War and internal displacement

October 1, 2024

Abstract

Where do people go when they are forcibly displaced by conflict? Existing work on conflict does not focus on predicting the locations people will choose to shelter, which makes it challenging to guide policy to better target aid. This paper studies how war displaces people geographically and over time. I develop a network model that exploits data with exact geographical coordinates and large spacetime variation during the war against ISIL between 2014 and 2017 in Iraq. I find the highest concentration of displaced people within two miles of conflict, decreasing with distance and time, and disappearing beyond seventy miles. People choose locations in highly populated areas within five miles of a main road. Areas with more diversity host more displaced people relative to areas with a clear ethno-religious majority.

JEL Classification: H56, D85, C31,

Keywords: conflict, displacement, geographical.

Word count: 14,005

Introduction

Forced displacement is one of the most important consequences of conflict, yet the forecasting tools required to anticipate and assist people fleeing from violence are limited. At the end of 2020, 82.4 million people around the world had been forced to flee their homes because of prosecution, war, or violence[1].

Academic research investigating displacement exist, but it is not always suitable to be readily utilized by policy makers when conflict happens. Policy recommendations tend to be derived from case studies and other literature produced by international aid organizations, governments, and NGOs (Chatty, 2003; Margesson et al., 2008; Chatty and Mansour, 2011a,b; Marfleet, 2011). The lack of displacement forecasting tools has led to devastating humanitarian disasters which could have been avoided. For example, in March 2003, the international humanitarian aid regime had estimated that more than a million Iraqis would be forcibly displaced because of the military conflict between the coalition led by the United States and the government of Saddam Hussein. By November 2004, few had seek shelter in the camps leaving the estimated exodus unfulfilled and, as a result, camps were dismantled, and pre-positioned food and other aid removed. In 2006, hundreds of thousands of Iraqis were unexpectedly displaced leading to a humanitarian crisis. This paper focuses on predicting forced displacement using precise geo-referenced data rather than the causal identification of the parameters estimated. In the context of forced displacement, prediction itself has very high policy value to mitigate the impact of violence and guide post-conflict recovery.

I develop a series of hypothesis about the relationship between the location of conflict events and displaced population; and between the location of forced displacement, and that location's time-invariant characteristics. I test these hypothesis using high-frequency data with precise geographical coordinates on the war in Iraq between 2014 and 2017. Conflict data is commonly available when a new war breaks out, but extraordinarily, data on the location of forced displacement was also publicly available through the International Organization for Migration Displacement Tracking Matrix. In 2014, millions of internally displaced persons were forced out of their homes when the Iraqi insurgency escalated into civil war as the Islamic State of Iraq and the Levant (ISIL) seized control over major Iraqi cities. Targeting civilians is commonly used as a tactic of war (Lupu and Wallace, 2023), and presumably a leading explanation on why the location of forcibly displaced people is not generally available to researchers even when the data asset exists.

The paper yields three primary results. First, I find that the highest concentration of internally displaced people happens within 2 miles of conflict and decreases with distance, disappearing beyond 70 miles. People remain extremely close to conflict, which might pose a challenge

to aid delivery as it may be captured by combatant factions, or be difficult to be delivered due to conflict intensity itself. I also find that displaced people tend to cluster in highly populated areas, within 5 miles of a main road. This finding is aligned with the proximity to conflict which tends to be in well connected and populated areas but important in the sense that it tells us that people are not leaving to less populated areas nearby, but rather highly populations nearby that them as locals deem safer than their usual residence.¿. Finally, I find that non-diverse areas host fewer displaced people relative to areas without a clear ethno-religious majority.¡This paragraph needs a small 'conclusion' sentence after each finding¿

Conflict and forced displacement location

I contribute to to the conflict literature by answering the fundamental question of ***where*** internally displaced people go as a function of conflict considering space and time. This paper is the first empirical paper studying the location of forced displacement as a response to conflict, and focuses on prediction using precise geo-referenced data rather than exploring the exact mechanism behind people's decision to flee. In the context of forced displacement, prediction itself has very high policy value to anticipate and respond to the crises internally displaced persons and refugees face. The findings offer insights which can guide future humanitarian policy on where and when aid is more effective.

Blattman and Miguel (2010) acknowledge that academic research on conflict tends to be limited and tends to focus on high-level questions, which are not very useful to guide policy. For instance, Bellows and Miguel (2009) find that individuals whose households directly experienced more intense war violence during the 1991-2002 Sierra Leone civil war are robustly more likely to attend community meetings, more likely to join local political and community groups, and more likely to vote; Cederman et al. (2009) examine ethnic wars and show that the conflict probability of marginalized groups increases with the demographic power balance compared to the group in power and that this probability increases with the distance from the group to the capital; or Voors et al. (2012) who examine the impact of exposure to conflict on social, risk, and time preferences.

Nonetheless, some examples of papers which target questions more relevant to policy makers can be found. Cortés et al. (2013) study which policies are more effective for reducing conflict in Colombia and find that increases in roads provision reduces conflict while education does not. Nunn and Qian (2014) study the effects of US food aid on conflict in recipient countries to find that an increase in United States' food aid increases the incidence and duration of civil conflicts. Jaeger and Paserman (2008) examine how violence against Israelis and Palestinians

affects the incidence and intensity of each side's reaction.Blanco and Vargas (2014) designed and implemented a randomized controlled trial to assess the extent to which informational barriers are responsible for the prevalent low take-up of government benefits among Colombian conflict-driven internally displaced persons. Angrist and Kugler (2008) find that rural areas in Colombia that saw accelerated coca production subsequently became considerably more violent. Rogall (2014) studies the 1994 Rwandan Genocide and shows that the genocide could have been stopped with military intervention targeting the various armed groups, which constituted only 10% of the perpetrators but were responsible for at least 83% of the killings. The persistence of violence worldwide and the fruitlessness of the efforts to stop it are is an indication that we there is still much research to conduct. In particular, Blattman and Miguel (2010) argue that more evidence of the educational, employment, and health impacts of conflict on armed group participants and civilians, including internally displaced persons (IDPs) is required.

Additionally, there is an extensive literature on migration which uses refugees in its analyses. However, these papers are concerned with the impact of forced migration on the native or receiving population/economy. Refugee waves are used as an exogenous source of variation in labor supply to identify the effects of *im*migration. These papers do not tend to have the decision process of the forced migrants as a central feature. At most, these papers address the decision to migrate of the refugees as a more or less sophisticated cost-benefit analysis of expected utility at home and expected utility abroad. The impact of the Mariel boatlift on the Miami labor market has been studied by many, for example Card (1990), Bodvarsson et al. (2008) and Borjas (2017). Friedberg (2001) studies the impact of the exodus of Soviet Jews on the Israeli labor market[2]. Angrist and Kugler (2003) use refugees from the Balkan wars to show that reduced flexibility in immigration laws increases the negative impact of immigration. Borjas and Monras (2016) study the three cases mentioned above: the impact of Cubans on Miami's labor market in 1980, the impact of Soviet Jews émigrés to Israel in the 1990s, the Balkan Wars refugees' impact on different European countries between 1991 and 2001, and the influx of French repatriates and Algerian nationals into France in the aftermath of the Algerian Independence War in 1962. Other recent examples include Del Carpio and Wagner (2015), Tumen (2016) and Konuk et al. (2016) all of whom study the impact of Syrian refugees on the Turkish labor market. Maystadt and Duranton (2018) study the effects of refugee flows originating from Burundi and Rwanda on Tanzanian households. Migration papers focused on forcibly displaced persons themselves can also be found in the literature. Åslund and Rooth (2007) investigate the long-term effects on immigrant earnings and employment from labor market conditions encountered upon arrival. Bevelander and Pendakur (2014) assess the employment and earnings trajectories of refugee and family reunion category immigrants in Canada and Sweden.

When it comes to internally displaced persons (IDPs), the literature is even more limited.

Some exceptions include Blanco and Vargas (2014) who designed and implemented a randomized controlled trial to assess the extent to which informational barriers are responsible for the prevalent low take-up of government benefits among Colombian conflict-driven internally displaced persons. Choi and Piazza (2014) assert that countries with large internally displaced populations are more likely to experience a higher rate of suicide terrorism. Vargas-Silva (2016) conducts a literature review on remittances sent to and from refugees and IDPs. There is however some work that focuses partly on the location of forcibly displaced population: Arias et al. (2014) provide evidence of the desire of internally displaced persons in Colombia to return to their land rather than staying in the reception areas. In a cross-county study using country aggregated data on both refugees and internally displaced persons, Melander and Öberg (2007) find that the threat perceived by potential forced migrants is more related to where the fighting is taking place than to the overall intensity of the fighting. Davenport et al. (2003) argue that potential migrants use observable information in their environment to assess the threat to their person and to form and revise beliefs about their personal security and that their decisions to stay or leave are based on these beliefs. In contrast, Balcells et al. (2012) claim that displacement is not simply a byproduct of violence: after a set of violent events, people flee. They argue that as refugees and IDPs are also political actors, their identities are crucial to their displacement. The authors show that in both Colombian and Spanish civil wars, higher levels of displacement are observed in locations where more sympathizers of rival armed groups reside.

Theory Section

Displacement Location

Some civilians travel long distances, sometimes crossing national borders and potentially becoming refugees, while others remain close to conflict. Some civilians relocate to international aid organizations camps, while others choose to stay with family and friends or make their own arrangements. Some forced displacement is temporary, but sometimes it is not. In this paper I am going to focus on internal displacement – within national borders – for two reasons. First, it is the most common type of displacement. In June 2024, UNHCR estimates that 68.3 million people are internally displaced worldwide, compared to 31.6 million refugees. Second, obtaining data on the location of those forcibly displaced is extremely challenging, but in an unprecedented opportunity, precise geo-location of internally displaced people in Iraq was available during the 2014-2017 war.

The most plausible explanation behind the lack of academic papers concerned with the location of those forcibly displaced might simply be the lack of data. While thanks to the efforts

of Armed Conflict Location & Events Data (ACLED), and Uppsala Conflict Data Program (UCDP) conflict data with precise geographical coordinates is now publicly available, data on the location of forcibly displaced people is hard to obtain. Data on forced displacement location is generally not available due to lack of resources to collect displacement data when it happens, but also because of security concerns as civilian targeting is a commonly used tactic of war and forced displacement might be the intended outcome behind it. Civilian targeting might be "discriminate" or "indiscriminate" if perpetrators purposely target non-military individuals based on their expected individual support based on ethnicity, religion, or political affiliation (Lupu and Wallace, 2023). While "discriminate" civilian targeting might aim to displace people who do not align with the perpetrators' identity, "indiscriminate" civilian targeting might also pursue the displacement of people as a way to pressure the incumbent political power. Additionally, the lack of forced displacement *location* data might explain why literature concerned with forcibly displaced persons themselves does not generally focus on forecasting the displacement location to better target aid.

Academic research on displacement is often not useful to policy makers during conflicts. Policy recommendations usually come from case studies and literature by international aid organizations, governments, and NGOs. The lack of displacement forecasting tools has led to preventable humanitarian disasters. For example, in 2003, aid organizations estimated over a million Iraqis would be displaced by conflict, but few sought shelter in camps, leading to their dismantling. In 2006, unexpected mass displacement caused a humanitarian crisis. This event does not mean that there were no internally displaced people between 2003 and 2006, but that much like in 2014, people chose not to go to camps and one of the reasons might have been that they were too far from their location of origin. This leads to my first hypothesis:

H1: *Displacement location hypothesis*: Internally displaced people remain close to their usual place of residence where conflict levels are high.

I acknowledge that important counter-arguments suggesting that the particular identity of internally displaced people and that of the different involved military factions will be crucial in many conflicts (Balcells et al., 2012) as opposed to conflict intensity (Davenport et al., 2003; Melander and Öberg, 2007). It is possible that only specific people are forced out of their homes, or at least, will be much more likely to fear prosecution and therefore leave. In the case of Iraq for example, we might be inclined to think that the Yazidi minority population escaping to Mount Sinjar might have had stronger incentives to flee than Sunni Muslims who might not have been as viciously targeted by ISIS. ISIS is after all an extremist Sunni jihadist group, with a particularly violent ideology that calls itself a caliphate and claims religious authority over all Muslims (RAND, 2024). The way Sunni Muslims perceived the danger coming from ISIS

might be different to the rest of the Iraqi population (Cockburn, 2015). Indiscrimiate violence however, might lead individuals to believe that they could be targeted regardless of their identity (Lupu and Wallace, 2023), and we observe that Sunni Muslims were also displaced. While this first hypothesis will not address faction or civilian identity directly, the second hypothesis will provide some empirical evidence regarding ethno-religious identity at the displacement location. Nonetheless, being able to predict where people will be displaced has important policy value even if not everybody is fleeing.

Displacement location characteristics

Thus far, I have focused on displacement location relative to conflict. However, at a similar distance from the origin of displacement, characteristics particular to each location are likely a factor when people fleeing conflict choose where to relocate. In some conflicts, hosting location characteristics might even be fundamental to understand and predict displacement patterns. These location characteristics might be the presence of a social network for people about to become displaced, or the new commute length to their usual occupation, the location's political identity or ethnic composition, its safety reputation, the presence of camps for conflicts were camps are utilized, or the rental housing availability among others.

Host location characteristics will also determine the quality of life of those displaced especially if displacement is prolonged. In fact, both refugees and internally displaced persons often face marginalization as a consequence of violation or lack of protection of human rights, including economic, social and cultural rights. International attention typically fades after the initial emergency phase, and the long term conditions of those displaced tend to worsen. Humanitarian assistance and the generosity of host communities are often overstretched, which is understandable as policy frameworks and institutional arrangements were initially designed for short-term solutions during the emergency phase, but are prolonged due to a lack of resources or better options (Christensen et al., 2009).

Displacement location urban characteristics

Often times displaced people relocate relatively close to conflict, potentially driven by the fact that many try to retain their usual occupation. Rather than relocating to rural areas, those forcibly displaced choose locations with high population density and good access to roads so they can continue with their lives while conflict is still unfolding. This may be because individuals can continue with their usual job by commuting to their usual work site, or because they can perform

their trade in an urban setting nearby. As a complement to the first hypothesis and together with the third that follows, the second theory aims to further understand the characteristics of the locations chosen by those displaced. I focus on two physical considerations which are now typically available for any geographical area around the world and therefore testable hypotheses: road networks, and population density which are usually or in part derived from satellite imagery and do not depend on the availability of national datasets[3]. This leads to my second hypothesis:

H2: *Displacement location physical characteristics hypothesis.*: Displaced people remain in highly populated areas close to a main road.

I acknowledge that important counter-arguments similar to those in H1 suggesting that individual political, ethnic, and religious identity constitute fundamental factors, are crucial in many conflicts. The driving mechanism behind this hypothesis is that people will try to continue with their usual socio-economic activity whenever possible. By staying close to their usual place of residence despite conflict (H1), and in a highly populated, well connected area, they are more likely to continue with their carer responsibilities and usual occupation — albeit a longer commute, or changing the neighbourhood where they conduct their trade.

Displacement location ethno-religious characteristics

This hypothesis seeks to address part of the limitations flagged for the first two hypotheses: political and ethno-religious identity also matters. The literature identifies two main forms of violence, discriminate and indiscriminate civilian targeting. In the case of the former, civilian and combatant identity constitutes a fundamental driver for displacement. With this second hypothesis, I am unveiling a third factor surrounding identity: the displacement location's political and ethno-religious identity. When a person is forcibly displaced due to perceived danger — which is a function among other things of their own identity but also the identity of the combatants posing the threat — the identity of the displacement location will also be important when deciding where to relocate. In particular, it could be the case that locations without a clear ethno-religious majority could be perceived as more welcoming, or safer as newcomers will be harder to single out if the conflict were to spread to those locations. This leads to the third hypothesis:

H3: *Safety in diversity hypothesis*: Diverse areas are more likely to host more displace people relative to areas with a clear ethno-religious majority.

I acknowledge that not observing displaced people's ethno-religious and political identity in the data remains a limitation of this paper. I test a third pillar of identity through location ethno-religious composition which in the case of Iraq is available based on past election voting

results. I do not think that displaced people's identity is not important as it was crucial in a few of the events and areas of the conflict used in this paper to test the hypotheses I proposed. In the 2014 war against the Islamic State in Iraq, while most people fled to highly populated and well connected areas, certain minorities escaped to rural areas. An insightful testimony of the Iraqi war against ISIL from the perspective of the Yazidi minority can be found in the book "My story of captivity, and my fight against the Islamic State" by Nobel Peace Prize winner Nadia Murad. Many of the Yazidis prosecuted by ISIL escaped to Mount Sinjar, a remote and rural area in the North of Iraq with religious significance to them. The Yazidi exodus might highlight the different displacement patterns between those who suffered discriminate attacks based on their ethnicity (Yazidi), as opposed to the Sunni population which was majority in most of the ISIL controlled territory north of Baghdad who might have suffered indiscriminate violence.

Research Design

To test these hypotheses, this paper uses publicly available geo-referenced data with a large spacetime variation showing on what day and on what geographic coordinates, displaced people and conflict occurred. Conflict data is drawn from the Uppsala Conflict Data Program (UCDP) georeferenced event dataset, and internally displaced persons (IDP) data is drawn from the International Organization for Migration (IOM) displacement tracking matrix unit in Iraq that tracked people specifically displaced by the conflict against the Islamic State of Iraq and the Levant (ISIL). Additionally, I use geographical information such as population density from Oak Ridge National Laboratory's LandScan and road networks from the United States National Imagery Mapping Agency (NIMA) made publicly available by the UN Office for the Coordination of Humanitarian Affairs (OCHA) country office in Iraq. Please see section 1 in the Annex for a detailed description of conflict and displacement data together with some descriptive statistics.

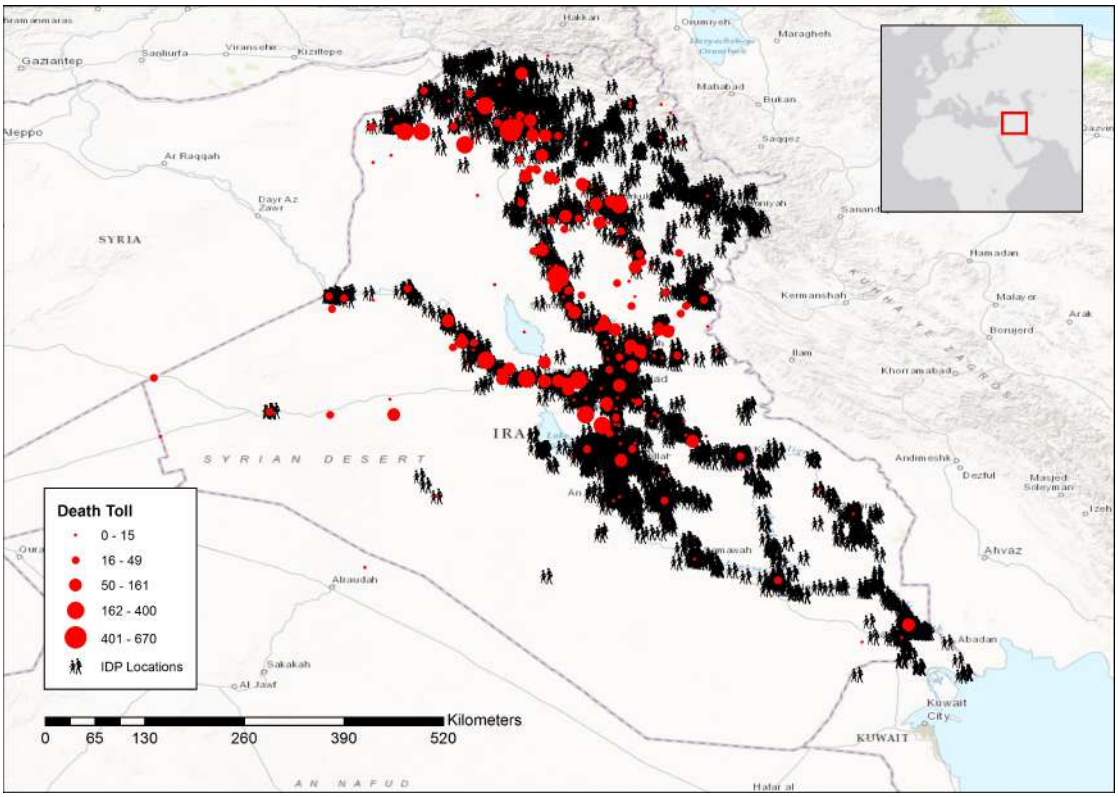
Data Advantages: Both conflict and IDP data offer great geolocation precision: point coordinates. This precision allows us to join other spatial data and allows the use of models that can accommodate the richness this information provides. Both conflict and IDP data have a lot of variation geographically and over time which allows me to disentangle which IDPs are caused by which conflict. I test the causality of this correlation measuring out-of-sample prediction errors.

Data Limitations: The main data limitation is that we do not know the location of origin for IDPs only where they are displaced. This feature poses a very important challenge to the estimation which is tackled using the great geographical and temporal variation of the data. Additionally, I do not observe individual's ethno-religious identity, only the identity of districts based on past election results.

Sample - 2014 Iraqi war against the Islamic State

The Arab Spring arrived to Iraq with the 2011 anti-government protests. In 2014, the Iraqi insurgency escalated into civil war when the Islamic State of Iraq and the Levant (ISIL) seized control over major cities in Iraq such as Tikrit, Fallujah and Mosul, creating hundreds of thousands of internally displaced persons (IDPs). The Iraqi war against ISIL was an armed conflict in which Iraqi armed forces, Kurdish Peshmergas, Turkmen Muslims, Assyrian Christians, Yazidis, Shabakis and Armenian Christians fought against the Islamic State of Iraq and the Levant[4].

Figure 1: Conflicts and IDPs - Iraq overview (2010-2015)



We can summarize the characteristics of the Iraqi Civil war between 2014 and 2017 with the following facts: Conflict and IDPs are spread geographically and over time. There have been periods of time with over 3 million IDPs. On average, 58% of those forcibly displaced are underage; 58% Sunni and 29% Shia. Conflict is much more intense from Baghdad to the north and this can explain that most displaced persons are Sunni, regardless of how religious identity shaped displacement. See Figures 1, 2 and 3 for a geographical overview of the data used in this paper.

The Iraqi population is clustered around the fertile land provided by the rivers Tigris and Euphrates, and the mountainous region of the Iraqi Kurdistan at the north of the country. The vast Arabian Desert in the south is largely unurbanized and unpopulated. In Iraq, 95-98% of the

population is Muslim. The two most important Islam branches in Iraq are Sunni (29-34%) and Shia (64-69%) (C.I.A. Iraq, The World Factbook). The Shia population is generally established on the southeastern half of the country, closer to Iran where Shia constitute 90-95% of the population. Sunni Muslims are established in the west half of Iraq, closer to Syria, Jordan, and Saudi Arabia where Sunni muslims are a religious majority. Turkmen Muslims, Assyrian Christians, Yazidis, Armenian Christians, can be found in the northern parts of Iraq near the border to Northern Syria and Turkey. Kurds are located in the Kurdistan autonomous region[5].

To the best of my knowledge, the only academic research studying IDPs in Iraq has been conducted by the social anthropologist Dawn Chatty and coauthors (Chatty, 2003; Chatty and Mansour, 2011b,a; Marfleet and Chatty, 2009). Additionally, the report for the US congress done by Margesson et al. (2008) analyses Iraqi refugees and IDPs during the 2007 crisis, including the conditions for those displaced in Iraq and the refugee situations in Syria, Jordan and elsewhere.

Figure 2: IDP Locations (April 2014)

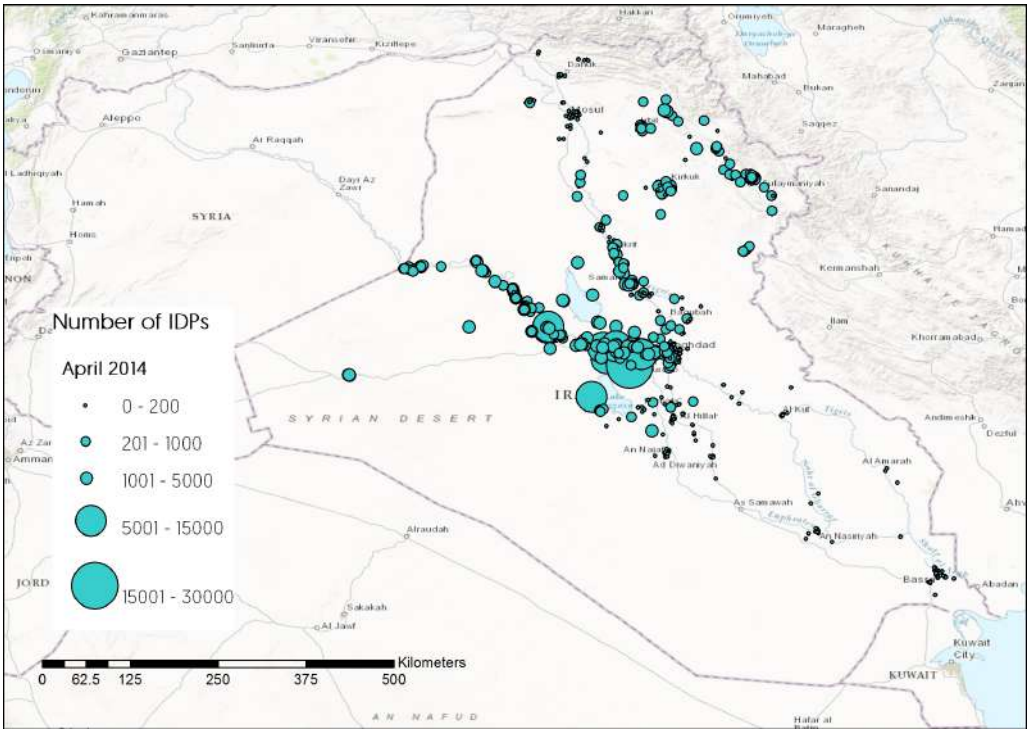
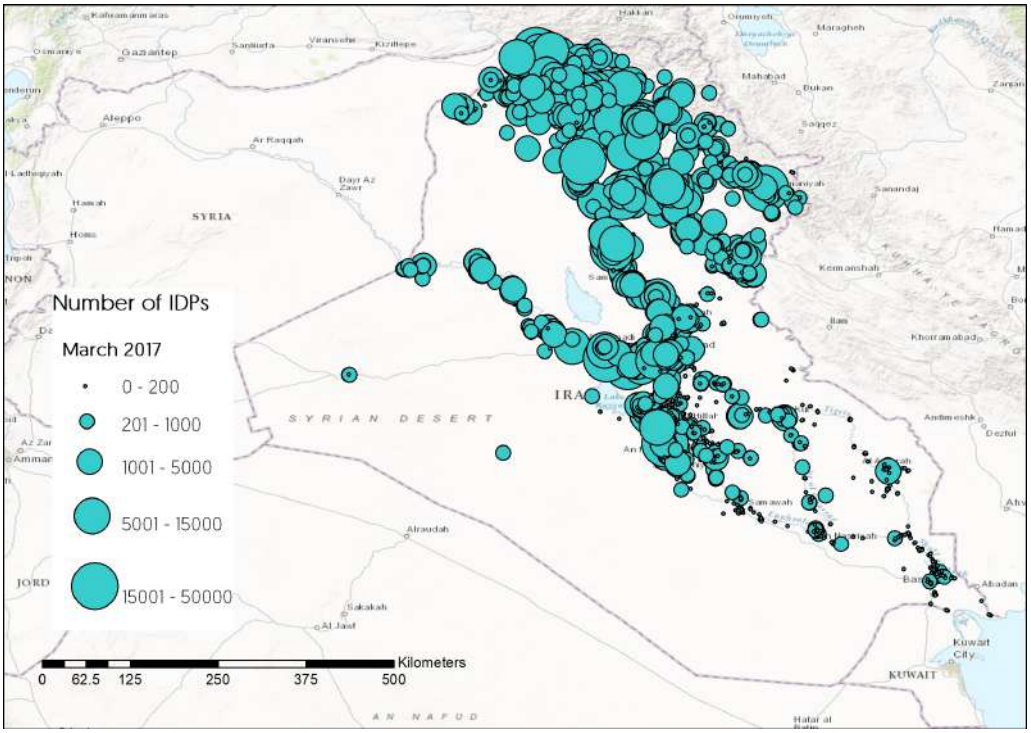


Figure 3: IDP Locations (March 2017)

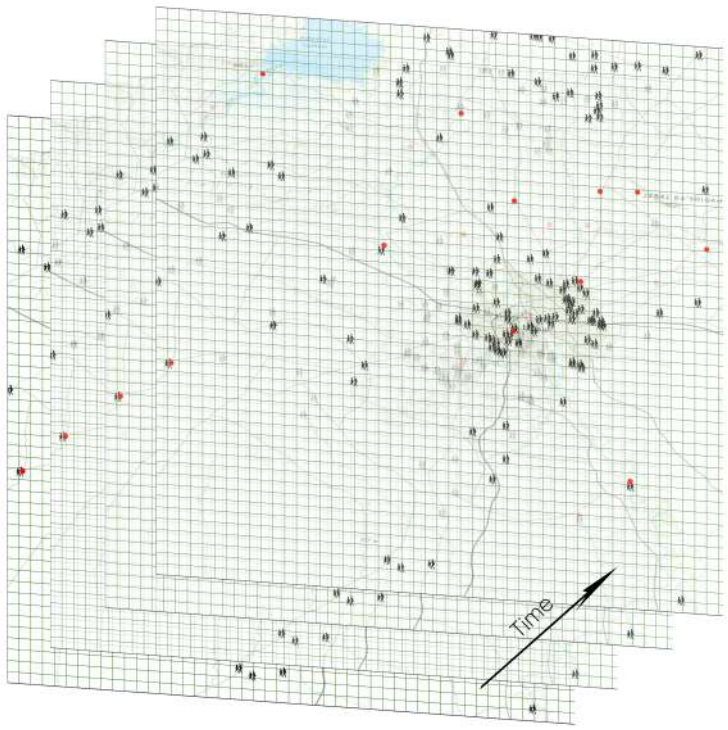


Methods

Data structure: Panel from grid

The first step is to construct a grid or lattice where we can organize the data with geolocation coming from different sources[6]. I take the populated surface over Iraq constructed from a high-resolution population data (LandScan 2011) and divide it into a grid of equally sized 4 mi^2 cells where each of the grid cells is linked to the rest of datasets using geographical coordinates[7]. Buhaug and Rød (2006) take a similar approach over Africa using a 100x100km grid size. The resulting dataset has a cross-section of 22,747 units (4 mi^2 cells, 2 miles by 2 miles). To construct the time dimension of our panel data we need to consider the event frequency in both IDP and conflict data. The data on conflict event is disaggregated at the day level for most recorded events in UCDP GED. Conflict data has 1,468 different time periods, while IDP location data is collected in a biweekly interval and has 61 available time periods. In our panel, I regularize the time dimension by constructing 67 different time intervals which correspond to the 1^{st} and 15^{th} of each month from April 2014 until December 2016. For the geographical merge, observations are assigned to the nearest panel unit. When we group all the observations in the 4 mi^2 grid cells, the 307 conflict locations boil down to 266 grid cells and the 4,634 IDP locations to 1,895 grid cells. For the temporal merge, both conflict events and IDP data figures are aggregated and counted to the closest future regularized time period. There are more regularized intervals than IDP waves because there are two missing IDP waves in the data. The resulting dataset is a panel with 22,747 units over 67 biweekly time periods where each unit is a 4 mi^2 grid cell ($N = 1,524,049$). Figure 4 illustrates the object used for the estimation.

Figure 4: Panel data structure illustration



The grid includes locations where our outcome of interest and regressors are always zero, where no conflicts nor displacement occur. These inactive units are included because we would like to let the data tell us why these locations were not chosen or chosen later by IDPs while other locations with the same characteristics from the econometrician's perspective without imposing a rigid structure in the model[8].

Having geocoded data this precise allows me to choose the unit of analysis, in this case, through the grid cell size. And while using a smaller unit fully exploits the data's precision, it will make spillovers between units more likely. Econometric i.i.d. models rely on the assumption that most – if not all – spillover effects are contained within the unit of analysis not to violate the independence assumption. The smaller the unit of analysis, the harder it should be to justify that the effects we are studying are contained within; using a small unit of analysis, we expect conflict to have an impact on the number of displaced people in units other than its own. I chose the smallest grid size which would make use of the precision of the geographical coordinates available in the data. Note that if IDPs do not travel great distances, a larger unit of analysis such as bigger administrative boundaries could be appropriate and not require to model spillovers. In this application however, larger levels of aggregation do not properly capture displacement movements as many IDPs travel across political boundaries even if they are only displaced for

less than 40 miles. Moreover, the prediction power of these models is much worse than the model I use – especially for out-of-sample prediction. See Results for a detailed discussion.

Statistical Model

Now that I have a panel, I use a model that allows conflict in any given grid cell to potentially have an impact on displacement in any other cell. The most ambitious and agnostic model we could run would be a model that estimates the correlation between the characteristics of any given pair of units in our panel. This is impossible as the number of observations is much larger than the time dimension ($N >> T$). In the networks literature it is standard to make assumptions to reduce the dimensionality of the estimation problem to make it feasible[9]. For example, when Currarini et al. (2010) estimate their network, they impose all spillover effects to be zero expect those for which the observations belong to the same ethnic group. Other papers such as Conley and Udry (2010), Ades and Chua (1997), or Case et al. (1993) use the notion of "neighbour" so that spillovers only happen between neighbouring farmers in Ghana, neighbouring countries, or neighbouring states in the US. In this paper, we are interested in the relationships between any two units in our panel and we would like to know which units affect which and how important the links between these units are. In contrast with the part of the networks literature which takes the links of the network as given to then estimate its intensity, we would like to be agnostic and estimate which links are relevant and therefore understand the range of influence any given unit has over all others. In contrast with the spatial econometrics literature, we do not want to impose a specific structure based on measures of distance between units to calculate the relationship of the weighted average of the characteristics of the rest of units. The latter would also impose that the importance of conflict diminishes over distance or time, which is a result of this paper, but not an assumption of the model.

Consider a linear model with a single regressor and time period to illustrate the point above:

$$\mathbf{Y}_t = \mathbb{N}_t \mathbf{X}_t + \epsilon_t \iff y_{i,t} = \alpha_1 x_{1,t} + \alpha_2 x_{2,t} + ... + \alpha_i x_{i,t} + ... + \epsilon_{i,t} \tag{1}$$

Where $\mathbf{Y}_t$ is a $n \times 1$ column vector in which the i^{th} element indicates the number of IDPs in the i^th grid cell of the panel at time t. $\mathbf{X}_t$ is an $n \times 1$ vector denoting the panel unit's i characteristic. The $n \times n$ network matrices $\mathbb{N}_t$ are constituted by the parameters of interest. The diagonal of $\mathbb{N}_t$ captures the contribution of conflict on a grid cell to the number of IDPs on that grid cell, while the off-diagonal parameters are the contribution of all other units in the panel. If we develop that matrix multiplication we obtain the expression at the individual level that follows, where α_k are the elements of $\mathbb{N}_t$ for every pair of units in the panel.

Allowing any given grid cell to potentially have an impact on displacement in any other cell, means allowing for the off-diagonal elements of $\mathbb{N}_t$ to be different than zero. I also do not want to impose a rigid structure by aggregating all elements in $\mathbb{N}_t$ and estimate a single parameter as I do not want to assume that I know that conflict near or far has more or less of an impact on displacement; I would rather let that be a result of the model instead of an assumption. However, to make this model feasible to estimate, I need to impose some restrictions to reduce the number of parameters.

Indifference Rings

To reduce the dimensionality of the problem, I assume there exist indifference rings around each location and events within the same indifference ring have the same impact on the outcomes of the origin. One can think of concentric rings and all conflict events within each of these rings will have the same influence over the center, irrespective of what direction relative to the center they are located. This way, the number of unique parameters to estimate for each regressor is reduced from n^2 to the number of indifference rings - from 517,426,009 parameters per regressor, per time period, to as many parameters per regressor per time period as indifference rings defined (25 in this paper).

I define a set of distances so that entire map of Iraq is covered and where the indifference ring widths to broaden as we move away from the unit of origin: I aggregate regressors by rings with a 2 mile width from 0 to 10 miles away from the origin, then 5 miles wide from 15 to 30 miles from the origin, a width of 10 from 40 to 100 miles from the origin, a width of 50 miles from 150 to 300 miles from the origin, and rings of 100 miles width from 400 to 700 miles from the origin. In miles, the set of distances in a pseudo-code notation is 2:2:10, 15:5:30, 40:10:100, 150:50:300, and 400:100:700. Note that the maximum distance between any two units of the panel is 699.3557 miles. I group the pair-distances of the N panel units to the threshold closest to the actual geographical distance.

This assumption would be problematic if people in Iraq are fleeing in a specific direction escaping from ISIL. For example, if people were escaping conflict always to the east as the battle front approached from the west (Syria). Data patterns however do not indicate that internally displaced persons are fleeing from conflict in a specific direction as if escaping from an advancing war front. First, we find that a substantial proportion internally displaced persons are within ISIL-controlled areas. Second, even in instances where civilians are fleeing from ISIL, they are not necessarily moving west. The locations most active in conflicts are Mosul and Baghdad and IDP sites surround both in every direction. Moreover, IDPs near Sinjar are also a good example of how this one-directional escape route is not the case. In Sinjar, we see many IDP sites form

in Mount Sinjar west of the city of Sinjar where the conflict takes place and deeper into ISIL controlled territory.

In terms of the equation presented above, and using the language of networks and spatial econometrics, I am constructing network model that treats the structure of interactions as unobserved by allowing all off-diagonal parameters to be different than zero, and accommodates data with specific geographical coordinates. I impose parameter restrictions on the network matrix ($\mathbb{N}$) as a function of the spatial weight matrix ($\mathbb{D}$)[10]. The contribution of conflict events to the number of displaced people at a given location is the same if conflict events happen at a similar distance[11]. These restrictions still allow for any unit in the panel to have spillover effects on any other unit. The number of unique parameters to estimate for each regressor is reduced from n^2 to the number of indifference rings. The method reduces the dimensionality of the parameter space dramatically and can be applied to any context in which the network structure is the object of interest, and a specific interpretation of the parameters is needed.

Imposing these parameter restrictions allows us to group our regressors under indifference rings and we can estimate the parameters of interest using these indifference rings. We can rearrange equation (1) using the parameters of interest for each indifference ring as a common factor to end up with the following model:

$$y_{i,t} = \beta_t^{d_0} x_{i,t}^{d_0} + \beta_t^{d_1} x_{i,t}^{d_1} + \beta_t^{d_2} x_{i,t}^{d_2} + ... + \epsilon_{i,t} \tag{2}$$

Where the $y_{i,t}$ is the number of IDPs in location i at time t, x_t^d is the sum of all conflict events at the indifference ring defined by distance d with respect to i at time t. The parameters of interest are β^d which are now associated to indifference rings rather than the panel units. Once this model is estimated, we can recover the entire network matrix by mapping these parameters back to their location in the matrix $\mathbb{N}$ based on the indifference rings we imposed.

Model Equations

Consider the generalization of the model presented where I now include t different time periods, and with T lags in a non-linear model:

$$\mathbf{Y}_t = f\Big(\mathbb{N}_t \mathbf{X}_t + \mathbb{N}_{t-1} \mathbf{X}_{t-1} + \mathbb{N}_{t-2} \mathbf{X}_{t-2} + ... + \gamma \mathbf{Z}\Big) + \epsilon_t \tag{3}$$

Where $\mathbf{Y}_t$ is a $n \times 1$ column vector in which the i^{th} element indicates the number of IDPs in the $i^t h$ grid cell of the panel at time t. $\mathbf{X}_t$ is an $n \times 1$ vector denoting the panel unit's number of conflict events at time t. The $n \times n$ network matrices $\mathbb{N}$ are constituted by the parameters

of interest. f is the Tobit/Rectifier function, $f(x) = \max(0, x)$. $\mathbf{Z}$ are characteristics of the location that do not change over time such as geographical controls and ethnic composition as it is observed in the data in a single period. Note that we are estimating a different matrix for each time lag and we are not imposing time stationarity or restricting the relationship between $\mathbb{N}_{t_1}$ and $\mathbb{N}_{t_2}$ in any way for any t_1 and t_2.

Once I apply the parameter restrictions, the model equation can be rearranged as follows:

$$y_{i,t} = f\Big(\sum_{d}\Big(\beta_t^d x_{i,t}^d + \beta_{t-1}^d x_{i,t-1}^d + \beta_{t-2}^d x_{i,t-2}^d + ...\Big) + \gamma z_i\Big) + \epsilon_{i,t} \tag{4}$$

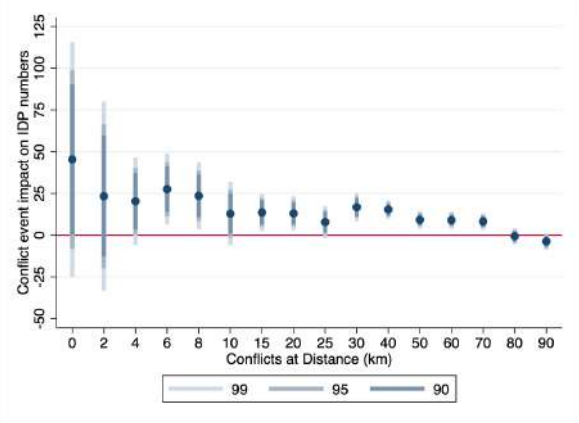
Where $y_{i,t}$ is the number of IDPs in location i at time t. The parameters of interest are β_t^d which capture the contribution of an additional conflict event that happens within the indifference ring at distance d and time period t to the number of displaced people at the origin cell i. I will test the hypotheses proposed interpreting the estimated β_t^d and γ. The main results will consist of the estimated β_t^d for distances d_0, d_1, d_2 at times t, $t-1$, $t-2$... (Hypothesis 1). For example, if $\beta_{-1}^4 = 57$, that would mean that an additional conflict event between 4 and 6 miles from origin, 1 month prior, is associated with an increase of 57 additional displaced people at origin. The model results goes beyond the original hypothesis 1 in the sense that it also provides results with respect to the time dimension. I can test that displacement occurs close to conflict, regardless of how far in time the conflict that triggered said displacement happened. I estimate a total of 250 parameters associated to the number of conflict events for the pairs of 25 distances and 10 time lags. γ captures the influence of the observable time-invariant characteristics z on the number of displaced people in its own location (Hypotheses 2 and 3). f is a Tobit function, to capture that conflict might not immediately trigger displacement, but rather after a few conflict events accumulate, people are displaced — the results are robust to linear models.

Results

Displacement location

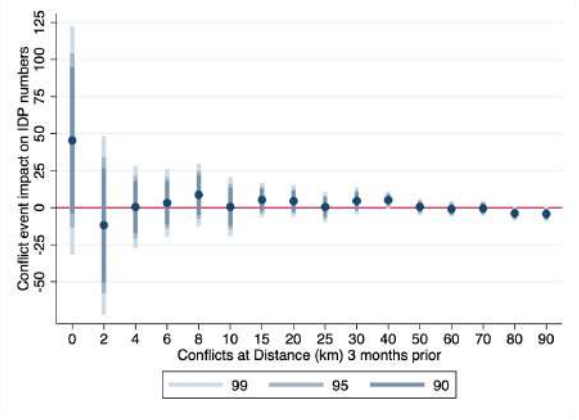
I test hypothesis 1 which proposed that internally displaced people remain close to their usual place of residence where conflict levels are high. I interpret the parameters associated to additional conflict events for each ring at distance d relative to displacement. I find the highest concentration of displaced people within two miles of conflict, decreasing with distance , and getting close to zero beyond eighty miles. Conflict further than 80 miles is not a determinant of new internal displacement. Figure 5 shows the estimation results for conflict intensity measured as the number of conflict events at a given location.

Figure 5: Baseline Model Estimation Results



The model results include also conflict at not only different distances but also different points in time. Figure 6 shows the coefficients associated to conflicts at the same set of distances but 3 months prior to displacement readings. I find that older conflict does not have a significant impact on displacement.

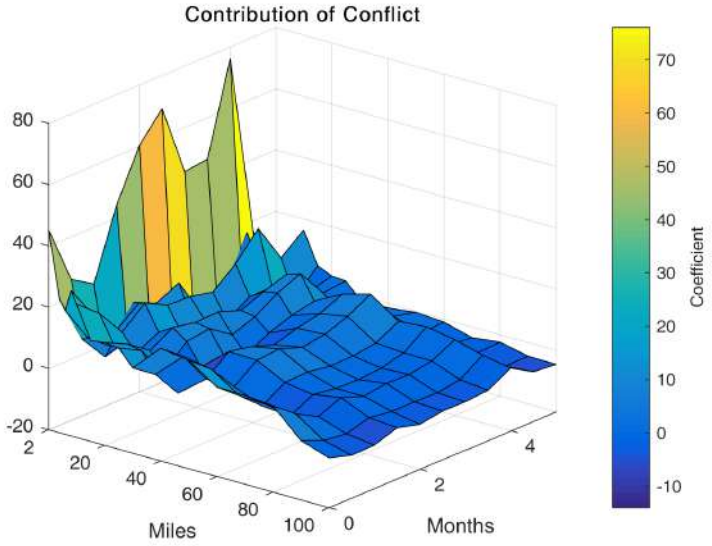
Figure 6: Baseline Model Estimation Results



Finally, I plot all coefficients for distance d and time t on Figure 7. The axis labeled "Months" displays the estimation results by the number of months between the conflict events and the displacement of people. The axis labeled "Miles" shows how the relationship between conflict intensity and the number of displaced people number changes at different distances. Figure 7 presents a subset of all estimated parameters for conflict intensity including only conflict events not older than 5 months and not further away than 90 miles, but please note that parameters for conflict further or older are not statistically different than zero. The results presented here result from a model where the independent variable is the stock of internally displaced people at a given locations at time t. These results are robust to using different model specifications such as a linear model with fixed effects, or network models using flows, or lagged stock of internally

displaced people. Please see Robustness Checks section in the Annex.

Figure 7: Baseline Model Estimation Results



Location physical characteristics and ethno-religious identity

I test hypotheses 2 and 3 which stated that displaced people remain in highly populated areas close to a main road, and that diverse areas host more displaced people relative to areas with a clear ethno-religious majority. I find that people choose locations in highly populated areas within five miles of a main road. The effect is stronger within two miles of a main road. Areas with a clear ethno-religious majority be it Sunni, Shia, a mixed of Sunni and Shia, or Turcoman host fewer internally displaced persons than areas without a clear ethno-religious majority. The effect is stronger in areas where Sunni are the majority. This effect is not found in areas with a Christian majority where the coefficient is positive but not statistically different than zero.

Table 1 presents the parameters associated to individual grid cell characteristics from the model. It shows that IDPs tend to cluster in highly populated areas within 5 miles of a main road, and to what extent the ethno-religious identity of receiving locations is a determinant for their likelihood to host IDPs relative to locations with no clear majority. The parameters interpretation is the following: An additional person at a location is associated with 28 additional internally displaced people hosted. Locations within two miles of a main road hosted on average 633 more internally displaced persons than locations further than five miles. Locations between two and five miles from a main road hosted on average 424 more internally displaced persons than locations further than five miles.

Table 1: Baseline model hosting locations' characteristics

	Coeff.	Std. Error	t	$P > \|t\|$
Ethno-religious identity:				
Sunni	-616.25	6.92	-89.1	0.000
Shia	-252.45	6.68	-37.77	0.000
Mixed Sunni Shia	-265.69	13.61	-19.52	0.000
Christian	63.20	35.36	1.79	0.074
Turcoman	-517.22	0.03	-23.45	0.000
Physical characteristics:				
Population	28.47	0.24	118.62	0
Road within 2mi	633.14	6.39	99.06	0
Road within 5mi	424.67	6.25	67.92	0
Conflict regressors	✓			
Observations	1,296,579			

Notes: Conflict regressors are omitted from this table. Ethno-religious majority at origin considered when at least 66% of the population voted for a clearly Sunni, Shia, or Kurd party in the 2011 census.

Locations with a clear Sunni majority host on average 616.25 fewer internally displaced people (IDP) than locations without a clear ethno-religious majority. Locations with a clear Shia majority host 252.45 fewer IDPs, locations with an even mix of Sunni-Shia host 265.69 fewer IDPs, locations with a clear Christian majority host 63.20 more IDPs (not stastistically significant at the 5% significance level), locations with a clear Turcoman majority host 517.22 fewer internally displaced people than locations without a clear ethno-religious majority.

The ethno-religious data comes from Berman et al. (2011) which use governorate-level returns in the December 2005 election. Diverse locations where the population did not vote for a clearly ethnic-specific party host more IDPs than locations with a Sunni, Shia, Mixed Sunni/Shia, or Turcoman majority party. Locations where at least 66% of the population in a governorate voted for a clearly Sunni party in the 2005 elections tend to host fewer IDPs than locations where Shia or Kurd political parties were elected. The parameters can be interpreted in the following way: Locations where a clearly Sunni party was elected in the corresponding governorate with at least 66% of the votes have on average 616.25 fewer displaced people than areas with where no elected party had two thirds of the total votes, or the elected party was not clearly Sunni, Shia, or Kurd. While statistically, the baseline corresponds to locations where there was no clear majority, most regions show elected parties with a clear majority. One should also take into account that most displacement happened north of Baghdad, and Shia populations are located to its south. For this reason, One could consider the baseline scenario to be given by the Shia or the mixed Sunni-Shia parameter at -252 and -265, respectively. The interpretation of the parameters would therefore be that Sunni areas tend to host around 360 displaced people fewer than more diverse areas;

Turcoman locations tend to host around 260 fewer displaced people, and Christian areas tend to host around 300 more internally displaced persons than areas with a Sunni majority. Iraq's ethno-religious geographical distribution based on the 2005 elections can be found in the Annex, Figure 16.

Prediction errors compared to other models

In this section, I show that the network model I use to test my hypothesis performs better than models using administrative boundaries. The network models estimated outperform specifications assuming independent units of observation, especially when we consider prediction errors using out-of-sample tests. Table 2 shows the distribution of the prediction error for several specifications. The first two columns correspond to models where the unit of analysis are the district and governorates in Iraq, and the last two columns correspond to the network models used in this paper.

Table 2: Prediction Error Distributions

	Assuming units are independent		Network Models	
	District	Governorate	4sqm-Linear	4sqm-Non-Linear
In-sample:				
Mean	0.00	0.00	0.00	-8.54
Std. Dev	8,232.89	28,152.38	104.99	235.20
95% Perc.	12,440.92	49,108.27	14.94	-1.00
Out-of-sample:				
Mean	686.50	3,823.30	-4.42	-18.47
Std. Dev	8,386.79	26,529.18	220.67	226.45
95% Perc.	17,510	49,497.91	4.88	-1.42
Observations	6,834	1,206	1,296,579	1,296,579
Fixed effects	Yes	Yes	Yes	No

When choosing the model to explain displacement as a function of conflict, the question is broader than what specific equation and what parameter restrictions should be made: Precise geographic data also allows us to choose the unit of analysis. Econometric models using administrative areas as the unit of analysis, rely on the assumption that observations are independent and identically distributed, and that spill-over effects are contained within these larger units of analysis, as otherwise, the existence of spill-over effects will violate the independent assumption (or at least that the spill-over effects are small enough to be considered zero). While choosing a smaller unit of analysis allows us to be more precise, it is not necessarily better. I have estimated that IDPs move close to their original position, which in retrospect could indicate that using dis-

trict or governorate as the unit of analysis is suitable. However, as there are so many spill overs, IDPs moving only 40 miles are already crossing a district boundary, these aggregated models are unable to capture such dynamics and their prediction power is much smaller than that coming from network models with a smaller unit of analysis, that allows for cross-unit spill-over effects. See Table 2.

Conclusions

Forced displacement is one of the most important consequences of conflict, yet the understanding to adequately support those fleeing violence could be improved. While academic research exists, it generally provides a wealth of theory and empirical insights that are not always useful to design policy and aid. There are first order simple questions that demand an answer. Where do people go when they are forcibly displaced by conflict?

In this paper, I focus on the impact of violent conflict on internally displaced persons in Iraq considering distance and time. Namely, how far away the conflict takes place and how long ago the conflict happened. I study all internal displacement within the country as a result of the war against ISIL from April 2014 until January 2017. I exploit the vast spacetime variation of both conflict and forced displacement data to map how conflict intensity overtime is correlated with the apparition of people in surrounding locations. To test my hypothesis, I use an econometric model where I impose the least possible restrictions to let the data show which conflict is triggering what displacement without imposing that conflict further away has more or less impact on displacement. I contribute to fill the gap in the conflict literature regarding internally displaced persons by answering the fundamental questions of where and when those displaced go as a function of conflict considering space and time. This paper is the first empirical paper studying forced movements as a response to conflict using geo-referenced data, and focuses on prediction using precise geo-referenced data rather than exploring the exact mechanism behind people's decision to flee. In the context of forced displacement, prediction itself has very high policy value to anticipate and respond to the crises internally displaced persons and refugees face. The findings offer insights which can guide future humanitarian policy on where and when aid is more effective.

The first key finding is that the highest concentration of displaced people happens within two miles of conflict, decreasing with distance and time, and disappearing beyond seventy miles. The findings show that aid to those forcibly displaced needs to be delivered in locations with high conflict intensity which is by itself a challenge. This challenge also affects the form of the aid that can provided to those forcibly displaced. While the provision of camps still constitutes

the most common form of aid, survey data shows that only three percent of internally displaced people seek shelter in a camp as most rent their own flats or stay with family or friends in accommodation in close proximity to conflict. Camps are generally located in areas further to conflict than where internally displaced person choose to relocate. This paper however is not able to distinguish whether camps are underutilized because of their location, or a preference for the type of shelter/accommodation they provide.

The second key result is that people choose locations in highly populated areas within five miles of a main road. The reason behind this choice, together with how close to their location of origin displaced people remain, might be the nature of their preferred shelter, and that people maintain their usual source of income. According to survey data collected in 2016, 43 percent of internally displaced were in rented housing, 30 percent with host families, 11 percent in abandoned buildings. Only 3 percent of Iraqi internally displaced people used camps to shelter (See Table 6 in the Annex). The most common reported sources of income during displacement were: the usual paid job (public) 31 percent, paid job (private) 22 percent, and informal commerce or inconsistent daily labor at 23 percent. Less than two percent of internally displaced people reported aid from organizations as their main source of income (See Table 7 in the Annex).

The third key result is that areas with more diversity host more displaced people relative to areas with a clear ethno-religious majority. With respect to the literature, while authors such as Davenport et al. (2003) argue that conflict is the main reason for displacement, Balcells et al. (2012) argue that the identity of those displaced is the fundamental factor. This discrepancy suggest that the religious, ethnic or political identity of the forcibly displaced might be important depending on the event. In the Iraqi war against ISIL, survey data collected in 2016 indicates that 'generalized violence and armed conflict' was reported as the main reason of displacement (93.22 percent of responses), compared to 'family members killed or threatened for ethnic/religious reasons' (1.62 percent), 'direct threats to family' (1.45 percent), or 'family members killed or threated for political affiliation' (0.17 percent). However, the prosecution of specific religious groups like the Yazidi points to ethno-religious identity as an important factor for displacement as well. I do not observe the identity of those forcibly displaced in the data, but I have information on the ethno-religious composition of the locations chosen by internally displaced people. The third key results is that more diverse locations host more internally displaced people, compared to locations that are strictly Sunni, Shia, Mixed Sunni-Shia, or Turcoman. There is a positive takeaway: diverse locations appear more welcoming as they host more displaced people. In the broader context of the existing literature, my findings suggest that the identity of internally displaced individuals may be crucial in certain situations, such as conflicts entwined with ethnic cleansing or political prosecution, while being less essential in others.

The external validity of the findings in this paper should be considered with caution. This paper provides insights into the relationship between conflict and internal displacement using specific geographical and temporal events. The extent to the generalization of these findings to other regions, even conflicts within the Middle East with similar war factions, remains an empirical question. Namely, the distance at which conflict is a predictor for displacement is inherently tied to Iraq's geography and road infrastructure. There are other factors such as community support that are only partially addressed in this paper through the ethnic composition, but the findings point to the opposite direction: more diverse locations are more likely to host displaced people. Finally, conflicts can be of different nature and severity; internal displacement as a response to ethnic cleansing might not be temporary in nature as the displacement studied in this paper. These findings might match forcibly displaced people patterns better in any of the war-plagued countries in Middle East than in other regions of the world, and while internally displaced people are different across countries, and there might be some differences in the nature of displacement, this paper offers some insights which would be useful to understand forced displacement globally. Generally speaking, internally displaced people in Iraq do not migrate, they are temporarily displaced and remain very near conflict to return home as soon as possible. Note that the distinction between internally displaced people and refugees is based on whether they cross a national border or not. In the absence of strict political borders, the findings of this paper might be applied to situations where refugees intend to return home as soon as possible even if they have crossed a national border very close to their location of origin. This situation is common between Middle East countries such as Lebanon, Jordan, Iraq, or Syria, but not exclusive to this world region.

Notes

[1]In 2020, UNHCR reported that an unprecedented 82.4 million people around the world were being forced away from their homes because of conflict and prosecution: 48 million internally displaced persons (IDPs), 31.3 refugees, and 4.1 million asylum-seekers. Like refugees, IDPs are individuals who have been forced to flee their home because of persecution, war, or violence. IDPs have a well-founded fear of persecution for reasons of race, religion, nationality, political opinion, or membership in a particular social group, but remain in their own country and do not cross an international border.

[2]The mass migration of Soviet Jews to Israel was as a consequence of the lifting of emigration restrictions in an unstable USSR and by the open immigration policy of Israel toward Soviet Jews.

[3]Please note however that the quality of the data might differ by country as often times data construction also benefits from embedding data from censuses in the case of population density, or anonymous tracking information in the case of road networks to update and improve the quality of the data.

[4]Please see "Iraqi War against ISIL in the Annex for a detailed account of the progression of the war.

[5]Figure 16 displays the distribution of Ethnic and Religious identities in Iraq.

[6]Readers not familiar with geographical information systems (GIS) can think of this type of data construction as layers which need to be properly placed on top of one another before they can be flattened out into a cross-section. Because each of these databases includes the geographical location for each observation, they are called GIS layers. In practice, we are able to load our information on a geographical software so that, with the proper coordinate system, we can align our layers correctly and start understanding the geographical relationship between the different layers.

[7]Using WorldPop or Iraqi night light intensity as an alternative data source to define populated areas yield the same results.

[8]Versions of the model only including "active" cells deliver the same results.

[9]Manresa (2016) imposes sparsity to achieve dimensionality reduction. In her case, this is sufficient as for the applications she has in mind, only a few star units have effects over the other units.

[10]The spatial weight matrix elements measure the distance – physical or social – between any two units. This method bridges networks and spatial econometrics by defining the network matrix parameter restrictions as a function of a spatial weight matrix, while still allowing for the network matrix to be estimated without imposing the rigid structure of a spatial weight matrix.

[11]$\mathbb{N}_{ij} = \mathbb{N}_{kl}$ if $\mathbb{D}_{ij} \approx \mathbb{D}_{kl}$. Where i, j, k, l are different units in the panel.

References

Ades, A. and H. B. Chua (1997). Thy neighbor's curse: regional instability and economic growth. *Journal of Economic Growth 2*(3), 279–304.

Allansson, M., E. Melander, and L. Themnér (2017). Organized violence, 1989-2016. *Journal of Peace Research*.

Angrist, J. D. and A. D. Kugler (2003). Protective or counter-productive? labour market institutions and the effect of immigration on eu natives. *The Economic Journal 113*(488), F302–F331.

Angrist, J. D. and A. D. Kugler (2008). Rural windfall or a new resource curse? coca, income, and civil conflict in colombia. *The Review of Economics and Statistics 90*(2), 191–215.

Arias, M. A., A. M. Ibáñez, and P. Querubin (2014). The desire to return during civil war: evidence for internally displaced populations in colombia. *Peace Economics, Peace Science and Public Policy 20*(1), 209–233.

Åslund, O. and D.-O. Rooth (2007). Do when and where matter? initial labour market conditions and immigrant earnings. *The Economic Journal 117*(518), 422–448.

Balcells, L., A. Steele, et al. (2012). Warfare, political identities, and displacement in spain and colombia. Technical report, Households in Conflict Network.

Bellows, J. and E. Miguel (2009). War and local collective action in sierra leone. *Journal of public Economics 93*(11), 1144–1157.

Berman, E., J. N. Shapiro, and J. H. Felter (2011). Can hearts and minds be bought? the economics of counterinsurgency in iraq. *Journal of Political Economy 119*(4), 766–819.

Bevelander, P. and R. Pendakur (2014). The labour market integration of refugee and family reunion immigrants: A comparison of outcomes in canada and sweden. *Journal of Ethnic and Migration Studies 40*(5), 689–709.

Blanco, M. and J. F. Vargas (2014). Can sms technology improve low take-up of social benefits? *Peace Economics, Peace Science and Public Policy 20*(1), 61–81.

Blattman, C. and E. Miguel (2010). Civil war. *Journal of Economic literature 48*(1), 3–57.

Bodvarsson, Ö. B., H. F. Van den Berg, and J. J. Lewer (2008). Measuring immigration's effects on labor demand: A reexamination of the mariel boatlift. *Labour Economics 15*(4), 560–574.

Borjas, G. J. (2017). The labor supply of undocumented immigrants. *Labour Economics 46*, 1–13.

Borjas, G. J. and J. Monras (2016). The labor market consequences of refugee supply shocks. Technical report, National Bureau of Economic Research.

Buhaug, H. and J. K. Rød (2006). Local determinants of african civil wars, 1970–2001. *Political Geography 25*(3), 315–335.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Industrial & Labor Relations Review 43*(2), 245–257.

Case, A. C., H. S. Rosen, and J. R. Hines (1993). Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of public economics 52*(3), 285–307.

Cederman, L.-E., H. Buhaug, and J. K. Rød (2009). Ethno-nationalist dyads and civil war: A gis-based analysis. *Journal of Conflict Resolution 53*(4), 496–525.

Chatty, D. (2003). Operation iraqi freedom'and its phantom million iraqi refugees. *Forced Migration Review 18*, 51.

Chatty, D. and N. Mansour (2011a). Displaced iraqis: predicaments and perceptions in exile in the middle east. *Refuge: Canada's Journal on Refugees 28*(1).

Chatty, D. and N. Mansour (2011b). Unlocking protracted displacement: An iraqi case study. *Refugee Survey Quarterly*, hdr012.

Choi, S.-W. and J. A. Piazza (2014). Internally displaced populations and suicide terrorism. *Journal of Conflict Resolution*, 0022002714550086.

Christensen, A., A. Christensen, and N. Harild (2009). *Forced displacement: The development challenge*. World Bank Washington.

Cockburn, P. (2015). *The rise of Islamic State: ISIS and the new Sunni revolution*. Verso Books.

Conley, T. G. and C. R. Udry (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review 100*(1), 35–69.

Cortés, D., D. Montolio, et al. (2013). Publicness of goods and violent conflict: Evidence from colombia. *Serie documentos de trabajo. No 137 (Abril 2013)*.

Croicu, M. and R. Sundberg (2015). Ucdp georeferenced event dataset codebook version 5.0. *Journal of Peace Research 50*(4), 523–532.

Currarini, S., M. O. Jackson, and P. Pin (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences 107*(11), 4857–4861.

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

Del Carpio, X. V. and M. C. Wagner (2015). The impact of syrians refugees on the turkish labor market. *World Bank Policy Research working paper* (7402).

Friedberg, R. M. (2001). The impact of mass migration on the israeli labor market. *Quarterly Journal of Economics*, 1373–1408.

Jaeger, D. A. and M. D. Paserman (2008). The cycle of violence? an empirical analysis of fatalities in the palestinian-israeli conflict. *The American Economic Review 98*(4), 1591–1604.

Konuk, B. B., S. Tumen, et al. (2016). Immigration and prices: Quasi-experimental evidence from syrian refugees in turkey. Technical report.

Lupu, Y. and G. P. Wallace (2023). Targeting and public opinion: An experimental analysis in ukraine. *Journal of Conflict Resolution 67*(5), 951–978.

Manresa, E. (2016). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*.

Marfleet, P. (2011). Displacement and denial: Idps in today's iraq. *International Journal of Contemporary Iraqi Studies 5*(2), 277–292.

Marfleet, P. and D. Chatty (2009). *Iraq's Refugees: Beyond" tolerance"*. University of Oxford, Refugee Studies Centre.

Margesson, R., J. M. Sharp, and A. Bruno (2008). Iraqi refugees and internally displaced persons: A deepening humanitarian crisis? DTIC Document.

Maystadt, J.-F. and G. Duranton (2018). The development push of refugees: Evidence from tanzania. *Journal of Economic Geography 19*(2), 299–334.

Melander, E. and M. Öberg (2007). The threat of violence and forced migration: Geographical scope trumps intensity of fighting. *Civil Wars 9*(2), 156–173.

Nunn, N. and N. Qian (2014). Us food aid and civil conflict. *The American Economic Review 104*(6), 1630–1666.

Rogall, T. (2014). Mobilizing the masses for genocide. *Unpublished Manuscript*.

Tumen, S. (2016). The economic impact of syrian refugees on host countries: Quasi-experimental evidence from turkey. *The American Economic Review 106*(5), 456–460.

Vargas-Silva, C. (2016). Literature review: Remittances sent to and from refugees and internally displaced persons.

Voors, M. J., E. E. Nillesen, P. Verwimp, E. H. Bulte, R. Lensink, and D. P. Van Soest (2012). Violent conflict and behavior: a field experiment in burundi. *The American Economic Review 102*(2), 941–964.

Annex

Conflict: Uppsala Conflict Data Project Georeferenced Event Data

UCDP GED provides the geographical location of conflict events that include several features. The features of interest for this paper are start and end dates, type of conflict, factions involved, combatant deaths and civilian deaths. The Uppsala Conflict Data Program was properly established at the Uppsala University Department of Peace and Conflict Research in the mid-1980s under the name Conflict Data Project. It continuously collects data on armed conflicts using conflict definitions designed so as to pick up the same phenomenon across time as well as across space. The coding rules are very strict, and the standards are set very high for inclusion of information. This makes the data useful for systematic studies of the origins of conflict, conflict dynamics and conflict resolution. Their data is free and publicly available online. The data used in this paper is the latest expansion of their data which is named UCDP's Georeferenced Event Data (UCDP GED) (Allansson et al., 2017). The basic unit of analysis for the UCDP GED dataset is the "event", an individual incident (phenomenon) of lethal violence occurring at a given time and place. In particular, an event is defined as *"An incident where armed force was by an organized actor against another organized actor, or against civilians, resulting in at least 1 direct death at a specific location and a specific date"* (Croicu and Sundberg, 2015) [12]. UCDP GED includes conflict data on 115 countries. There were 6,495 registered conflict events from the January 1st 1989 until December 31st 2016 in Iraqi territory[13]. In this paper, I will focus on the period 2014-present which corresponds with the Iraqi Civil War. As we can see in Figure (8), the levels of violence in Iraq were already very high even before the outbreak of the Syrian Civil War in 2011 with 6,810 conflict-induced deaths and 2,651 civilian deaths from January 2011 to December 31st 2013. However, the Iraqi Civil War with 37,222 conflict-induced deaths and 8,358 civilian deaths from January 2014 to January 2017, crystallized a wave of violence that doubled the figures from the Iraq War in 2003. During the 1095 day period between January 1st 2014 and December 31st 2016, there were 2,062 conflict events recorded in UCDP GED at 825 different dates.

During the Iraqi Civil War (2014-) most observations (98.4%) involve ISIL as either a direct clash against the Government of Iraq Military forces (75.58% of observations) or a one-sided civilian massacre (22.82% of observations). Table 3 displays the frequency of the different faction-dyadic values in Iraq in the period 2010-2016 as recorded in UCDP GED[14]. The best temporal precision available in the UCDP GED dataset events is the day. Conflict events are spread over all locations with an important concentration in Mosul and Baghdad. Nonetheless, this concentration is never higher than 17%. This dispersion is relevant as we are interested in the

movement of IDPs as a response to a generalized level of conflict all over Iraq. If all the reported conflicts were located in the same biggest most exposed locations we would not not be able to explain the movements of people locally.

Figure 8: Monthly conflict-induced deaths (2010-2016)

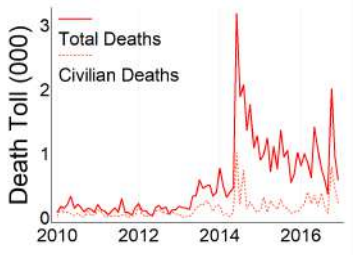


Figure 8 displays total and civilian deaths over time aggregated by month from January 2010 until December 2016. We can see that while conflict-induced deaths have been substantial since at least 2010, conflict intensity increases greatly in 2014.

IDP: International Organization for Migration's DTM

The IDP data used in this paper comes from the Displacement Tracking Matrix (DTM) master list which is IOM's information system to track the Iraqi displaced population. DTM's master lists collect data on numbers and locations of IDPs using an extended network of over 4,000 key informants. Once a location is identified, IOM's Rapid Assessment and Response Teams (RARTs)—composed of 140 field staff— visit and directly confirm the new IDP location and collect more detailed information on the displaced population. The DTM is an ongoing data collection system that identifies and routinely updates figures through contacts with key informants. DTM Iraq collects information on people displaced after December 2013, which they understand to be strictly ISIS-induced IDPs. The data collected is on IDPs defined according to the Guiding Principles on Internal Displacement: internally displaced persons (IDPs) are *"persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized state border."* (United Nations, 1998). The unit of observation is the location. Locations' boundaries are determined on the basis of key informants and RARTs' knowledge and evaluation. When the teams identify a new location where IDPs are living, the location is added to the database and a unique identifier (place_id) is created and assigned. If one location no longer hosts IDPs, the number of families living in the areas and all the other attributes are zeroed; the record is kept in the database as an inactive location that

used to host IDPs but no longer does so.

DTM master lists are fully updated in one calendar month, which means that information on all locations is updated once a month. In two weeks, approximately 50% of the locations are updated, data is sent to the IOM Information Management Unit, and the dataset with partial updates is released after quality control, while the teams continue to update information from the remaining locations. By the end of the month, the update is complete and the DTM report is published with fully updated information on IDPs and returnees.

Features include the GPS coordinates of the place of displacement where the population is identified, the governorate of origin for IDPs, the wave of displacement (the DTM conventionally identifies six displacement waves or periods: pre-June 2014; June-July 2014; August 2014; post-September 2014; Post April 2015; and Post March 2016 onwards), the shelter type (the DTM identifies four shelter category: camps; critical shelter arrangements (informal settlements, religious buildings, schools; unfinished or abandoned buildings; and other formal settlements/ collective centers); private dwellings (host communities, rented houses, and hotels/motels); and unknown (applies to locations not accessible when the shelter type cannot be identified). The master lists collects information on the total number of families displaced to a location at the time of data collection thus at every round of updates, the new count replaces the old count. The new count can be lower/higher than the previous count if the inflow is smaller/bigger than the outflow, or it can be zero if all IDPs/returnees left the place where they were previously identified. Once a location stops hosting IDPs or returnees, the DTM does not track personal IDP movements; that is, if specific families returned home, moved to a different shelter in the same location, or moved to a different location still away from home. Instead, the DTM methodology is designed to regularly monitor and update all IDP locations, thus enabling a continuous countrywide coverage of the main characteristics of the IDP population.

Figure 9: Total number of IDPs (2014-2017)

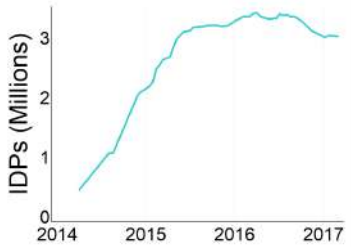
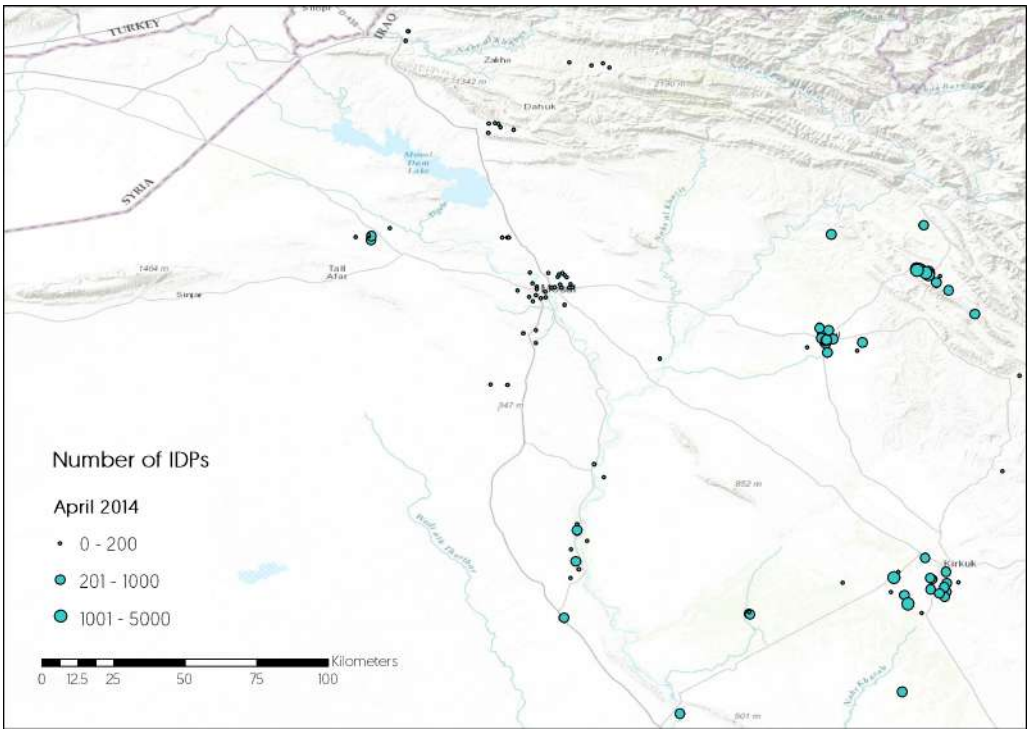


Figure 9 displays the evolution of the number of IDPs from the first published DTM list until January 1st 2017. Coinciding with the Iraqi Civil War outbreak in January 2014, the number of internally displaced persons increased dramatically reaching over 2,000,000 displaced people in 2015 and almost 3,500,000 IDPs by January 2016.

Figure 10: IDP Locations near Mosul (April 2014 and March 2017)



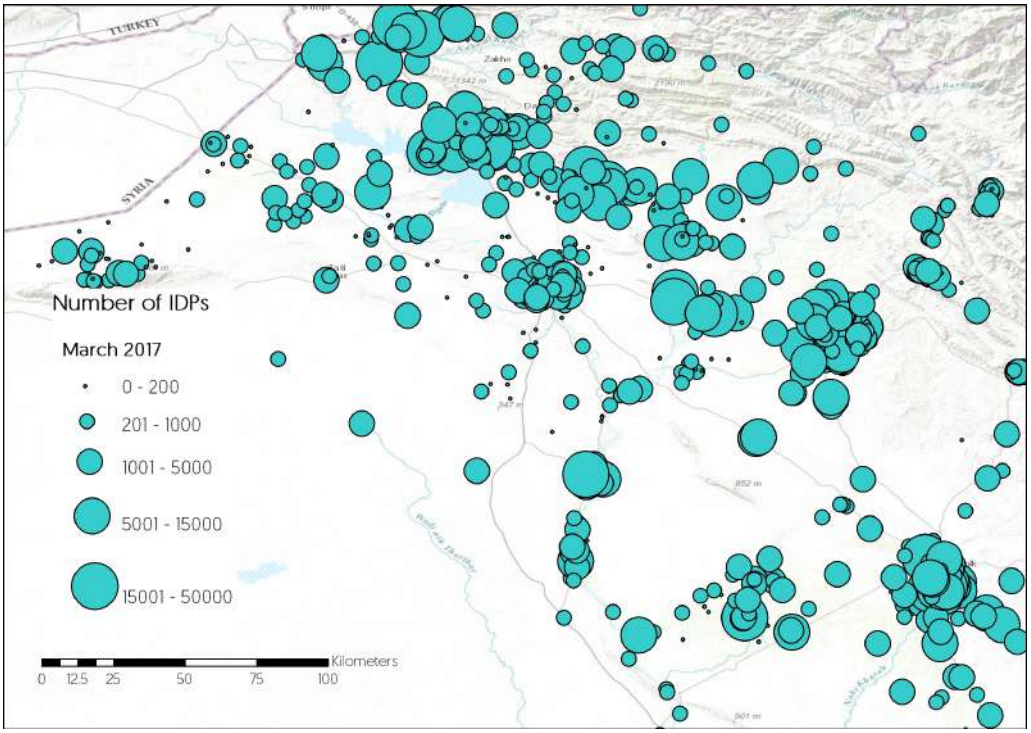


Figure 11: IDP Locations near Baghdad (April 2014)

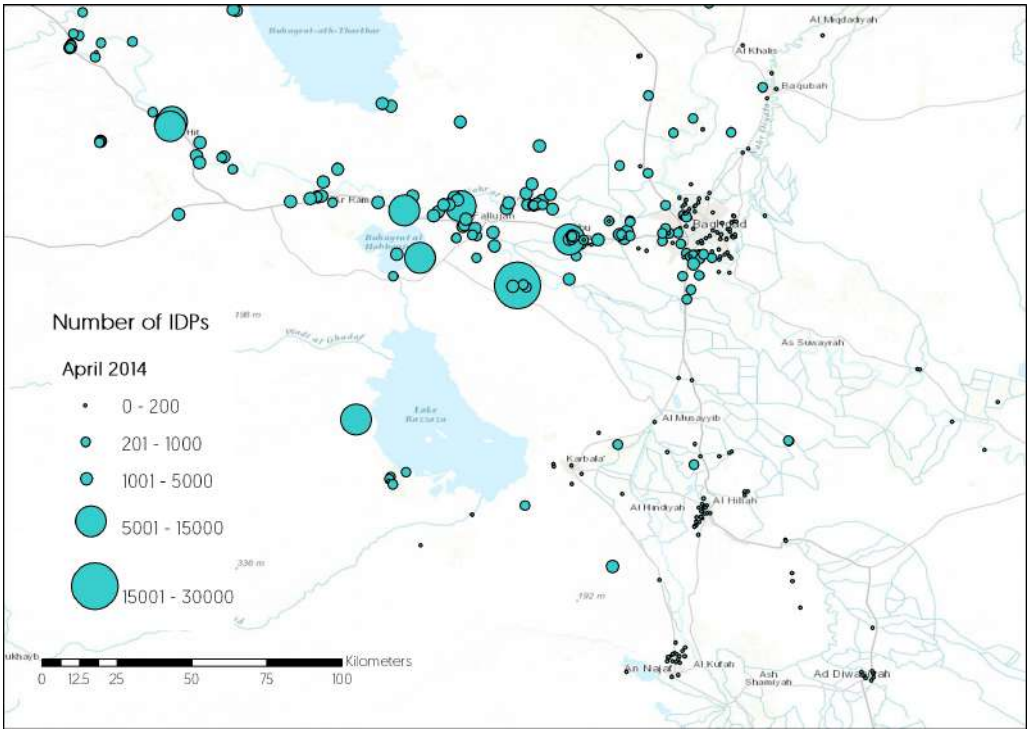
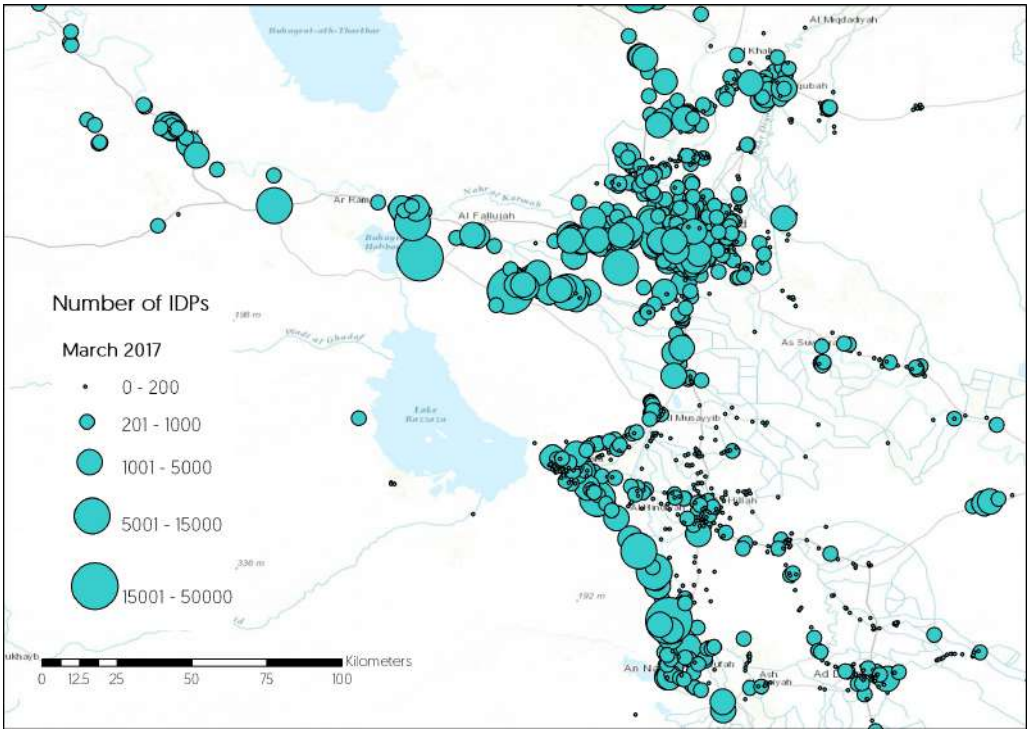


Figure 12: IDP Locations near Baghdad (March 2017)



Figures 2 and 3 provide an overview of the geographical extension of the internally displaced population in Iraq through the Iraqi Civil War. In April 2014, there were 815 IDP locations. Small IDP sites could be found in the governorates of Salah al-Din, Erbil, Kirkuk and Sulaymaniyah. In addition, very few displaced persons are tracked in the Shi'ite governorates of Southeast Iraq. Most of the IDP concentration is located in the province of Al-Anbar where the biggest IDP locations were established. Two of the biggest IDP sites are located near Ramadi with over 5,400 IDPs each, two more IDP sites were located near Hit with over 5,000 IDPs each. Finally, near Fallujah, there were two IDP sites with over 6,000 and 9,900 IDPs respectively and the biggest IDP site recorded to date with 25,536 displaced individuals. In April 2014, there were over 74,000 displaced families which constituted almost half a million displaced individuals. In March 2017, there were 3,660 active IDP locations and 118 of them hosted over 5,000 individuals. As we can see in Figure 3, southeastern Iraq still hosts few displaced individuals relatively to the conflict-intense areas in Northern Iraq and the surroundings of Ramadi, Fallujah and Iraq. The biggest recorded IDP site in March 2017 was the Amriyat Al-Fallujah camp which hosted over 44,000 individuals. Other huge IDP locations were the Khazer M1 camp in Ninewa with over 31,000 individuals, the Hay Al Jamia-Mahalla IDP site in Kirkuk, the Habbaniya tourist city camp in Fallujah, the IDP site in Khabat in Erbil and the Chamishku camp in Dahuk, all of them with over 28,000 displaced persons. In March 2017, there were over half a million displaced families which constituted over three million displaced individuals. Figures 11, 12, display the same time periods as figures 2 and 3 but instead of showing the whole Iraqi geography, they focus on the regions surrounding Baghdad and Fallujah. Figure 10 shows the location and size of IDP locations around Mosul in April 2014 and March 2017.

Geographical controls

The geographical data used in this paper comes from LandScan Population 2011, ESRI Database and ESOC, WorldPop, and Nighttime Light series by NOAA for population density. Open Street Map (OSM) for road networks. Government administrative data was collected from the ESRI Database. Ethnicity data is drawn from Berman et al. (2011). This data plotted in a series of maps can be found in the Annex: Figures 13, 14, 15, and 16.

Iraqi War against ISIL

The purpose of this section is to provide the reader with an overall understanding of the nationwide spread and intensity of the ongoing conflict in Iraq as well as important dates of international intervention. Between June 2014 and December 2016 alone, there were 2,062 different conflict events recorded in the UCDP GED dataset. To illustrate the overall level of conflict in Iraq during this period, a description of important events is presented.

Year 2014: By June 2014, ISIL had seized control over 70% of the Al Anbar Governorate, including the cities of Fallujah, the border crossing Al Qaim and half of Ramadi[15]. This campaign triggered the first major IDP wave in Iraq. Moreover, following large scale offensives in northern Iraq, ISIL took control of Mosul, the second largest city in the country. During the same month, ISIL also captured parts of Kirkuk and Diyala provinces and Tikrit. During this offensive in northern Iraq, IS advanced to within 30km of the Iraqi Kurdish capital of Erbil. In August 2014, ISIL captured Sinjar causing almost 200,000 civilians (most of them Yazidis) to flee from the fighting in the city. About 50,000 people took refuge in the Sinjar mountains where they were trapped without food, water or medical care. On August 7th, President Obama authorized targeted airstrikes in Iraq against IS along with airdrops of aid. By August 13th, US aerial bombings and Kurdish land forces broke the IS siege of Mount Sinjar. On August 18th, Kurdish Peshmerga ground troops together with Iraqi Special Forces and the US air campaign reclaimed the Mosul Dam. By the end of August, western forces began airstrikes and humanitarian air drops to prevent potential massacres against the Shi'a Turkmen minority in Amerli. In mid-October 2014, IS forces captured the city of Hit causing an estimated 180,000 civilians to escape from the area, some of them were already internally displaced persons from the Anbar IDP wave of June 2014. On October 21st, ISIL seized control north of Mount Sinjar again, thus cutting the area's escape route to the Kurdish-controlled areas. A month later, Iraqi forces succeeded in retaking control of the city of Baiji and its nearby oil refinery. By the end of December, ISIL siege of the Sinjar Mountains was broken again, enabling more Yazidis to be evacuated. In 2014, there were 2,900 civilians deaths and a total death toll of 15,010 recorded in the UCDP GED.

Year 2015: In January 2015, the Mosul offensive began with Kurdish Peshmerga forces conquering the surroundings of the city. Five days later, the province of Diyala was recaptured by Iraqi forces. On March 2nd, IS was expelled from the city of Tikrit. By May 20th, ISIL seized full control over Ramadi after a 5 day battle. On July 13th, Iraqi forces started the Anbar offensive with the objective of reconquering important cities in the province of Al-Anbar. On October 22nd, Kurdish forces regained control of Baiji from ISIL. The government of Turkey started bombing alleged Kurdistan Workers' Party (PKK) bases in northern Iraq. In November, Kurdish forces re-conquered Sinjar from ISIL as it had shifted hands back in August. In 2015, there were 1,885 civilians deaths and a total death toll of 11,251 recorded in the UCDP GED.

Year 2016: In February 2016, the Battle of Fallujah began and would last until June 2016 with the retreat of ISIL from the city. In March, the Mosul offensive began as the coalition forces tried to take back Mosul city. During 2016, a substantial number of airstrikes killed many ISIL fighters in the two main military operations in Al-Fallujah and Mosul. In 2016, there were 3,550 civilians deaths and a total death toll of 10,961 recorded in the UCDP GED.

Aggregated survey data not used in the estimation model

In addition to the biweekly DTM master lists described above, two location assessment surveys were conducted in September 2015 and March 2016. On top of the information available through the master lists, these surveys include reasons of displacement, future intentions, security incidents and demographic information (age and sex). The nationally aggregated data coming from group assessment surveys are not used in the estimation but provide information which is relevant to understanding the circumstances under which the IDPs choose a specific location, and offer insights to contextualize and interpret the results coming from the estimation model.

Looking at the reported reasons for displacement, future intentions and the security incidents will alleviate some of the estimation concerns which will arise once I present the methodology. Table 4 displays the reported reason of displacement reported by IDP families in March 2016. The survey shows that generalized violence was the most important reason for displacement reported by 93.22% of the IDP families interviewed. The second to most important reported reason was *"family member killed in general violence"*, reported by 2.78% of the families. The third most important reason is *"family members killed or threatened for ethnic/religious reasons"* reported by 1.62% of the IDP families. These reports offer some empirical evidence showing that the exact identity of the displaced people might not be so important in this case as opposed to the cases in Colombia and Spain discussed by Balcells et al. (2012).

Table 5 displays the future intentions reported by the displaced families. Table 5 shows that 93.8% of the displaced families intend to return to their place of origin while *"resettle in a third location"* and *"return to an area of past displacement"* was only chosen by 0.49 and .22% of the displaced families. Finally, only 8 families out of 466,383 reported they intended to move outside of Iraq. These reports offer additional empirical evidence supporting the findings about the intention of IDPs in Colombia found by Arias et al. (2014).

Figures and Tables

Conflict

Table 3: Conflict events in Iraq (2010-2016)

Factions involved	**Freq.**	**Percent**
Government of Iraq - IS	2,554	75.58
IS - Civilians	771	22.82
Government of Turkey - PKK	25	0.74
Government of Iraq - Civilians	9	0.27
Government of Iran - PJAK	8	0.24
IS - JRTN	6	0.18
Jaysh al-Mukhtar - MEK	4	0.12
Government of Iraq - Ansar al-Islam	2	0.06
Total	3,379	100

Aggregated Data: Displaced Persons Survey Results

Table 4: Reason of Displacement (March 2016)

Reason of Displacement	**Freq.**	**Percent**
Generalized violence and armed conflict	434,758	93.22
Family member killed in general violence	12,953	2.78
Family members killed or threatened for ethnic/religious reasons	7,550	1.62
Direct threats to family	6,776	1.45
Evacuated / displaced by the government (to safer sites)	2,450	0.53
Family members killed or threatened for political affiliation	770	0.17
Evicted from property	716	0.15
House damaged / destroyed	164	0.04
Unknown Reason	104	0.02
Dam manipulation / flooding	80	0.02
Evicted by private owners (They wanted their property back)	36	0.01
Lack of access to sustainable income	16	0.00
Lack of access to basic services	5	0.00
Evacuated / displaced by government	5	0.00
Total Families	466,383	100

Table 5: Intentions (March 2016 Group Assessment)

Intentions	**Freq.**	**Percent**
Return to place of origin	437,463	93.80
Waiting on one or several factors	12,982	2.78
Locally integrate in current location	11,441	2.45
Resettle in a third location	2,294	0.49
Unknown intention	1,186	0.25
Return to Area of Past Displacement	1,009	0.22
Moving outside of Iraq	8	0.00
Total Families	466,383	100

Table 6: Shelter Type (March 2016 Group Assessment)

Shelter Type	**Freq.**	**Percent**
Rented housing	200,004	42.88
Host Families	139,206	29.85
Unfinished/Abandoned building	50,731	10.88
Religious Building	26,308	5.64
Informal settlements or collective shelters	21,487	4.61
Camp	14,398	3.09
School Building	6,144	1.32
Hotel/Motel	5,796	1.24
Other	1,241	0.27
Unknown	1,068	0.23
Total Families	466,383	100

Table 7: Main source of income (July-September 2016 Survey)

IDPs' source of income at location (choose 3)	**Freq.**	**Percent**
Paid job (public)	1,036	30.73
Informal commerce or inconsistent daily labor	780	23.14
Paid job (private)	735	21.8
Agriculture / farming / herd animal raising	206	6.11
Savings	198	5.87
Pension	109	3.23
Money from family and/or friends in Iraq	85	2.52
Loans	65	1.93
Aid from national institutions	59	1.75
Aid from international institutions	37	1.1
Business	28	0.83
No source of revenue	24	0.71
Income from rent of house or land	4	0.12
Money from family and/or friends abroad	4	0.12
Other	1	0.03
Total	3,371	100

Geographical data

Figure 13: Iraqi Population Density

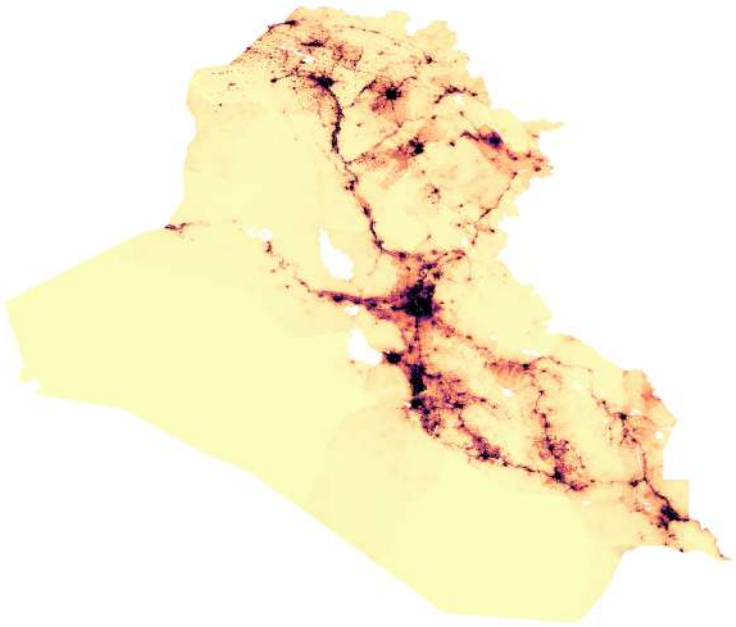


Figure 14: Iraqi Night lights

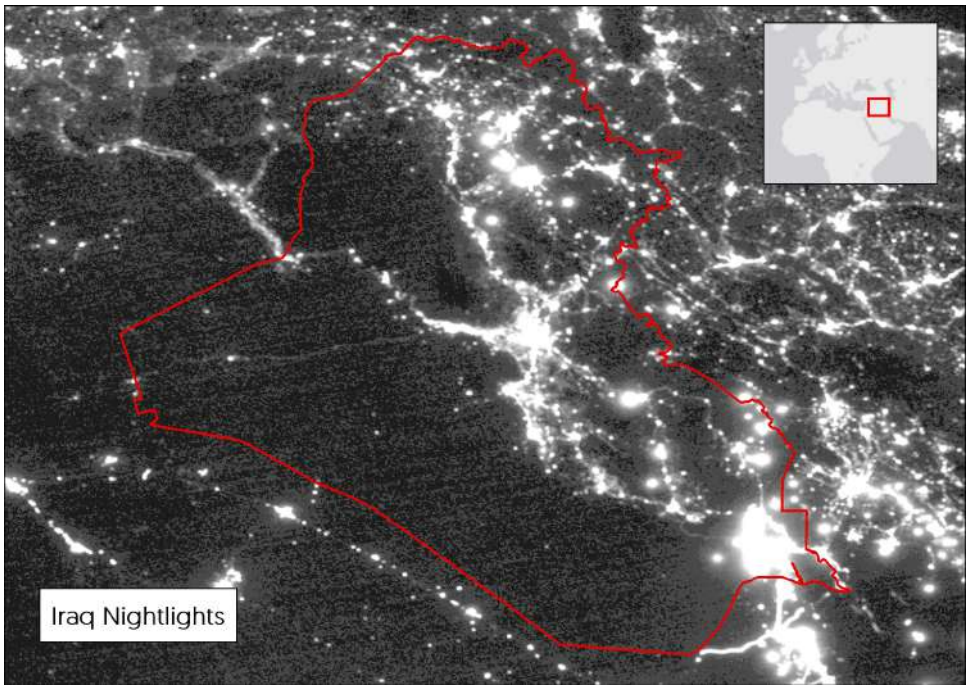


Figure 15: Iraqi roads

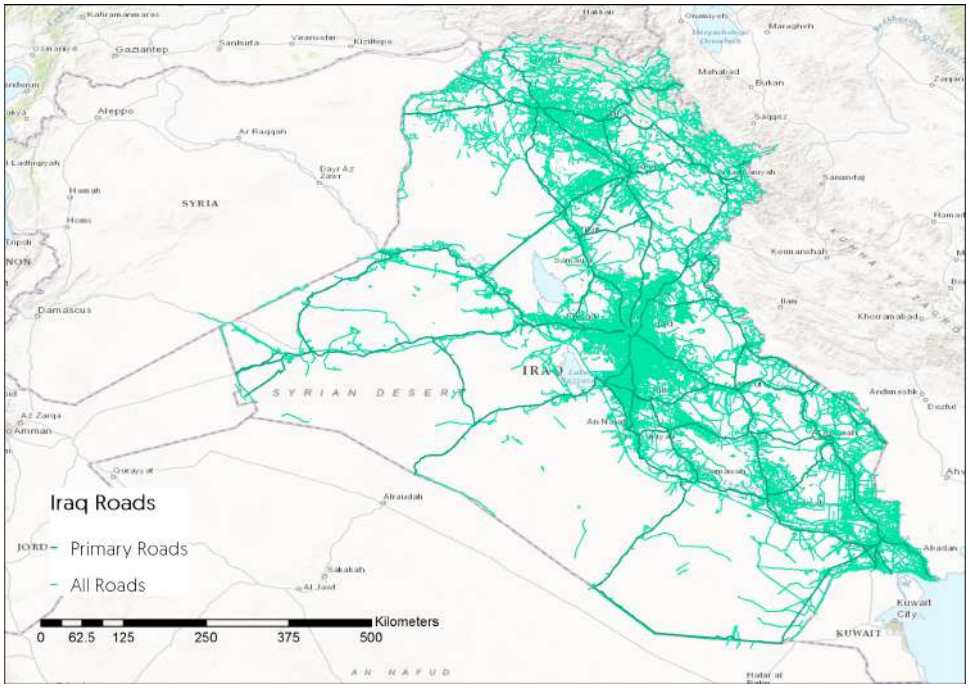
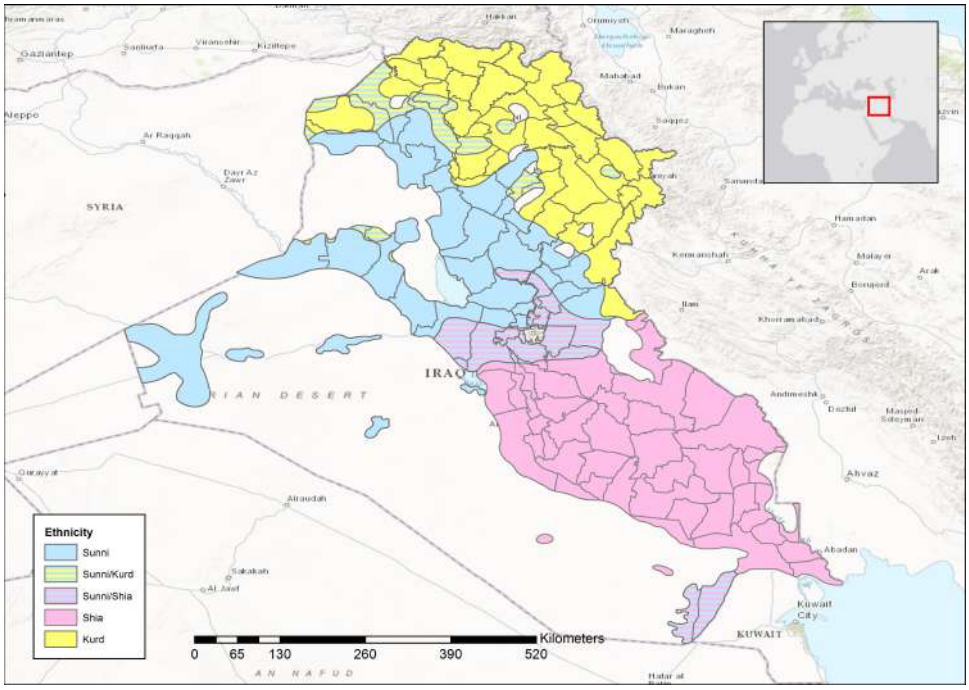
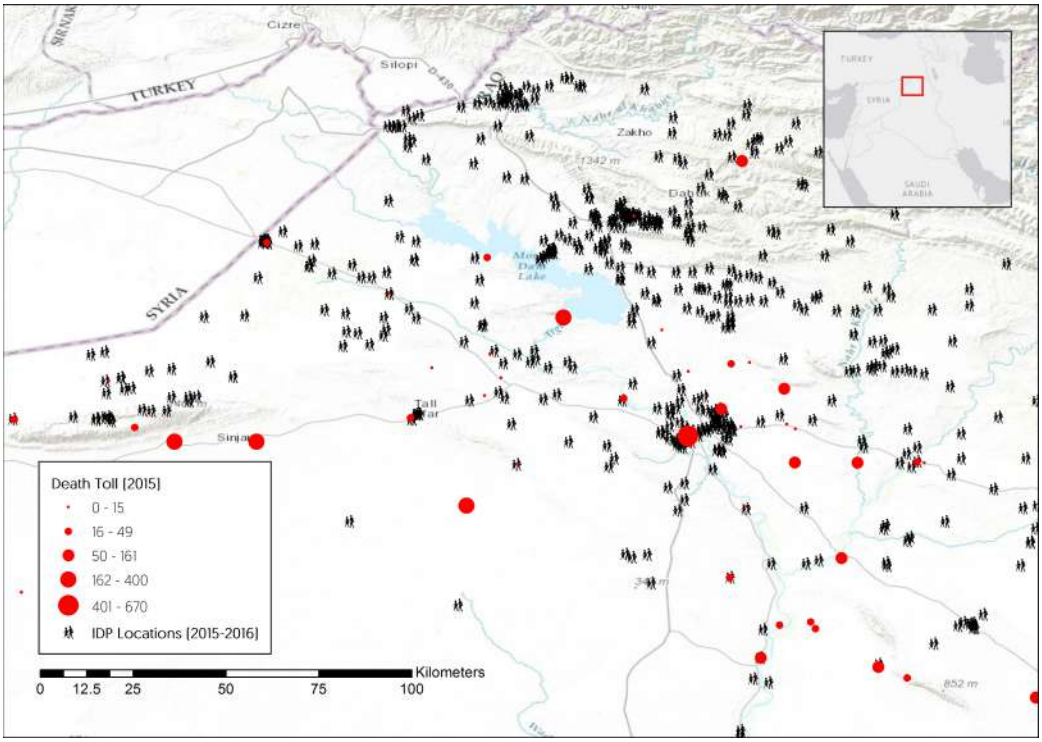


Figure 16: Iraq Ethno-Religious distribution (2005)



Database Construction

Figure 17: Conflicts and IDPs - Mosul area (2015-2016)



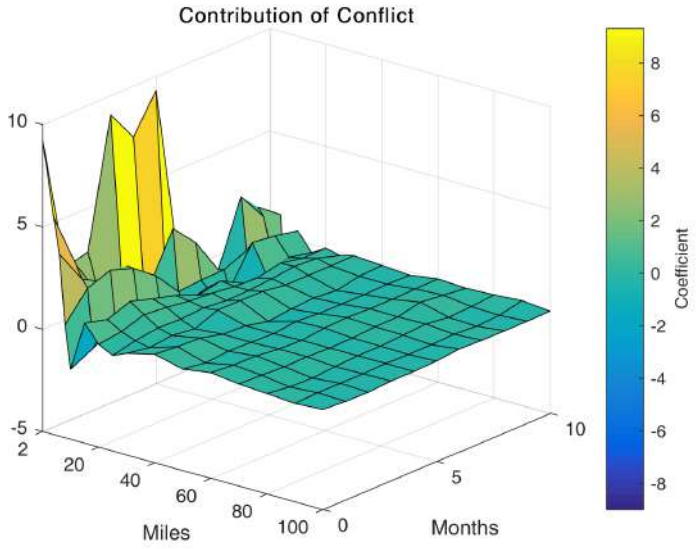
Robustness Checks

Model using linear specification and location fixed effects

$$\mathbf{Y}_t = \eta + \sum_k \mathbb{N}_k^H \mathbf{X}_{k,t}^H + \epsilon_t, \forall t \in \{1, T\}$$

Where $\mathbf{X}_{k,t}^H$ is the history of $\mathbf{X}_{k,t}$ where $\mathbf{X}_{k,t}^H = \mathbf{X}_{k,t} + \mathbf{X}_{k,t-1} + \mathbf{X}_{k,t-2} + ... + \mathbf{X}_{k,0}$ and $\mathbb{N}_k^H$ is the n by n parameter matrix associated to each of the elements in this sum.

Figure 18: Linear Model with Fixed Effects Estimation Results

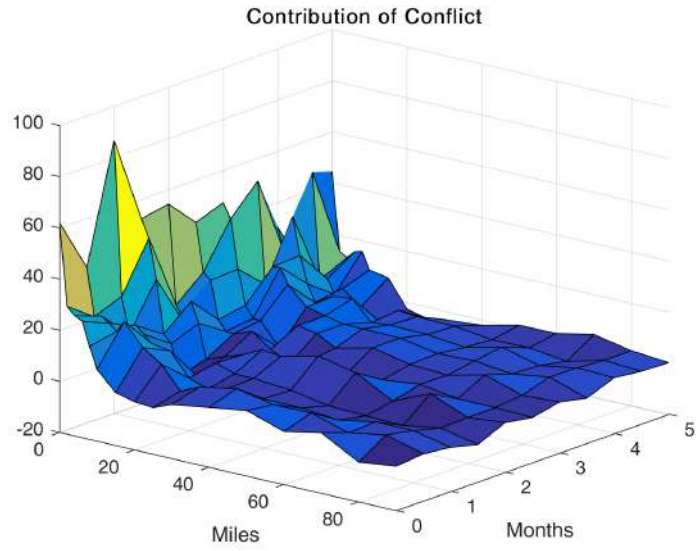


The results are robust to a different specification removing the Tobit/Rectifier function and adding fixed effects for each grid cell of the panel.

Model using Flows instead of Stock of IDPs

In this section presents the results of the baseline model results only using the flow of IDPs instead of the stock of IDPs.

Figure 19: Flow Model Estimation Results

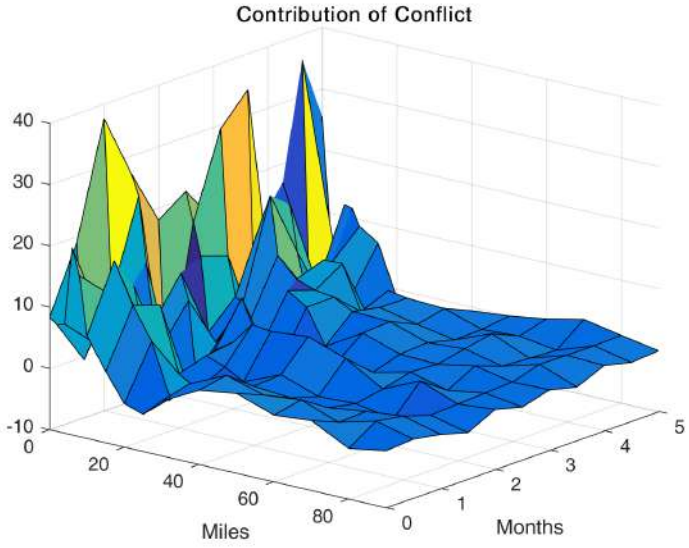


The results are similar to the ones obtained in the baseline model indicating that the model is robust to choice of flow and stock to measure internally displaced people.

Model using Flows and Lagged Stock of IDPs

In this section presents the results of the baseline model only using the flow of IDPs and the lagged stock as a control.

Figure 20: Flow and Lagged Stock Model Estimation Results



The results are similar to the ones obtained in the baseline model indicating that the model is robust to choice of flow, while controlling by the lagged stock to measure internally displaced people.