(0.21)	
Forced Migration$_{t\text{-}1}$	0.50**		0.35*	
	(0.15)		(0.15)	
Interstate War	0.19		0.23	
	(0.29)		(0.32)	
Conflict Duration	−0.02		−0.02	
	(0.01)		(0.02)	
log(Intercommunal Deaths)	0.00		0.01	
	(0.03)		(0.04)	
log(Natural Disaster Victims)	0.03		−0.04	
	(0.04)		(0.03)	
log(Civilian Deaths)	0.11**		0.11**	
	(0.03)		(0.03)	
Ethnic War		0.17**		0.17**
		(0.06)		(0.06)
Population Density		0.00		0.00
		(0.00)		(0.00)
Island		−0.34**		−0.39**
		(0.13)		(0.13)
Civil War$_{Neighborhood}$		0.05		0.05
		(0.07)		(0.07)
Intcom.Viol$_{Neighborhood}$		−0.07		−0.05
		(0.06)		(0.07)
Civil.Targ$_{Neighborhood}$		−0.04		−0.04
		(0.08)		(0.08)
Borders		−0.01		−0.02
		(0.01)		(0.01)
Ref:Forced Migrants$_{t\text{-}1}$		0.25**		0.30**
		(0.05)		(0.08)
σ		0.41**		
ρ		0.60**		
invMillsRatio				0.50**
Adj. R^2		0.14		0.13
Num. obs.	227	365	227	365

Note: *p<0.05; **p<0.01

Appendix B: Conditional Effects Models with Log Counts

	Model 7_{MLE}		Model 8_{2step}	
	Selection	Outcome	Selection	Outcome
Civil War Intensity	−0.19		−0.11	
	(0.21)		(0.21)	
Forced Migration$_{t\text{-}1}$	0.45**		0.35*	
	(0.15)		(0.15)	
Interstate War	0.21		0.23	
	(0.31)		(0.32)	
Conflict Duration	−0.02		−0.02	
	(0.02)		(0.02)	
log(Intercommunal Deaths)	0.00		0.01	
	(0.04)		(0.04)	
log(Natural Disaster Victims)	−0.03		−0.04	
	(0.04)		(0.03)	
log(Civilian Deaths)	0.10**	−0.05	0.11**	−0.05
	(0.03)	(0.02)	(0.03)	(0.02)
Civilian Deaths X Ethnic War		0.18*		0.19*
		(0.08)		(0.07)
Ethnic War		−0.03		0.17**
		(0.10)		(0.06)
Population Density		0.00		0.00
		(0.00)		(0.00)
Island		−0.41**		−0.41**
		(0.13)		(0.13)
Civil War$_{Neighborhood}$		0.05		0.05
		(0.07)		(0.07)
Intcom.Viol$_{Neighborhood}$		−0.04		−0.05
		(0.06)		(0.06)
Civil.Targ$_{Neighborhood}$		−0.06		−0.04
		(0.08)		(0.08)
Borders		−0.02		−0.02
		(0.01)		(0.01)
Ref:Forced Migrants$_{t\text{-}1}$		0.04**		0.30**
		(0.01)		(0.08)
σ		0.37**		
ρ		0.40*		
invMillsRatio				0.17*
Adj. R^2		0.14		0.15
Num. obs.	227	365	227	365

Note: *p<0.05; **p<0.01

Appendix C: Three-Stage Heckit

	$\text{Selection}_{\textit{Stage 1}}$
(Intercept)	−0.75
	(0.70)
log(Population)	0.16*
	(0.07)
Polity	−0.00
	(0.02)
Polity^2	0.00
	(0.00)
Regime Stability	−0.20
	(0.23)
Ethnic Fractionalization	0.38
	(0.37)
log(Mountainous Terrain)	0.03
	(0.07)
Infant Mortality	0.33*
	(0.14)
$\text{Civil War}_{\textit{Neighborhood}}$	−0.13
	(0.08)
Peaceyears	−2.57**
	(0.24)
Peaceyears2	0.57**
	(0.07)
Peaceyears3	−0.04**
	(0.01)
Slow Growth	0.18
	(0.19)
AIC	266.86
Log Likelihood	-120.43
Num. obs.	2,720

$^{**}p < 0.01$, $^{*}p < 0.05$

Appendix C: Three-Stage Heckit

	$\text{Selection}_{Stage\ 2}$	Outcome
Civil War Intensity	−0.10	
	(0.20)	
Forced Migration$_{t\text{-}1}$	0.41	
	(0.15)	
Interstate War	0.15	
	(0.30)	
Conflict Duration	−0.03	
	(0.02)	
log(Intercommunal Deaths)	0.05	
	(0.18)	
log(Natural Disaster Victims)	−0.09	
	(0.15)	
log(Civilian Deaths)	0.70**	
	(0.03)	
Ethnic War		0.17**
		(0.06)
Population Density		0.00
		(0.00)
Island		−0.40
		(0.13)
$\text{Civil War}_{Neighborhood}$		0.03
		(0.07)
$\text{Intcom.Viol}_{Neighborhood}$		−0.04
		(0.06)
$\text{Civil.Targ}_{Neighborhood}$		−0.05
		(0.07)
Borders		−0.02
		(0.01)
Ref:Forced Migrants$_{t\text{-}1}$		0.30**
		(0.06)
invMillsRatio	−0.03	0.56**
Adj. R^2		0.14
Num. obs.	365	227
Note:		*p<0.05; **p<0.01