UC Berkeley
CEGA Working Papers

Title

The Dynamics of Refugee Return: Syrian Refugees and Their MigrationIntentions

Permalink

https://escholarship.org/uc/item/736760c2

Authors

Alrababa'h, Ala'
Masterson, Daniel
Casalis, Marine
et al.

Publication Date

2020-12-03

DOI

10.31235/osf.io/7t2wd

The Dynamics of Refugee Return: Syrian Refugees and Their Migration Intentions

Ala' Alrababa'h, Daniel Masterson, Marine Casalis, Dominik Hangartner and Jeremy Weinstein

CEGA

Center for Effective Global Action

Working Paper Series

Center for Effective Global Action

University of California

eScholarship
Univesity of Californiz

The Dynamics of Refugee Return:
Syrian Refugees and Their Migration Intentions [*]

Ala' Alrababa'h[1,2], Daniel Masterson[2,3], Marine Casalis[2],

Dominik Hangartner[2,4,5], Jeremy Weinstein[1,2]

Word Count: 12,004

[1]Department of Political Science, Stanford University

[2]Immigration Policy Lab, Stanford University and ETH Zurich

[3]Department of Political Science, University of California, Santa Barbara

[4]Center for International and Comparative Studies, ETH Zurich

[5]Department of Government, London School of Economics and Political Science

November 9, 2020

Abstract

Despite the importance of understanding how refugee crises end, little is known about when and why refugees return home. We study the drivers of refugees' decision-making using original observational and experimental data from a representative sample of 3,003 Syrian refugees in Lebanon. We find that conditions in a refugee's home country are the primary drivers of return intentions. Refugees' decisions are influenced primarily by safety and security in their place of origin, their economic prospects, the availability of public services, and their personal networks. Confidence in information is also important, as several drivers of return only impact intentions among people who have high confidence in their information. By contrast, the conditions in refugee–hosting countries—so-called "push" factors—play a much smaller role. Even in the face of hostility and poor living conditions, refugees are unlikely to return unless the situation at home improves significantly.

1 Introduction

Mass forced displacement has proven to be an enduring challenge in contemporary international politics. With a record high of 26 million refugees worldwide and a dearth of effective policy responses, the consequences of persistent refugee crises are profound. Forcibly displaced people face joblessness and food insecurity, lack legal status, and experience hostility and violence in host countries. The governments of hosting countries also struggle to meet the additional demands that large influxes of refugees place on public services and infrastructure (The World Bank, 2017). The consequences of forced migration are particularly acute in developing countries, where more than 85% of refugees reside, because of tight government budgets, weak state capacity, and limited public infrastructure (UNHCR, 2019). Despite the significant challenges that refugee crises pose to refugees themselves, hosting countries, and international donors, effective responses are lacking. Over the last decade less than 1% of refugees received citizenship in a host country each year and only 1–2% were resettled annually (UNHCR, 2019, pp. 28-33). The vast majority of refugees remain in a permanent state of limbo, neither able to integrate locally or to return home.

Refugee crises also have significant political implications. Anti-refugee attitudes can emerge as critical dimensions of political conflict in hosting countries (Lazarev and Sharma, 2017; Alrababa'h et al., 2020; Ghosn, Braithwaite and Chu, 2019; Braithwaite et al., 2019), and exposure to refugee populations can increase xenophobia (Hangartner et al., 2019) and the popularity of far-right parties (Dinas et al., 2019). The presence of refugees also shapes policymaking. For example, hosting countries often impose restrictive policies on refugees' legal and economic status in order to accelerate people's choices to return home. The political attitudes of refugees themselves can play an important role in postwar politics in their home country (Fabbe, Hazlett and Sinmazdemir, 2019). And the movement of refugees can also have security implications, as it may contribute to conflict spillovers and the spread of rebel groups and networks (Salehyan and Gleditsch, 2006; Lischer, 2006; Zhou and Shaver, 2019). More recent work explores the security implications of refugee return (Van Leeuwen and Van Der Haar, 2016; Camarena, 2018; Schwartz, 2019) and the choices of internally displaced people about return to their hometown (Metivier, Stefanovic and Loizides, 2018; Camarena and Hagerdal, 2020).

How do refugee crises end? One possibility is that they end when conflicts terminate. But the historical evidence suggests the situation is far more complicated, both because conflicts wax and wane in their intensity and many refugees choose not to return even as violence

ceases (UNHCR, 2004). To answer this question comprehensively, we need an understanding of when and why refugees choose to return home.

But this is a challenging issue to explore empirically. Existing administrative data on refugee return is incomplete: in the past, many returns went unrecorded and the definition of return varied across organizations and across countries, making systematic analysis difficult and raising concerns about selection bias. Moreover, data collection is challenging with mobile populations. The often-unexpected timing of return means that coordinating field research with real-world events is difficult, especially in contexts of on-going violence. Individuals may also return to violent areas where data collection is either practically infeasible or unethical because it puts participants at risk. This can be a source of differential attrition making it difficult to identify the systematic drivers of return choices.

In this paper, we begin by conceptualizing the decision-making process of refugee households, as they weigh migration choices in order to maximize well-being and minimize risk in an environment of limited information. We inform theory-building through extended interviews and focus groups with refugees and humanitarian organizations working in the context of forced displacement. We posit that people weigh push factors in the host country against pull factors in their home country, considering the possible risks and cost of traveling home. When facing the prospect of return, families must consider not only the risks from generalized violence, but also more personal factors including the threat of targeted violence against returnees, the risk of political persecution, where family and friends are living, and the local economic opportunities in the post-war economy.

We tackle the empirical challenges of studying return with original cross-sectional survey data from a nationally representative sample of 3,003 Syrian refugee households in Lebanon. We use this data to examine predictors of return intentions and preparations, to explore the role of information, and to identify differences in the drivers of short- and long-term return intentions. We supplement this analysis of observational data with a conjoint experiment in order to isolate the causal effect of conditions in Syria and Lebanon on return intentions. Finally, we explore the generality of our findings using a second original survey of Syrian refugees in Jordan.

The Syrian refugee crisis in Lebanon provides a useful case for examining the dynamics of refugee return. When we launched our study in October 2019, active conflict in Syria was diminishing and many governmental and humanitarian organizations had begun discussing and even facilitating returns. At the same time, conditions across Syria varied

widely—many areas remained insecure, and overall prospects for safety, economic recovery, and service provision were uncertain. Moreover, Syrian refugees in Lebanon experienced highly differentiated living conditions, local government policies, and levels of community hostility. In some municipalities, governments actively targeted refugees for harsh treatment and prominent politicians were calling for accelerating their return. We leverage this wide variation in prospects in the country of origin and well-being in the host country to learn about the drivers of return intentions.

The research offers three major takeaways about the drivers of refugee return. First, there is strong evidence that pull factors play a larger role in shaping refugees' return choices than push factors—conditions in Syria are highly predictive of return intentions, whereas conditions in Lebanon are not. Perceptions of safety in Syria are a powerful predictor of return intentions, as are economic conditions and the availability of public services. Personal networks in Syria also play an important role. By contrast, conditions in Lebanon do not significantly shape return intentions, even though some Syrians confront extremely challenging living conditions in the country. Second, refugees' confidence in their information about the situation at home is important for translating underlying preferences into actual plans to return. We find evidence that the role of a number of drivers of return—regime control, economic prospects, access to services, and networks in Syria—are moderated by whether people have high confidence in information about conditions in Syria. Third, the results reaffirm the fundamental humanitarian nature of refugee crises. Despite having been displaced for nearly a decade, people who have fled the violence and societal devastation of civil war generally want to return home when those threats dissipate. Forcible displacement is not a cover for economic migration.

This article advances the research agenda on refugees in two ways. First, it focuses attention on the dynamics of refugee decision-making in the Global South. Much recent research on the causes (e.g., Holland and Peters, 2020) and consequences (e.g., Hangartner et al., 2019) of refugee migration studies contexts in the Global North. However, the majority of refugees are hosted by countries in the Global South, where nearly 55% of refugees live, most often in countries bordering their home country (UNHCR, 2019). Our study seeks to address this mismatch between the geographic distribution of refugees and the areas that have received attention in academic research by squarely focusing on refugees in Lebanon and Jordan—two of the countries most affected by the Syrian refugee crisis.

Second, we explicitly center the agency of people living as refugees. Too often research

conceptualizes refugee crises as a policy problem, rather than a behavioral phenomenon with important political implications. This paper represents one of the first attempts to model and empirically examine the decision problem facing refugees from their own perspective, not from the perspective of policymakers in Europe and the United States (e.g., Dinas et al., 2019). It is difficult to imagine designing effective policy responses, either by hosting governments or the international community, without an understanding of how refugees themselves consider their choices.

More broadly, we contribute to three bodies of research. First, we extend the study of the economics of migration to refugee populations. While a significant body of work examines decisions to migrate, destination choices, and factors promoting or impeding migration (Massey et al., 1993; Fitzgerald, Leblang and Teets, 2014; Bazzi, 2017), the literature has focused primarily on labor migration, and may have limited relevance in contexts of forced migration. Refugees' initial migration choices differ from those of labor migrants (Davenport, Moore and Poe, 2003), suggesting that the drivers of return may differ across the two populations as well.

Second, refugee return is critical for understanding post-conflict reconstruction in both home countries as well as in neighboring states (Blattman, Hartman and Blair, 2014; Bahar et al., 2018; Schwartz, 2019; Ruiz and Vargas-Silva, 2020). Our results underscore that the process of refugee return cannot be rushed. It depends on improving the situation on the ground in the home country (security, economic opportunity, and services), and promoting access to accurate and trusted information for those contemplating return. To the extent that a return of displaced people is a key contributor to economic recovery, post-conflict governments must create hospitable conditions to draw refugees back.

Third, research on immigrant–native dynamics often focuses on host populations in the receiving country, examining the effect of immigration and refugees on local labor markets (Scheve and Slaughter, 2001; Hainmueller and Hiscox, 2010; Malhotra, Margalit and Mo, 2013; Goldstein and Peters, 2014), political attitudes and behavior (Sides and Citrin, 2007; Hopkins, 2010; Hainmueller and Hopkins, 2014), and tensions, discrimination, and conflict (Dancygier, 2010; Adida, 2014; Hangartner et al., 2019). We contribute to this literature by exploring the possibility that these host country dynamics may in turn shape choices about return. Our results show that refugees may be willing to live with extreme hardship in the absence of a viable opportunity to return to their home country.

The structure of the article is as follows. We begin with evidence from historical data

about the relationship between conflict and refugee return. Next, we present a theory of refugee return that highlights the role of push and pull factors as well as mobility costs and the quality of information. We then introduce the case of the Syrian refugee crisis before describing our research design and empirical strategy. We turn to our analysis, presenting results from multiple data sources and empirical approaches. A concluding section discusses the relevance of our findings for both researchers and policymakers, and identifies new questions that should be explored as part of a research agenda on refugee return.

2   When Do Refugees Return?

In considering the question of how refugee crises end, a natural place to begin is the cessation of violence. Indeed, an influential conventional wisdom holds that the return of refugees is largely a post-conflict phenomenon.[1] This presumption is plausible as forced migration is often driven by violence (Davenport, Moore and Poe, 2003; Adhikari, 2013; Schon, 2019; Holland and Peters, 2020), and the end of hostilities in the home country may be a major driver of refugee return. In fact, when conflicts come to an end, refugees often find themselves pushed to return by aid agencies in neighboring countries that see their funding dry up and by governments that feel overstretched by hosting displaced persons (Crisp, 2019). At the same time, many governments in post-conflict countries actively seek the return of refugee populations as they work to catalyze an economic recovery.

How strong is the evidence for this conventional wisdom? Figure 1 presents the mobility patterns from four of the largest refugee crises in recent decades, focusing on Sudan, Afghanistan, Somalia, and Iraq. The figure displays the number of refugees over time (from UNHCR data) and when each country was experiencing active conflict (from UCDP/PRIO Armed Conflict Dataset, defined as at least 25 battle-related deaths in a given year) (*UNHCR Refugee Population Statistics Database*, 2019; Gleditsch et al., 2002; Pettersson, 2019). The observed pattern is consistent with violence driving displacement and the end of active fighting spurring return migration. Specifically, we see that the number of displaced refugees often increases sharply when a conflict begins and tends to decrease after the conflict ends. However, the data also reveal significant variation in the number of refugees during conflicts, sometimes increasing and other times decreasing. Moreover, the number of refugees often

---

[1]See, e.g., Jallow, Heinbecker and Malik (2004); UNHCR (2008); Schwartz (2019) But see Stein and Cuny (1994); Camarena (2018).

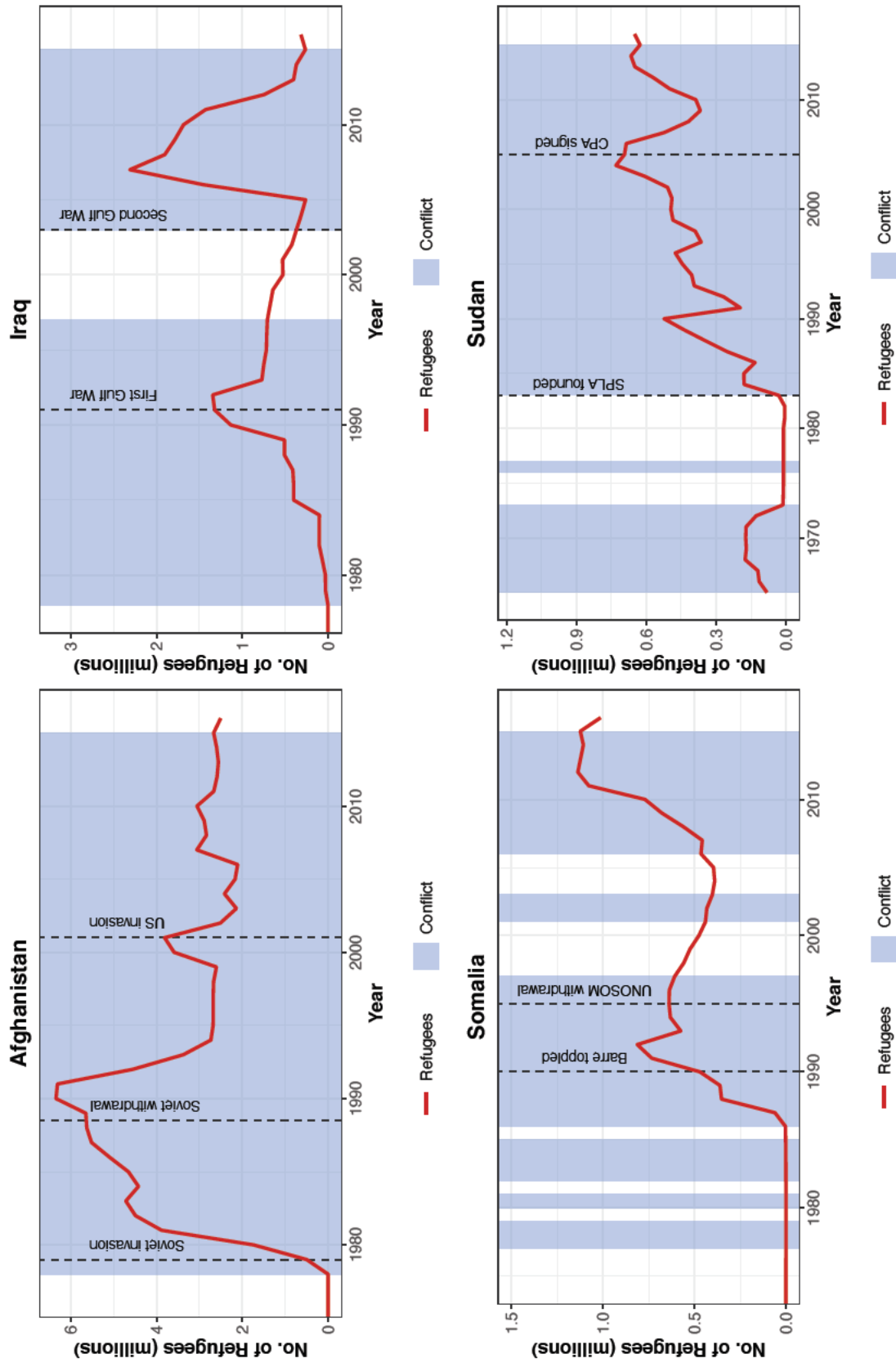decreases very slowly after conflicts come to an end.

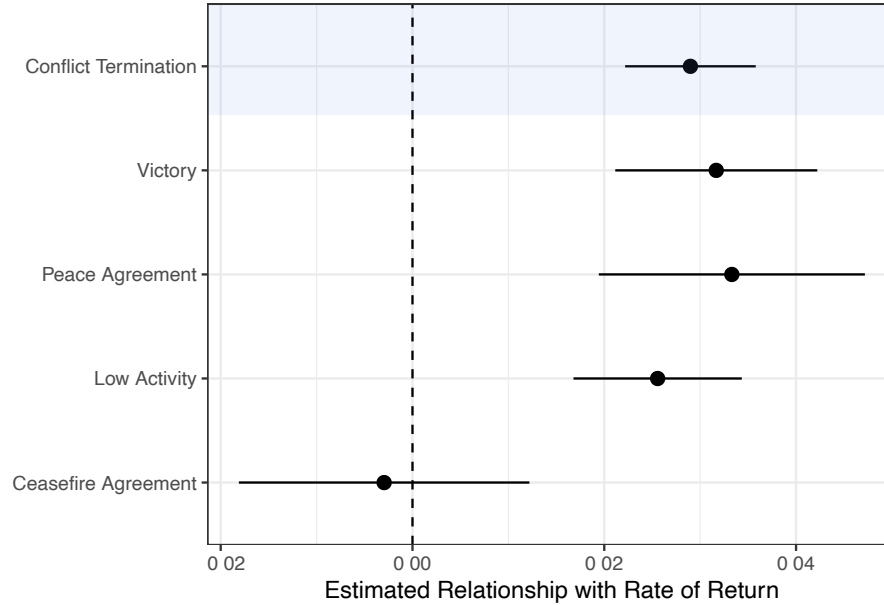Figure 1: Conflict and Number of Refugees Over Time

Figure 2: Relationship between conflict outcome and the rate of return. Point estimates from regressions of refugee return on conflict termination with corresponding 95% confidence intervals.

The data from these four refugee crises suggest that conflict dynamics alone may be insufficient to understand the decisions refugees make about return. The trends observed in these four prominent cases, however, may differ from more general patterns. Therefore, we look more systematically at the relationship between conflict and return with basic time-series regression models.[2] Using historical conflict data from PRIO and refugee migration data from UNHCR between 1951 and 2016, Figure 2 shows the association between conflict termination and the share of refugees who returned in that year.

In the top part of the figure, the estimate suggests that the end of conflict is associated

---

[2] We fit two models: (1) $Y_{it} = \alpha_i + \gamma_t + \rho P_{it} + \theta T_{it} + \lambda X_{it} + \epsilon_{it}$ and (2) $Y_{it} = \alpha_i + \gamma_t + \rho P_{it} + \beta_1 V_{it} + \beta_2 P_{it} + \beta_3 L_{it} + \beta_4 C_{it} + \lambda X_{it} + \epsilon_{it}$. Years are indexed in $t = 1951, ..., 2016$ and γ_t denotes year fixed effects. $i = 1, ..., N$ indexes dyads of a refugee-hosting country and a refugee-sending country, and α_i denotes dyad fixed effects. Y_{it} denotes the logged number of returnees in dyad i in year t. P_{it} is an indicator for whether the refugee-sending country in dyad i was experiencing peacetime in year t. T_{it} is an indicator for whether the refugee-sending country in dyad i experienced the end of conflict in year t. V_{it} is an indicator for whether the refugee-sending country in country dyad i saw the end of conflict through a decisive victory in year t. Similarly P, L, and C, respectively, denote the end of conflict due to a peace agreement, low activity, and a ceasefire agreement. X_{it} is a control for the logged number of refugees in dyad i in year t. ϵ is a mean-zero disturbance term.

with a 3% increase in the rate of return relative to ongoing conflicts. We also can examine the explanatory power of conflict disaggregated by the form of termination. Compared to ongoing conflicts, decisive victories and peace agreements are both associated with about 3% increases in the rate of return. Conflict termination due to a reduction in hostilities is associated with an approximately 2.5% increase in the rate of return. All these estimates are statistically significant at conventional levels. Ceasefire agreements, in contrast, do not have a detectable association with the rate of return.

The formal end of hostilities is clearly an important predictor of refugee return, but it also leaves a great deal of variation unexplained. Given the fact that violence is rarely distributed uniformly across geographic space or time, it is not surprising that refugee numbers wax and wane during conflict and do not automatically drop when a war notionally comes to an end. Or it may be the case that a particular party to the conflict is victorious while another one is defeated, leaving some refugees concerned about their safety if they were to return. Further, the highly dynamic patterns of refugee return depicted in Figure 1 suggest that factors other than conflict at a macro-level are shaping refugee return.

With so much variation in refugee return left unexplained by an aggregate measure of conflict, we pivot to exploring the dynamics of return through the lens of household decision-making. Approaching return migration as a function of household decisions requires that we consider people's preferences, the environment in which they live, the context to which they might return, as well as other factors including the costs of moving and people's access to information. A focus on households enables us to consider the impact of macro-level changes in a home country or a host country, meso-level sub-national conflict processes, as well as micro-level measures of household experiences, beliefs, and resources.

We define return as moving from a *host* country to one's *home* country with no immediate plans to depart again. Our focus is on the binary choice of whether to return to the home country, thereby setting aside other migration-related choices that refugees face such as internal migration within a host country, location choice within their home country after return, and formal or informal migration to a third country. We offer this definition with an awareness that during war and in its aftermath, the process of return may not be straightforward. Some people may return only to find that the situation in their home country necessitates migrating again in search of safety and a livelihood.

We focus on four major drivers of return: (a) push factors, or the situation in the host country (b) pull factors, or the dynamics in the country of origin, (c) the costs of mobility,

and (d) the role that information plays in how households evaluate the costs and benefits of return.[3]

We developed this framework in two stages. First, we drew on canonical models of individual migration (e.g., Roy, 1951; Borjas, 1987; Heckman and Honore, 1990), treating migration as a binary decision based on a comparison of expected utility between living conditions in a country of residence and a prospective destination. Second, we established substantive grounding for the theory through extended interviews and focus groups with Syrian refugees and staff at humanitarian organizations in Lebanon in the summer of 2018. The qualitative research was essential for adapting canonical models of individual migration to the forced displacement context and to household decision-making.

Push factors: The situation in the hosting country is one primary determinant of people's choices about return migration (Dustmann and Weiss, 2007). Historically, many refugee-hosting countries—even those that are initially receptive to refugee migration—gradually ramp up anti-refugee rhetoric and undercut refugees' legal residency and right to work. Often, the rationale behind such restrictions is that harsher living conditions will incentivize refugees to return home. In the face of such policies, refugee households must actively consider the relative safety of their host country in comparison to their home country. They must also weigh the economic, legal, and social hardships they confront while living as a refugee. In many contexts, refugees live on the margins of society, lacking legal status, being denied basic rights, and confronting discrimination by state authorities. We expect that a range of push factors in the host country will shape decision-making about whether to return including a household's economic situation, access to humanitarian aid, availability of public services, extent of social integration and acceptance, and legal status (e.g., Yahya 2018, Mhaissen and Hodges 2019).

Pull factors: Household decision-making also depends in important ways on the environment in the home country. Refugees must consider the current conditions in their country

---

[3]We ignore an important aspect of refugees' decision problem in our two-country setup. Namely, some people living as refugees may consider migration to third countries when deciding whether to return home or not. We leave this possibility for future research for two reasons. A two-country setup allows us to focus on the essential aspects of the decision to return or not. Our fieldwork and survey data suggest that staying in Lebanon or returning to Syria are by far the two most prominent options that Syrian refugees in Lebanon consider. Second, studying the impact of prospects in third countries faces a significant endogeneity problem. Refugees' prospects in a third country are likely less driven by exogenous conditions in that country than by endogenous personal characteristics.

of origin as well as their expectations about how the situation will evolve. For example, will violence pick up again and would it affect their particular region? Moreover, might they be at risk of targeted persecution or arrest upon return? The threats that people confront come not only from armed conflict, but also from potential retribution. These political dynamics, alongside overall levels of violence, might exert a powerful impact on choices regarding return migration. As households assess their safety if they were to return, they may consider current levels of violence in their hometown, their connections or proximity to existing political divisions, and expectations about continued violence and persecution by the government or armed groups.

As people consider whether to return, they will also evaluate the conditions of the post-war economy and whether they believe they will be able to meet their family's basic needs. When people flee a country, they leave behind homes, jobs, and assets, and refugees' economic prospects upon return will likely vary widely across individuals. The war itself may impact people's economic prospects, by generating disputes about the ownership of housing and other assets (Schwartz, 2019) or when governments claim and distribute property for the sake of demographic engineering or rewarding allies (McNamee, 2018). War often also contributes to the destruction of infrastructure and systems for delivering public services. In a post-war context, patterns of reconstruction and restarting service delivery may reflect geographic and political divisions (Croese, 2017), thereby influencing local livelihoods and choices about return.

Mobility costs: Households must also weigh the financial costs and physical risks associated with moving (Hunt and Mueller, 2004). These costs and risks likely vary depending on one's destination and household characteristics such as family size (Rossi, 1980). Long-distance travel, in particular, may be expensive and require passage through unsafe territory. Return migrants might also face the prospect of being stopped at military checkpoints run by armed groups that charge tolls or taxes, steal possessions, or detain, interrogate, or abuse travelers (e.g., Stork and Abrahams, 2004). Given these concerns, households facing higher mobility costs may be less likely to return.

Information: Finally, we argue that decisions about whether to return are influenced by a fourth factor—information—which may shape how households evaluate the costs and benefits of return. While refugees are generally able to accurately assess their well-being in the host country where they live, beliefs about the conditions in their place of origin are only a guess based on the limited information they have (Munshi, 2003). After months or years

away from home, people need to seek out and piece together information—often incomplete, often contradictory—in order to form expectations about what life would be like if they were to return (Batista and Cestari, 2016). The confidence that people have in their beliefs about the quality of life back home is likely to shape in important ways how they evaluate the costs and benefits of return.

Because refugees are driven from their homes by violence, the cessation of conflict can be a major impetus to return. However, recognizing that conflicts rarely end cleanly and that insecurity remains in many contexts, it is important to evaluate return decision-making at a more micro-level. Examining these decisions at a household level suggests a more complex model based on a range of factors going beyond naive models of war termination.

3   The Syrian Refugee Crisis in Lebanon

To test the drivers of refugee return intentions, we focus on the Syrian refugee crisis in Lebanon. Given the diversity of localities in which Syrians have settled and the heterogeneity in conditions in Syria, this is an excellent case for examining the role of push and pull factors, mobility costs, and information in shaping return intentions. Lebanon, in particular, provides a critical test of the importance of push factors, given the documented hostility, discrimination, and violence that many Syrians have faced in Lebanon. In addition, the context provides meaningful variation in prospects in Syria, mobility costs, and access to information. Syrians in Lebanon vary widely in their characteristics and backgrounds, originating from all of Syria's regions and spanning the country's pre-war socioeconomic spectrum.

What began in Syria in 2011 with street demonstrations and calls for political reforms collapsed into a devastating civil war. Estimates of the death toll range from 371,222 to more than 570,000, and large sections of the country's major cities were destroyed by government bombardment. The war led to an enormous refugee crisis, with millions of people fleeing to Lebanon, Jordan, Turkey, Iraq, Egypt and beyond. As of late 2019, when our study was conducted, more than five million Syrians had fled to neighboring countries and more than six million were displaced inside Syria. Approximately 930,000 Syrians lived in Lebanon, alongside 4.5 million native residents, in a small country with a land area three-quarters the size of the US state of Connecticut.[4]

---

[4]We use UN registration numbers, which provide a conservative estimate of displacement.Refugee population: UNHCR Operational Portal. https://data2.unhcr.org/en/situations/syria. IDP data: UN-

One driving assumption behind Lebanon's national policy agenda for Syrian refugees is that exploitation, vulnerability, and material hardship will force Syrians to leave the country (Janmyr, 2016). Syrians in Lebanon face widespread hostility, confront significant restrictions on the right to work, and have only limited legal status in the host country. Most Syrians in Lebanon lack reliable access to education, healthcare, stable housing, and safe transportation (see, e.g., Mourad, 2017; Lehmann and Masterson, 2020). They live primarily in urban and peri-urban settings, with 15% in Lebanon living in camps, informally managed by NGOs, as the UN did not establish official refugee camps in the country. The situation of Syrians in Lebanon is similar in many respects to the hardship that refugees worldwide face; notably, many governments restrict refugee rights in order to accelerate return and, worldwide, less than one third of the world's 25.9 million refugees live in camps.

As the Syrian government regains control of much of the country, tens of thousands of Syrians have begun returning home, even as violence continues to displace more people. State and non-state actors in Lebanon have begun taking steps to facilitate and push for the return of refugees, tensions between Lebanese and Syrians remain high, and discussions about the return of refugees are increasingly prominent in journalistic and policy circles. Looking to Syria, the war has devastated the country's infrastructure and public services, including the water supply, electricity, schools, and healthcare. Many people fear the persecution and violence that may result from government retribution and collective punishment in the postwar period. People who escaped Syria during the conflict may be especially prone to retaliation by the regime upon return. Men aged 18-42 are subject to military conscription in Syria, and serving in the Syrian military is likely to put conscripts in dangerous situations for years to come, where they may have to kill innocent people or be killed in combat or insurgent attacks. Even if the ultimate victor in the war is no longer in question, the specter of future violence remains.

4   Research Design and Data

4.1   Survey Design

We use original survey data from interviews with a nationally representative sample of 3,003 Syrian refugee households living in Lebanon. The survey measured a wide range of household

---

HCR Refugee Data Finder. https://www.unhcr.org/refugee-statistics/. And IDMC. https://www.internal-displacement.org/countries/syria. Data accessed November 15, 2019.

characteristics, predictors of return, and migration intentions, and also included a conjoint experiment to identify drivers of return intentions. The research team contracted a Lebanese survey firm to conduct data collection, and authors participated in all stages of research including enumerator training, survey piloting, and oversight of data collection. Data collection took place from August to October 2019.

To obtain a representative sample of Syrian households in Lebanon, we used stratified random sampling to ensure variation in Syrian and Lebanese demographics in localities and households sampled. A household head (either gender) served as survey respondent. For a detailed discussion of sampling and random walk protocols, please refer to Appendix Section 1.

4.2   Measuring Return Intentions

Measuring return intentions is challenging, and survey instruments must account for the different time horizons across which households consider decisions in addition to the uncertainty that people face. Capturing intentions precisely is also difficult in the absence of concrete behaviors consistent with stated intentions. As a result, we also measure preparations to return, as a behavioral manifestation of return intentions.

We asked respondents about their return intentions in three ways:

- Return intentions: "Do you plan to return to Syria in the next 12 months?"

- Return preparations: We asked a battery of questions about legal, financial, and logistical steps to prepare for return, and use it to calculate a preparations index with polychoric principal component analysis (PCA).

- Long-term return intentions: "Do you hope to move back to Syria and live there one day?"

It is worth noting that our key outcomes are stated intentions and preparations to return, not a retrospective measurement of actual return choices. Such forward-looking outcomes are the relevant quantity of interest in this case, when policymakers are designing and implementing programs to address the situation before people choose to return. Relatedly, a foundational principle of return policy is ensuring its *voluntary* nature, which requires placing people's *intentions* to return at the center of planning.

4.3   Measuring Drivers of Return

We measure four key concepts that we hypothesize will drive return decisions: (1) well-being in Lebanon, (2) prospective well-being in Syria, (3) information, and (4) mobility costs. To measure concepts 1-3, we draw on data from multiple survey questions and use PCA to construct indices to capture aspects of respondents' living situation in Lebanon, prospects in Syria, and access to information. We present the full set of PCA inputs in Appendix Sections 3 and 4.[5] In both Syria and Lebanon, we measure economic well-being, using data on assets and earning potential in each country, and their current employment, earnings, and aid in Lebanon. We also examine the availability of services, including education, healthcare, water, and electricity, in Lebanon and Syria. We analyze the size of social networks and the number of friends and family in Lebanon and Syria. We examine people's ability to move freely and safely around Lebanon, and their integration in the country using the measures from the IPL-12 integration scale (Harder et al., 2018). To construct an index for the security situation in Syria, we focus on both general factors, such as whether there is still fighting, and personal factors, such as whether a family has any draft-aged men and whether the respondent personally experienced violence. We also construct an index for regime control, including detailed questions on which parties currently and formerly controlled a respondent's hometown.[6] The index for information includes whether the respondent spoke regularly with family or friends in Syria about the situation as well as measures of people's confidence in the information they have about safety, jobs, services, and conscription in their hometown.

A number of these concepts have some overlap in what they measure—most notably confidence in information and social networks. In addition to measuring people's confidence in information directly, we ask questions about the size of refugees' networks in the host and the home country. Networks of family and friends may directly impact people's return choices independent of information they provide, in the sense that many people want to live in the same place as others in their close network. Family and friends may also serve as important sources of information about conditions in one's hometown.

We study mobility costs using two metrics: travel distance to one's hometown and house-

---

[5]Although PCA inputs were pre-specified, some survey questions were mistakenly listed in the PAP for inclusion in two indices. We therefore departed from the PAP in these cases in order to maintain mutually exclusive index inputs. See Appendix Section 6 for a list of these changes.

[6]We deviated from the PAP in separating regime control from safety conditions in Syria. We discovered after data collection that these two concepts did not have a clear relationship, suggesting that they do not deserve inclusion in the same index.

hold size from our survey data. We calculate travel distance from each survey respondent's town of residence in Lebanon to their hometown in Syria, via the Beirut–Damascus highway and border crossing, using the Google Maps API. Our fieldwork revealed that this was the only legal border crossing open at the time of research and that a majority of Syrians moving back travel via official routes.

4.4   Conjoint Experiment

To complement the analysis of observational data, we present a conjoint survey that experimentally manipulates potential drivers of return intentions. This allows us to identify how conditions in Lebanon and Syria, individual circumstances, and social networks influence respondents' thinking about return, while better controlling for potential confounds.

In the conjoint, the enumerator informed respondents: "I will now present you with five conditional scenarios. Please listen to these scenarios carefully and answer the questions about them." Respondents were then read a sequence of five separate vignettes, and after each round, they were asked the following question: "Under these conditions, would you return to Syria?"

The vignettes were presented as follows:

> *Imagine that one year from now, regarding the security situation in Syria, [IN-SERT FROM (1) BELOW]. t appears that in [INSERT HOMETOWN], [INSERT FROM (2)]. As for conscription, [INSERT FROM (3)]. In Lebanon, [INSERT FROM (4)]. Finally, regarding your friends and relatives, are [INSERT FROM (5)].*

In the vignettes, each of the numbered attributes below was randomly given one of the lettered values. The full text of each vignette read as follows:

1. **Safety in Syria:**   (a) Your hometown is quite safe; (b) Your hometown remains insecure; (c) All of Syria is quite safe

2. **Economic conditions in Syria:**   (a) There are many job opportunities; (b) Public services, such as health centers and schools, are relatively easy to attain; (c) There are few job opportunities; (d) Public services, such as health centers and schools, are difficult to attain

3. **Personal safety:**   (a) Military conscription has stopped; (b) Military conscription is still in place

4. **Conditions in Lebanon:**   (a) You have a good job in Lebanon; (b) You do not have a good job in Lebanon; (c) Health centers and schools in Lebanon are available and affordable; (d) Health centers and schools in Lebanon are unavailable and unaffordable

5. **Network effects:**   (a) Most of your friends and relatives are in Lebanon; (b) Most of your friends and relatives are in Syria; (c) Most of your friends and relatives are in Jordan, Turkey, and Iraq

5   Results: Observational Data on Return Intentions

We begin by looking at the distribution of responses about return intentions in Figure 3. We find that return intentions are increasing with the time horizon. Only 5% of Syrians plan to return in the next 12 months, that is, before approximately September 2020, and about a quarter of Syrians anticipate returning before September 2021. 63% plan to return at some point in the future.[7] To put these numbers in context, the median year of arrival for respondents was 2013, meaning that the median respondent had been displaced for more than 6 years at the time of data collection.

   To study how cross-sectional differences shape return intentions, we examine the predictive power of a range of potential drivers of refugee return, using the indices we described earlier. We estimate the following regression model:

$$Y_i = \alpha + \beta T_i + \gamma X_i + \epsilon_i \ , \tag{1}$$

for each outcome Y and a vector of indices T. Each index is the first principal component from a PCA analysis of the measures detailed in Section 2. We also adjust for a vector of control variables, X, including household-level covariates and locality-level fixed effects.[8] Finally, ϵ is a mean-zero disturbance term. We also run a series of regression models similar

---

[7]We impute missing values in our data using multivariate imputation by chained equations, discussed in Appendix Section 3.3.

[8]In regressions including travel distance on the right-hand side, we drop controls for location in Lebanon and hometown in Syria, since travel distance is calculated using these two geographic variables.
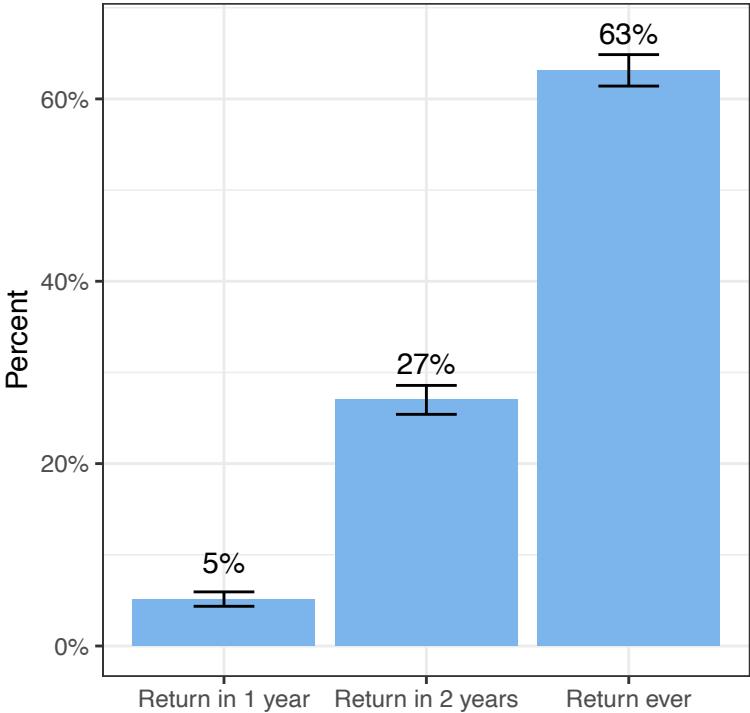
Figure 3: Return Intentions (Short, medium, and long-term)

to Equation 1, but where the vector of indices T is replaced with each respective index in one model.[9]

5.1 Analysis of Drivers of Return Intentions
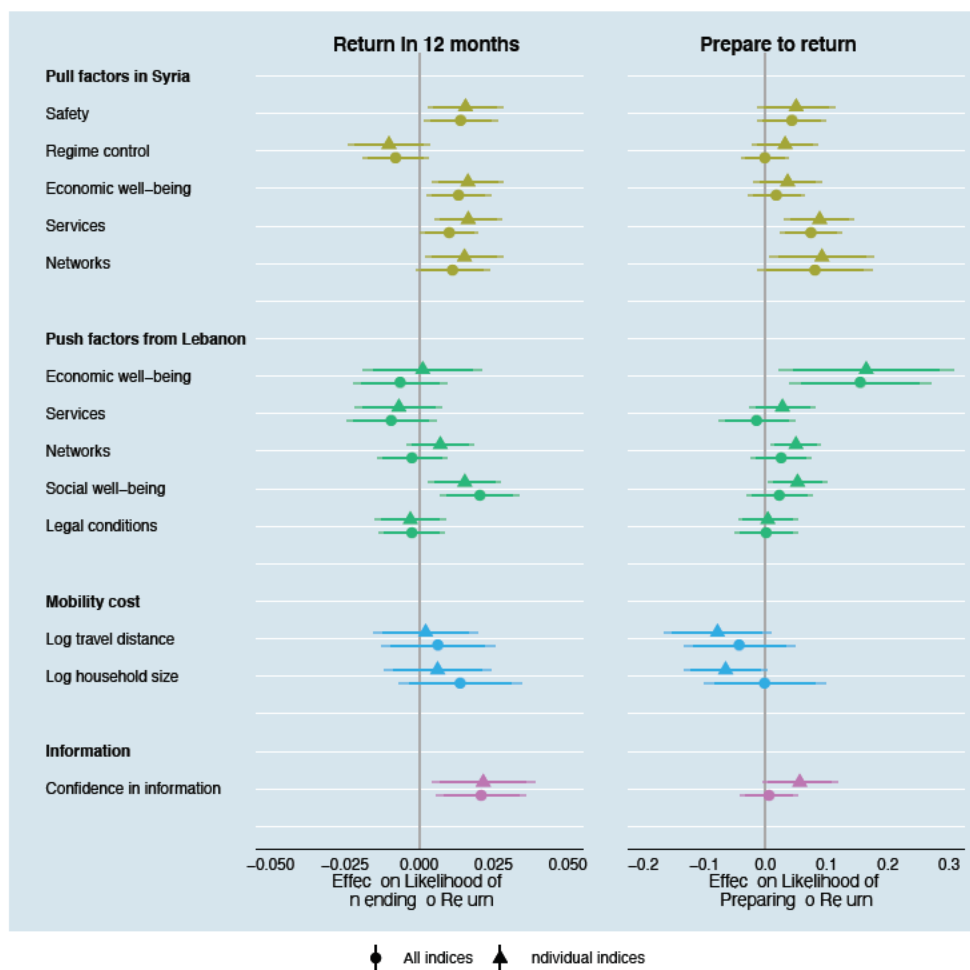
Figure 4: Index Results—Return Intentions and Preparations

---

[9]Figure 4 involves two deviations from the PAP, discussed in detail in Appendix Section 6. First, we present results with the pre-specified "All indices" regression with a minor adjustment to avoid multicollinearity. Safety in Syria and services in Syria exhibit high correlation. Therefore, we estimate a regression that includes all indices except services, and run a separate regression that includes all indices except safety. Other coefficients are drawn from the first regression, and Appendix Section 7.1 shows that results are substantively identical when we do the opposite. Second, the "Individual indices" regression was not pre-specified. We added this analysis to check for robustness to the possible threat of multicollinearity. See Appendix Sections 6.1 and 7.7 for discussion and tests of multicollinearity.

We present results for the drivers of return intentions in Figure 4. Each dot represents the point estimate for the relationship between a given index, labeled on the y axis, and a metric of return, labeled at the top of each panel. Circles represent point estimates drawn from our main model in Equation 1, and triangles represent point estimates drawn from models with each respective index in one model. The independent variables are grouped into four categories: people's prospective living situation in Syria, people's living situation in Lebanon, mobility costs to return to Syria, and people's confidence in the information they possess about Syria. The horizontal line around each point estimate shows the 90% and 95% confidence intervals (dark and light, respectively). Standard errors are clustered at the locality level, following from the sampling strategy. Indices are normalized to have mean zero and standard deviation one, and the point estimates present the change in the probability of return intentions that corresponds to a one standard deviation shift in an index. As shown in Appendix Section 7.3, results are robust to using additive indices rather than PCA indices.

Looking at Figure 4, we see strong evidence for a relationship between conditions in Syria and intentions to return within 12 months (first panel). We see that safety in Syria, economic prospects in Syria, the availability of public services in one's hometown, and respondents' family and friend networks in Syria are positively and significantly associated with return. For each of these indices, we see that a one standard deviation shift in the index corresponds with about a 2 percentage point increase in return intentions. In light of the small fraction of refugees (only 5%) who plan to return in the next year, this constitutes a large increase in return intentions in percentage terms (roughly 40%). Control by the Syrian government correlates negatively with intentions to return, although we cannot rule out a null relationship at either the 90% or 95% level.

The relationship between conditions in Syria and preparations for return (second panel) is less clear. Point estimates are consistently positive, but only the availability of services and the size of social networks are statistically significant. Security in Syria, regime control and economic prospects have no detectable relationship with preparations to return.

The results on push factors in Lebanon are quite different. First, looking at the left panel, we do not find a clear correlation between well-being in Lebanon and return intentions. We cannot rule out a zero association for most of the indices. The one index that demonstrates a statistically significant association with return intentions is social well-being. In contrast to the lack of evidence for a role of push factors in shaping return *intentions*, the second panel

reveals evidence for an association between conditions in Lebanon and return *preparations*. We find that higher levels of economic well-being, networks, and social well-being in Lebanon exhibit a detectable *positive* correlation with having taken steps to prepare to return to Syria in at least one specification. The direction of the relationship is not what we expected ex ante, based on a theory of preparations driven by a simple utility comparison between conditions in Lebanon and prospects in Syria. The finding highlights that our theory's focus on migration costs and incentives may have overlooked migration capacities. Indeed, return is a complex and daunting process, and people with more financial and social resources may be better able to undertake a safe, voluntary return.

Looking now at the next group of drivers, we see in the first panel that the results do not provide evidence of a relationship between mobility costs and return intentions. In the second panel, we find a negative association between mobility costs and preparations for return, significant at the 10% level, in the "individual indices" specification. Looking at the bottom row of Figure 4, we see that confidence in information about one's hometown is positively associated with both intentions and preparations. Information access have may both a direct effect on return intentions and a moderating role, a possibility we formally test in the next section.

Before concluding that pull factors are more powerful predictors than push factors in shaping return intentions, we explore two additional tests. First, we fit predictive models based on push factors and pull factors using 10-fold cross validation. We present the results in Appendix Section 7.6, and find that predictive models based on pull factors consistently demonstrate higher predictive power than models based on push factors. Second, we test whether Syrians' conditions in Lebanon exhibit identifying variation at both the national and local levels. If Syrians' conditions in the country were fairly homogeneous, then a null relationship between push factors in Lebanon and return intentions would be substantively trivial. Reassuringly, the data are not consistent with this concern. In Appendix Section 4.1, the interquartile ranges of the descriptive statistics demonstrate wide variation in the living conditions of Syrians in Lebanon. In Appendix Section 7.4, we re-run all models that controlled for locality level fixed effects, but without this control variable included. Our findings are robust to this alternative specification, suggesting that our null findings for the role of push factors in Lebanon are not driven by a lack of identifying variation in living conditions within localities.

5.2   Information and Return Intentions

To further explore the relationship between information and return, we examine whether information moderates the role of perceived conditions in Syria. Specifically, we examine whether conditions in Syria have a larger effect on people's intentions when they have high levels of confidence in their information about the situation in Syria.

$$Y_i = \alpha + \beta_1 T_i + \beta_2 \left(T_i \times \mathbb{1}(I_i > 0) \right) + \gamma X_i + \epsilon_i \tag{2}$$

Equation 2 is similar to the "individual indices" specification of Equation 1, but includes a multiplicative interaction term between each index T and information confidence. The indicator function, $\mathbb{1}(I_i > 0)$, denotes whether a respondent i had an index value for information confidence above the mean. X denotes the same vector of covariates as in Equation 1. Figure 5 presents regression results, displaying the estimated marginal effect of a one standard deviation change in each index for people with low (below-mean) confidence in information compared to high (above-mean) confidence in information.

The results in Figure 5 suggest that the relationship between conditions in Syria and return intentions and preparation is shaped by respondents' confidence in their information sources for some key factors. Specifically, we find evidence that information is a significant moderator for regime control, economic prospects, services, and networks. To be more precise, we find evidence that both regime control and economic well-being have a differential relationship with return intentions depending on respondents' information confidence. Next, we see a differential relationship between the availability of services in people's hometowns and both return intentions and return preparations. Last, we see a differential relationship between networks in Syria and return preparations depending on information confidence.

5.3   Long-term Return Intentions

In most refugee crises, some subset of the population will face protracted displacement, continuing to live as refugees for years, decades, and even generations. In some cases, households may face continued constraints that prevent their return. In others, after a prolonged period of displacement, people may simply decide to permanently settle where they are. Our data position us to consider not only near-term plans to return, but also long-term intentions by exploring the characteristics of people in our data who report that they intend to never return to Syria.
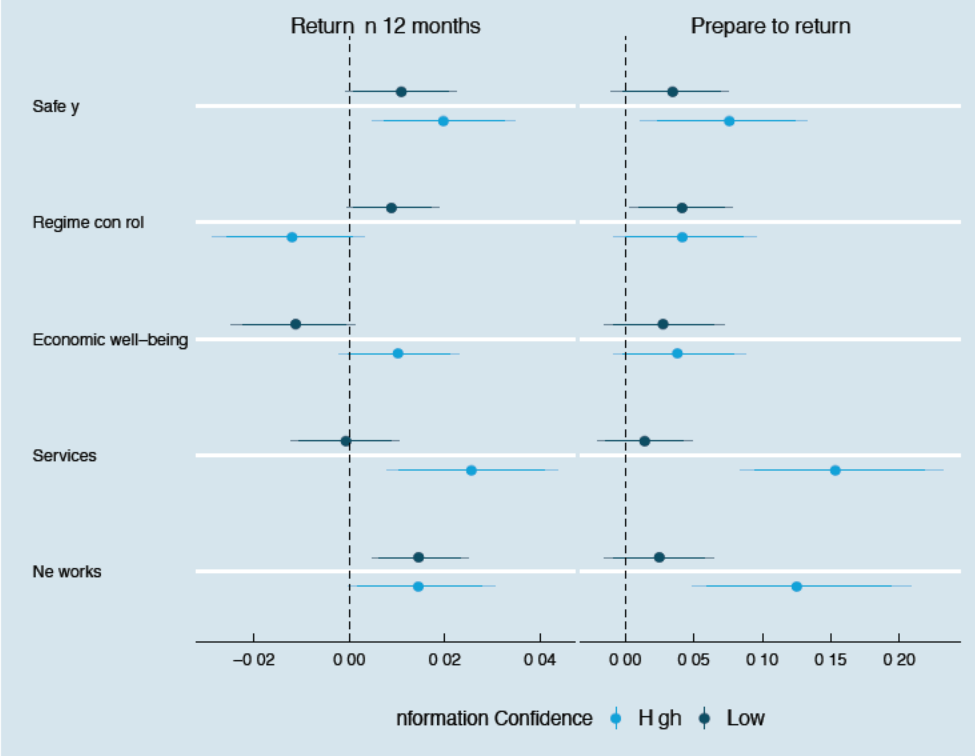
Figure 5: Interactive Effects of Information Confidence on Intentions and Preparations. Estimated marginal effects are presented with the corresponding 95% confidence intervals, with standard errors clustered by locality in Lebanon.

Figure 6 presents the same regression analyses defined in Equation 1 but with intentions to ever return as the outcome. The results offer evidence about the predictors of protracted displacement. We saw in Figure 3 that approximately 40% of respondents stated that they do not intend to ever return, suggesting the possibility that hundreds of thousands of Syrians might be displaced for the long term in Lebanon. The results suggest that the trends identified for 12-month return intentions and return preparations shown in Figure 4 are generally consistent, although more nuanced, when we consider intentions to ever return.

Aligning with our main results, we find evidence of a consistent relationship between conditions in the home country and life-long return intentions. We see statistically significant relationships with security, economic well-being, and networks. Public services in Syria exhibits a positive point estimate similar in magnitude to the other positive predictors, and the estimate is significant at the 90% level in one specification. We do not find evidence of a relationship between regime control and people's long-term intentions. In short, those who
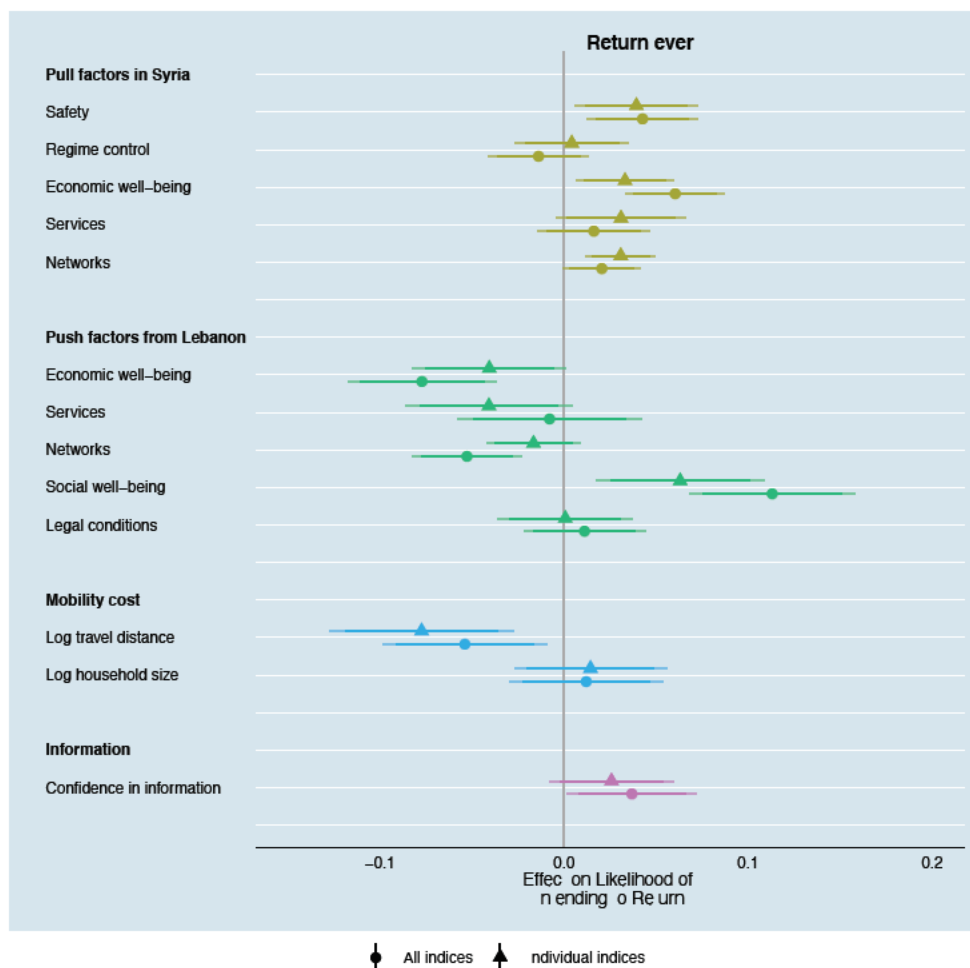
Figure 6: Index Results—Long-term return intentions

perceive better circumstances in Syria are more likely to indicate a desire to ever return, whereas those who do not are more likely to indicate that they will never go back to Syria.[10]

The results do not offer evidence of a straightforward relationship between current well-being in Lebanon and people's intentions to ever return—suggesting a complex relationship between well-being in the host country and protracted displacement. Some dimensions of well-being in Lebanon correlate positively, others negatively, with long-term return intentions. Specifically, the evidence on the role of host country well-being suggests that people in better economic situations, with broader personal networks, and with more access to ser-

---

[10]We also analyzed predictors of people's two-year return intentions, as pre-specified. Point estimates demonstrate a similar pattern although confidence intervals are wider than the one-year and lifelong return intentions analyses. These results are presented in Appendix Section 7.2.1.

vices are *less* likely to ever return, and people with higher levels of social well-being and integration are *more* likely to plan to ever return.

Looking at the third category, we see similar weak evidence of a negative relationship between mobility costs and refugees' long-term return intentions. In contrast to the 12-month results but similar to return preparations, we find a detectable negative association for travel distance. Conditional on our battery of control variables, people from more distant parts of Syria have lower long-term return intentions.

The bottom row presents the estimated relationship between confidence in information and intentions to ever return, which is consistent with the results in Figure 4. Both the sign of the relationship is the same and the magnitude is similar to analysis of 12-month return intentions. We see that the point estimate is statistically significant at the 95% in the one specification.

6   Results: Conjoint Experiment

The analysis of observational data strongly suggests that pull factors are more predictive of return intentions than push factors. Yet, our correlational estimates might be affected by other drivers not included in the model. In this section, we present the results of a conjoint experiment designed to provide greater leverage on the causal effects of these drivers on return intentions.

We follow the standard approach for analyzing conjoint experiments, using OLS regressions to estimate the Average Marginal Component Effect (AMCE) for each attribute (Hainmueller, Hopkins and Yamamoto, 2014). Figure 4 displays the effects for respondents' answers to the question: "Under these conditions, would you be willing to return to Syria?"

The results from the conjoint experiment are consistent with the analysis of observational data. Conditions in Syria play a more important role in shaping people's return intentions than conditions in Lebanon. Results suggest that safety is the most powerful driver of return, with security in one's hometown increasing return intentions by 35 percentage points and nationwide security increasing return intentions by 42 percentage points. The fact that safety in one's hometown has nearly as large of an effect as nationwide safety, suggests that the majority of variation in people's consideration in security is driven by conditions in their hometown, highlighting the local nature of security concerns in postwar environments. The availability of jobs and public services in Syria increases the likelihood of return by

up to 8 percentage points. An end to military conscription also plays an important role in shaping people's return intentions, and increases the likelihood of return by approximately 18 percentage points.

Both a good job and access to public services in Lebanon play small but statistically significant roles in people's return intentions. Someone with a good job in Lebanon is 2 percentage points less likely to return, and if someone has access to public services they are 3 percentage points less likely to return. Despite the statistical significance of these results, the differences in magnitudes between push and pull factors is striking.

In the bottom of the figure, we see the effect of network on people's responses. People were nearly 5 percentage points more likely to say that they would return to Syria if they have family and friends there (compared to having people outside of Syria and Lebanon). In contrast, we see a precisely estimated null effect for having family and friends in Lebanon on people's return intentions. These network results align with our earlier findings about the relative importance of the conditions in the home country compared to the hosting country.
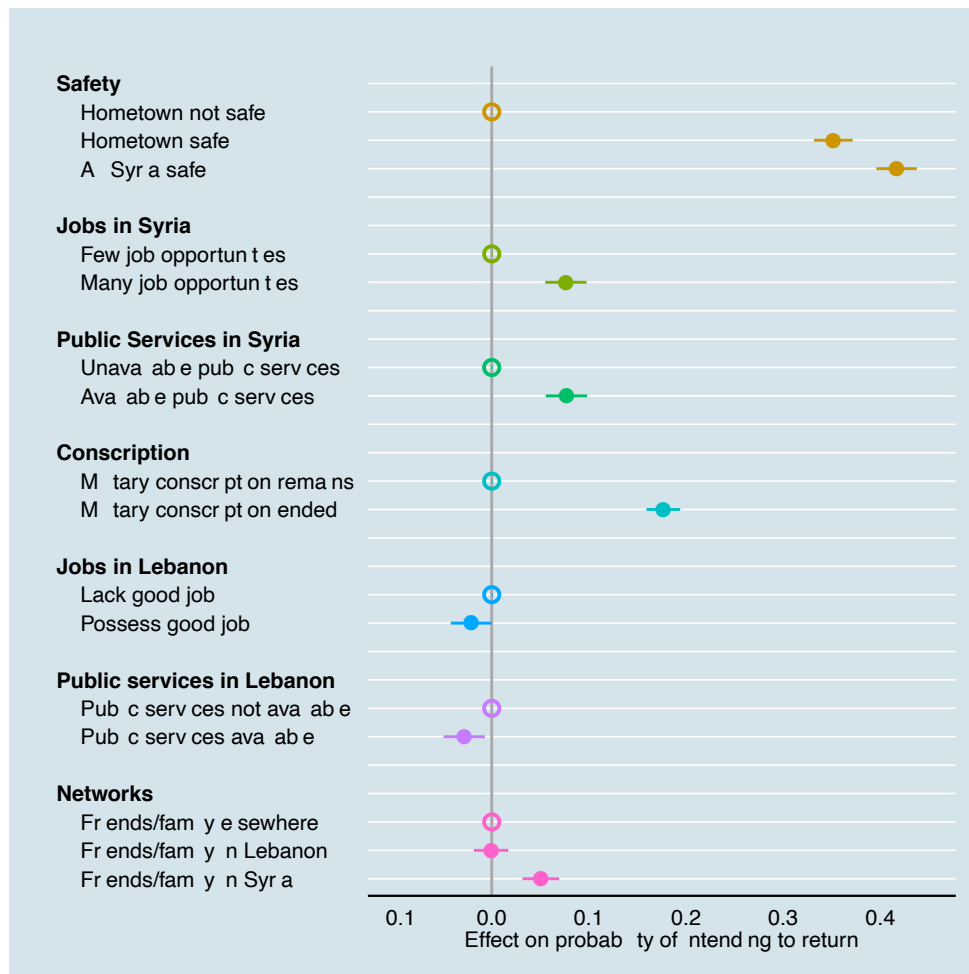
Figure 7: Conjoint Experiment Results. Each dot represents the effect on the probability that respondents would return to Syria in a given hypothetical situation, presented with its corresponding 95% confidence interval. The empty circles at $x = 0$ are reference categories. We cluster standard errors at the respondent level.

7   Beyond Lebanon: Return Intentions in Jordan

Given the magnitude of the Syrian refugee crises, refugee families are widely distributed, including in three primary hosting countries: Lebanon, Jordan, and Turkey. In order to ascertain whether our results are driven by unique circumstances among Syrians in Lebanon, we ran a separate survey with 1,286 Syrian refugees living in Jordan. This data offer a test of the external validity of our findings to the broader population of Syrian refugees. Our sampling strategy selected individuals from the four metropolitan areas in Jordan with the largest refugee populations: Amman, Irbid, Mafraq and Zarqa (including Azraq town). In the summer of 2019, enumerators interviewed a random sample of Syrians who received services from the NGO CARE during the study period. The participants were recruited from Syrian refugees living outside of camps, as do 83% of Syrians in Jordan (Brown et al., 2019).

The two cases make for a valuable comparison given some key similarities and critical differences. Similar to Lebanon, Jordan hosts a large number of Syrian refugees relative to its population, and public discourse in the country widely frames refugees as having large negative economic and fiscal impacts. In contrast, the baseline rate of return intentions for Syrians in our sample in Jordan is very low. When we asked Syrian refugees in Jordan if they plan to ever return to Syria, we find that a large majority of respondents (around 75%) reported that they never want to return to Syria. Further, unlike Lebanon, national political discourse in Jordan at the time of the survey was not pushing aggressively for Syrians to return. Therefore, this data enable us to examine whether our results from Lebanon pertain only to a context with major political pressure to return and where a large share of people hope to return home someday.

The difference in baseline return intentions between our samples in Lebanon and Jordan is likely driven by different selection into displacement to the countries. Similar to Lichtenheld (2020)'s theory of assortative displacement, we can imagine that refugees sort into host countries based on a range of personal characteristics, including relationship to a war's armed groups in the country of origin, and these views may influence the return decision. First, we see a difference between the two samples in self-reported level of security in respondents' hometowns. As of summer 2019, 51% of the sample in Jordan said that their place of origin continues to be very dangerous. In contrast, only 28% of respondents in Lebanon said so, when we conducted our survey there a few months later in September–October 2019. Second, our fieldwork suggests that the political attitudes of Syrians living in Jordan tend to be more

anti-regime whereas the Syrian population in Lebanon is more divided in its views toward the Syrian government, which aligns with public opinion surveys on the topic (Corstange, 2018).

Using our data from Jordan, we construct indices for dimensions of people's well-being in Jordan and prospective well-being in Syria. We then regress return intentions on the indices, as defined in the "individual indices" specification of Equation 1, to estimate the impact of each factor on peoples' stated plans to ever return to Syria.[11]

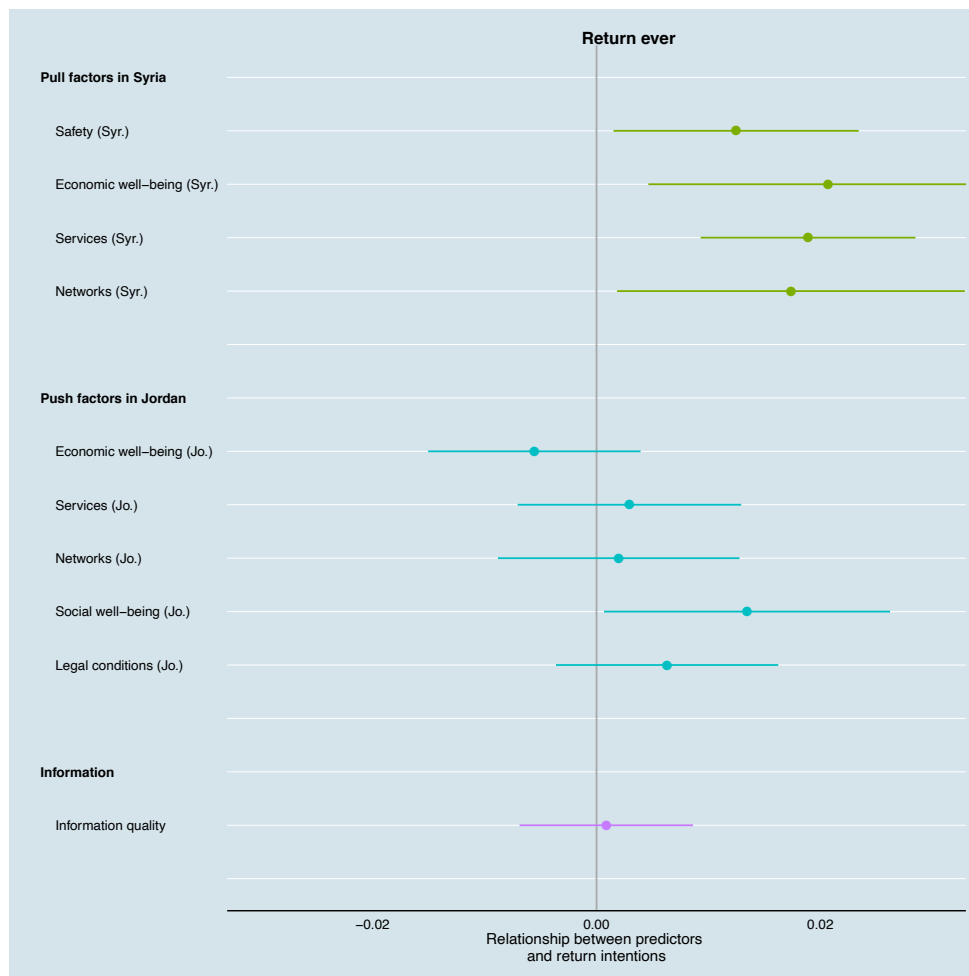Figure 8: Index Results in Jordan

Figure 8 presents results from analysis of the Jordan data. Despite the sizeable difference in baseline return intentions and the political climate, the drivers of return intentions in Jordan are strikingly similar to Lebanon. First, prospective conditions in Syria play an

---

[11]The list of questions used in each index are included in Section 9 of the Appendix

important role. We see that conditions in respondents' place of origin in Syria—specifically safety, economic prospects, and public services—are positively correlated with return intentions. Also, having family and friend networks in Syria is positively correlated with return intentions.

Second, in line with results from Lebanon, we do not find strong evidence that conditions in Jordan drive return intentions. First, we see in Figure 8 that economic conditions, access to public services, personal networks, and legal conditions are not strongly associated with return intentions. In contrast, social well-being is the one dimension of conditions in Jordan where we detect a relationship with return intentions. This aligns with findings in Lebanon, where we saw that people with higher social well-being were more likely to intend to return within 12 months, to have prepared to return, and to intend to return ever.

Finally, looking at the impact of information, we do not find evidence of a relationship between information and return intentions in Jordan. This contrasts with the evidence we found in Lebanon for the importance of information for Syrians' decision making about return.[12]

8 Discussion and Conclusion

We develop a framework for understanding the dynamics of refugee return by adapting canonical models of individual migration, viewing them through the lens of refugee household decision-making. We inform our theory-building with extended interviews and focus groups with refugees and humanitarian organizations in the context of forced displacement. Based on this framework, we identify four major drivers of return: push factors in the host country, pull factors in the home country, the cost of mobility, and the role of information. We test the relative importance and interactions of these drivers in the context of the Syrian refugee crisis in Lebanon (and Jordan) using observational and experimental survey data from a representative sample of refugees.

We find strong evidence that the return decisions of Syrian refugees are shaped primarily by the situation in the home country (pull factors), rather than the dynamics in the hosting country (push factors). The most important pull factors are perceptions of safety, economic opportunity, the availability of public services, and the size of personal networks. Strikingly,

---

[12]Unfortunately, we were not able to ask respondents in Jordan for the name of their hometown or district in Syria, which prevents us from analyzing the role of mobility cost.

we do not find evidence of a strong relationship between conditions in the hosting country and people's return intentions. If anything, people with better conditions in Lebanon were *more* likely to prepare to return—a finding that we did not anticipate, but which is consistent with the fact that moving internationally requires significant resources, especially under circumstances of instability and violence. We find evidence that the strength of pull factors is moderated by the confidence that respondents have in their information about the situation in Syria. Perceptions of the situation in Syria are more predictive of return intentions if respondents are confident in the quality of the information they have.

As we consider the broader import of these findings, it is useful to reflect on benefits and limits of a study conducted in Lebanon (and Jordan). On the one hand, Lebanon offers a number of features ideal for studying this question. The country exhibits wide variation in push and pull factors, and the country is also a most-likely case for identifying the role of push factors, given the severity of the hardship that many Syrians face in the country.

In contexts of longer-term displacement, the drivers of return may exhibit different effects on migration intentions. The median respondent in our data has been in Lebanon for 6 years, but many refugees worldwide have been displaced even longer. As displacement extends across generations, people who grew up in a host country may possess economic and social ties there, while holding fairly limited ties to the country that their parents or grandparents fled. Protracted displacement may also shape people's identity in ways that change the choice set they actively consider for migration. For example, inter-generational displacement may impact people's identity or the institutional arrangements governing migration, such that people simply no longer consider returning.

The dynamics of return may also differ for refugees living in Western countries, where about 15% of refugees worldwide reside. These refugees, most living far from their country of origin, fall outside the scope conditions of this study for three reasons. First, people who make the choice and have the opportunity to travel to distant countries likely differ in important ways from refugees who flee to neighboring countries. Second, forcibly displaced people who arrive in Western countries often face a very different legal environment than those in the Global South. Among people who have been resettled or granted asylum in Western countries, we might expect that legal security and the well-being it facilitates would lead to lower rates of return. We cannot test this because neither Lebanon nor Jordan grants asylum to Syrians. Third, our data exhibits wide variation in travel distance from Lebanon to Syria, but it is worth noting that much longer trips, such as across continents, would be

necessary for some refugees in other contexts to return home. Mobility costs may play a stronger role in shaping migration choices once the distances extend across the globe.

Our findings highlight a number of important questions for a growing research agenda on refugee crises and the dynamics of return. First, our results underscore the importance of further research on the role of well-being in the host country in refugee migration. The null relationship we find between host-country conditions and return intentions could emerge if well-being in the host country has countervailing negative and positive impacts on people's willingness to return and ability to do so. People with greater resources while living as refugees may be both better off in exile but also more able to afford the costs of moving back home. Our results provide evidence for this possibility, where we see that better conditions have very small or null effects on people's intentions to return, but a positive association with concrete steps to prepare to return. More rigorous tests of the mechanisms linking host country well-being and return remain on the agenda for future research.

Second, very little is understood about how refugees acquire and assess information about the situation in their home country. It is intuitive that high-quality information will condition migration choices, especially given the potential negative consequences of returning prematurely to a difficult context (e.g., military conscription, arrest or kidnapping, torture, or death). While refugees cannot predict the future in their host country with certainty, they can at least seek to understand current conditions. But the uncertainty that they have about the situation at home may lead them to underweight outcomes in the home country relative to those in the host country (Kahneman and Tversky, 1979). Uncertainty, and therefore underweighting, is reduced when people have access to better information about the situation in the home country.

Third, future research should explore the degree to which return intentions predict people's subsequent migration choices. Existing work on labor migration may provide a baseline expectation. Worldwide, Tjaden, Auer and Laczko (2019) use data from six consecutive years in the Gallup World Poll (GWP) from 2010 to 2015. They find that a 1 percent change in emigration intentions leads to an approximately 0.8 percent change in actual out-migration. In another study, looking at migration to OECD countries, Docquier, Peri and Ruyssen (2014) show that emigration intentions correlate highly with migration flows, but the relationship is heterogeneous based on education level. People with a college degree are much more able to realize a desire to migrate to an OECD country. However, these existing studies focus on economic migration and not forced displacement. In the latter context, there may be a gap

between intentions and choices because of legal and documentation challenges.

Finally, refugees' decision-making about return is a product of both individual-level factors and aggregate shocks (e.g., economic crisis or civil war in the hosting country). In considering the role of push factors, it will be important to examine how aggregate shocks condition household decision-making. Future work would also do well to explore the interaction of shocks and household characteristics to learn about differential impacts. A limitation of the evidence presented in this paper, given the cross-sectional nature of the data, is that the variation in host country conditions does not include such shocks.

This research provides two key takeaways for policymakers and humanitarian organizations. First, the results reaffirm the fundamental humanitarian mandate of the refugee protection regime. Refugees are not economic migrants by another name. They intend to return to their home country when threats to their physical, economic, and social well-being have decreased, and when they feel that they possess credible information about these changes. Even after years in a host country, people's migration choices do not appear to be driven by opportunities in the hosting country. We find that more than two thirds of Syrians in Lebanon want to return home, and prospects for a good job and access to public services do not influence people's likelihood of staying. The evidence is clear that policies that deny rights to refugees or broader anti-refugee hostility are unlikely to drive people to return.

The findings also offer lessons for how humanitarian agencies can support refugee well-being while also promoting safe, voluntary return. Given the strong desire of refugees to return home, efforts to deliver humanitarian assistance and provide economic opportunities are unlikely to incentivize refugees to remain in the host country. More traditional development programs that support refugees' economic integration could benefit both refugees and host-country economies and free refugees from a reliance on aid. This would, in turn, allow humanitarian agencies to focus their attention and resources on emergencies, rather than struggling to continue providing assistance in protracted displacement situations. In addition, as policymakers focus on conflict resolution in refugee-sending countries, our findings highlight the importance of access to information. To end refugee crises, credible information dissemination is critical, as refugees need a good understanding of the conditions at home before they are willing to consider return.

References

Adhikari, Prakash. 2013. "Conflict-Induced Displacement, Understanding the Causes of Flight." *American Journal of Political Science* 57(1):82–89.

Adida, Claire L. 2014. *Immigrant exclusion and insecurity in Africa.* Cambridge University Press.

Alrababa'h, Ala', Andrea Dillon, Scott Williamson, Jens Hainmueller, Dominik Hangartner and Jeremy Weinstein. 2020. "Attitudes toward migrants in a highly impacted economy: evidence from the Syrian refugee crisis in Jordan." *Comparative Political Studies* .

Bahar, Dany, Andreas Hauptmann, Cem Özgüzel and Hillel Rapoport. 2018. "Let their Knowledge Flow: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia.".

Batista, Cátia and Francesco Cestari. 2016. "Migrant Intentions to Return: The Role of Migrant Social Networks." (1602).

Bazzi, Samuel. 2017. "Wealth heterogeneity and the income elasticity of migration." *American Economic Journal: Applied Economics* 9(2):219–255.

Blattman, Christopher, Alexandra Hartman and Robert Blair. 2014. "How to Promote Order and Property Rights under Weak Rule of Law? An Experiment in Changing Dispute Resolution Behavior through Community Education." *American Political Science Review* 108(01):1–21.

Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 4(77):531–55.

Braithwaite, Alex, Tiffany S Chu, Justin Curtis and Faten Ghosn. 2019. "Violence and the perception of risk associated with hosting refugees." *Public Choice* 178(3-4):473–492.

Brown, Harry, Nicola Giordano, Charles Maughan and Alix Wadeson. 2019. "Vulnerability Assessment Framework." *UNHCR* .

Camarena, Kara Ross. 2018. "Refugee Return and Food Insecurity: Explaining Increases in Violence in Ongoing Civil Wars." *Working paper* .

Camarena, Kara Ross and Nils Hagerdal. 2020. "When Do Displaced Persons Return? Postwar Migration among Christians in Mount Lebanon." *American Journal of Political Science* .

Corstange, Daniel. 2018. "The Syrian Conflict and Public Opinion Among Syrians in Lebanon." *British Journal of Middle Eastern Studies* .

Crisp, Jeff. 2019. "Repatriation principles under pressure." *Forced Migration Review* (62):19–23.

Croese, Sylvia. 2017. "State-led housing delivery as an instrument of developmental patrimonialism: The case of post-war Angola." *African Affairs* 116(462):80–100.

Dancygier, Rafaela M. 2010. *Immigration and conflict in Europe.* Cambridge University Press.

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

Dinas, Elias, Konstantinos Matakos, Dimitrios Xefteris and Dominik Hangartner. 2019. "Waking Up the Golden Dawn: Does Exposure to the Refugee Crisis Increase Support for Extreme-Right Parties?" *Political Analysis* pp. 1–11.

Docquier, Frédéric, Giovanni Peri and Ilse Ruyssen. 2014. "The cross-country determinants of potential and actual migration." *International Migration Review* 48(1):37–99.

Dustmann, Christian and Yoram Weiss. 2007. "Return migration: theory and empirical evidence from the UK." *British Journal of Industrial Relations* 45(2):236–256.

Fabbe, Kristin, Chad Hazlett and Tolga Sinmazdemir. 2019. "A persuasive peace: Syrian refugees' attitudes towards compromise and civil war termination." *Journal of Peace Research* 56(1):103–117.

Fitzgerald, Jennifer, David Leblang and Jessica C Teets. 2014. "Defying the law of gravity: The political economy of international migration." *World Politics* 66(3):406–445.

Ghosn, Faten, Alex Braithwaite and Tiffany S Chu. 2019. "Violence, displacement, contact, and attitudes toward hosting refugees." *Journal of peace research* 56(1):118–133.

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

Goldstein, Judith L. and Margaret E. Peters. 2014. "Nativism or Economic Threat: Attitudes Toward Immigrants During the Great Recession." *International Interactions* 40(3):376–401.

Hainmueller, J., D.J. Hopkins and T. Yamamoto. 2014. "Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments." *Political Analysis* 22(1):1–30.

Hainmueller, Jens and Daniel J. Hopkins. 2014. "Public Attitudes Toward Immigration." *Annual Review of Political Science* 17(1):225–249.

Hainmueller, Jens and Michael J Hiscox. 2010. "Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment." *American Political Science Review* 104(01):61–84.

Hangartner, Dominik, Elias Dinas, Moritz Marbach, Konstantinos Matakos and Dimitrios Xefteris. 2019. "Does exposure to the refugee crisis make natives more hostile?" *American Political Science Review* 113(2):442–455.

Harder, Niklas, Lucila Figueroa, Rachel M Gillum, Dominik Hangartner, David D Laitin and Jens Hainmueller. 2018. "Multidimensional measure of immigrant integration." *Proceedings of the National Academy of Sciences* 115(45):11483–11488.

Heckman, James J and Bo E Honore. 1990. "The empirical content of the Roy model." *Econometrica: Journal of the Econometric Society* pp. 1121–1149.

Holland, Alisha C and Margaret E. Peters. 2020. "Explaining Migration Timing: Political Information and Opportunities." *International Organization* .

Hopkins, D J. 2010. "Politicized places: Explaining where and when immigrants provoke local opposition." *The American Political Science Review* .

Hunt, Gary L and Richard E Mueller. 2004. "North American migration: returns to skill, border effects, and mobility costs." *Review of Economics and Statistics* 86(4):988–1007.

Jallow, Amadou Tijan, C. Heinbecker and S Malik. 2004. *Handbook for repatriation and reintegration activities*. UNHCR.

Janmyr, Maja. 2016. "Precarity in Exile: The Legal Status of Syrian Refugees in Lebanon." *Refugee Survey Quarterly* 35(4):58–78.

Kahneman, D and A Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica* 47(2):263–291.

Lazarev, Egor and Kunaal Sharma. 2017. "Brother or burden: An experiment on reducing prejudice toward Syrian refugees in Turkey." *Political Science Research and Methods* 5(2):201–219.

Lehmann, M Christian and Daniel Masterson. 2020. "Do Cash Transfers Reduce Anti-Refugee Violence? Evidence from Syrian Refugees in Lebanon." *American Political Science Review* 114(4):1335–1342.

Lichtenheld, Adam G. 2020. "Explaining Population Displacement Strategies in Civil Wars: A Cross-National Analysis." *International Organization* 74(2):253–294.

Lischer, Sarah Kenyon. 2006. *Dangerous Sanctuaries: Refugee Camps, Civil War, and the Dilemmas of Humanitarian Aid*. Cornell University Press.

Malhotra, Neil, Yotam Margalit and Cecilia Hyunjung Mo. 2013. "Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact." *American Journal of Political Science* 57(2).

Massey, Douglas S, Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino and J Edward Taylor. 1993. "Theories of international migration: A review and appraisal." *Population and development review* pp. 431–466.

McNamee, Lachlan. 2018. "Mass Resettlement and Political Violence: Evidence from Rwanda." *World Politics* 70(4):595–644.

Metivier, Sean, Djordje Stefanovic and Neophytos Loizides. 2018. "Struggling for and Within the Community: What Leads Bosnian Forced Migrants to Desire Community Return?" *Ethnopolitics* 17(2):147–164.

Mhaissen, Rouba and Elena Hodges. 2019. "Unpacking Return: Syrian Refugees' Conditions and Concerns." *SAWA for Development and Aid* .

Mourad, Lama. 2017. ""Standoffish" Policy-making: Inaction and Change in the Lebanese Response to the Syrian Displacement Crisis." *Middle East Law and Governance* 9(3):249–266.

Munshi, Kaivan. 2003. "Networks in the modern economy: Mexican migrants in the US labor market." *The Quarterly Journal of Economics* 118(2):549–599.

Pettersson, Therese. 2019. "UCDP/PRIO Armed Conflict Dataset Codebook v 19.1.".
URL: *https://ucdp.uu.se/downloads/*

Rossi, Peter Henry. 1980. *Why families move.* Sage Publications, Inc.

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers* 3(2):135–146.

Ruiz, Isabel and Carlos Vargas-Silva. 2020. "The impacts of refugee repatriation on receiving communities." *Journal of Economic Geography* .

Salehyan, Idean and Kristian Skrede Gleditsch. 2006. "Refugees and the Spread of Civil War." *International Organization* 60(02):335–366.

Scheve, Kenneth F. and Matthew J. Slaughter. 2001. "Labor Market Competition and Individual Preferences Over Immigration Policy." *Review of Economics and Statistics* 83(1):133–145.

Schon, Justin. 2019. "Motivation and opportunity for conflict-induced migration: An analysis of Syrian migration timing." *Journal of Peace Research* 56(1):12–27.

Schwartz, Stephanie. 2019. "Home, Again: Refugee Return and Post-Conflict Violence in Burundi." *International Security* 44(2):110–145.

Sides, John and Jack Citrin. 2007. "European opinion about immigration: The role of identities, interests and information." *British journal of political science* 37(3):477–504.

Stein, Barry N and Frederick C Cuny. 1994. "Refugee repatriation during conflict: Protection and post-return assistance." *Development in Practice* 4(3):173–187.

Stork, Joe and Fred Abrahams. 2004. "Sidelined: Human Rights in Postwar Iraq." *Human Rights Watch World Report 2004* .

The World Bank. 2017. "Syrian Refugees and Their Hosts in Jordan, Lebanon, and the Kurdistan Region of Iraq: Lives, Livelihoods, and Local Impacts." *Unpublished Report* .

Tjaden, Jasper, Daniel Auer and Frank Laczko. 2019. "Linking migration intentions with flows: Evidence and potential use." *International Migration* 57(1):36–57.

UNHCR. 2004. "EC/54/SC/C: Protracted Refugee Situations." *Executive Committee of the High Commissioner's Programme* .

UNHCR. 2008. "Implementation Strategy: UNHCR's Role in Support of the Return and Reintegration of Displaced Populations.".

UNHCR. 2019. "Global Trends Forced Displacement in 2018.".

UNHCR Refugee Population Statistics Database. 2019.
     URL: *https://www.unhcr.org/refugee-statistics/.*

Van Leeuwen, Mathijs and Gemma Van Der Haar. 2016. "Theorizing the land–violent conflict nexus." *World Development* 78:94–104.

Yahya, Maha. 2018. "Unheard Voices: What Syrian Refugees Need to Return Home." *Carnegie Middle East Center, and Carnegie Endowment for International Peace* .

Zhou, Yang-Yang and Andrew Shaver. 2019. "Reexamining the Effect of Refugees on Civil Conflict: A Global Subnational Analysis." *Working Paper* .

Appendix

The Dynamics of Refugee Return:
Syrian Refugees and Their Migration Intentions[*]

Ala' Alrababa'h[1,2], Daniel Masterson[2,3], Marine Casalis[2],

Dominik Hangartner[2,4,5], Jeremy Weinstein[1,2]

[1]Department of Political Science, Stanford University

[2]Immigration Policy Lab, Stanford University and ETH Zurich

[3]Department of Political Science, University of California, Santa Barbara

[4]Center for International and Comparative Studies, ETH Zurich

[5]Department of Government, London School of Economics and Political Science

November 9, 2020

1

Contents

1   Sampling Strategy

We conducted stratified multistage sampling. In the first stage, we selected localities based on two dimensions: the prevalence of Syrians and the majority sect. We drew Syrian population data from UNHCR registration records and Lebanese population data from voter registration records, since there is no current census available. In the second stage, we used a random walk procedure to select households within each locality. In the third stage, we selected a head of household   defined as a person regardless of gender of who takes part in decisions for the household. We offer a detailed discussion below.

1.1   Definitions

Household: a group of persons who eat and live together, and are related by blood or marriage.

Head of household: head of household   a person with primary decision-making responsibility in the household, who knows the details of the house, its economic situation, and reasons for important household decisions.

Dwelling: A building or residential unit. It may include one or more households, as in the case of compounds or apartment buildings and may include tents or informal settlements.

Mixed-households: households including both Syrians and host community, either where a Syrian is married to a member of the host community, or where Syrians are living in a host community household and eating out of the same pot.

Syrian prevalence: describes the number of Syrians living in a specific area in re-

lation to the total population (Syrians and non-Syrians) living in that specific area. This is calculated as: Number of Syrians in the area / total population (Syrians + non-Syrian) in that area.

Syrian density: describes the number of Syrians living in one specific sub-area (e.g. neighbourhood) in relation to the total number of Syrians living in the whole area to be surveyed (e.g. city). This is calculated as: Number of Syrians in one specific area / Number of Syrians in the whole area.

Majority sect: describes the largest sectarian group according to Lebanese voter registration data.

Mixed area: describes an area (e.g., village) where no single sectarian group constitutes more than 50% of the village population.

1.2 Defining the Population of Interest

Sampling refugees requires a clear definition of the 'refugee' population of interest. As described in UNHCR's sampling guide for non-camp settings there are three options for how to define refugees for the purpose of sampling:

1. Registered refugees, including only those who have gone through registration;

2. Registered refugees and asylum seekers, including only those who have gone through registration or have asylum-seeker status; or

3. Including all individuals from the same country of origin, regardless of their status.

These distinctions are relevant because not all individuals from the a given country of origin will seek refugee status. Also there are a range of residency and visa types that Syrians use when living in Lebanon.

For this study we use the third definition, guided by the aim of the survey. We seek to draw inferences about all individuals from Syria in Lebanon as of the study's start date, regardless of their legal status and access to services and support. This study includes Syrian citizens, regardless of whether they registered with UNHCR, and also Palestinian refugee from Syria (PRS), regardless of whether they registered with UNRWA.

We conducted a stratified sample of localities and random walk recruitment of households. This incorporated an approximate map and list of villages with approximate number of Syrians and Syrian families in each villages, or at least the relative distribution of Syrians in the various areas   i.e., the size of the Syrian population in different areas proportionate to each other (e.g. one area may have about 3 times refugees than the other, etc.) and the approximate prevalence of Syrians relative to Lebanese population (Syrians and non-Syrians) in the various areas.

1.3   First sampling stage: Locality selection

The sampling frame for the first stage is the list of localities published by the Lebanese Council for Development and Reconstruction (CDR) in 2018 and the 2018 UNHCR data on the number of registered Syrians by locality. Each locality is identified by way of its administrative affiliation   Kaza and Mohafza   this is important to avoid confusing towns in different parts of the country with similar or identical names. The database reports the total population in each village, as well as the number of Lebanese, Palestinian, and Syrian population in each.

The localities were sorted into nine strata depending on their prevalence of Syrian population and the majority Lebanese sectarian group, as follows:

- Dimension 1: Prevalence of Syrian refugees

– Low prevalence: where the Syrian population accounted for less than 20% of the total population.

– Medium prevalence: where the Syrian population is between 20% and 50% of the total population.

– High prevalence: where the Syrian population accounted for over 50% of the total population.

- Dimension 2: Sectarian component of Lebanese population

  – Sunni Lebanese majority: More than 50% of the documented Lebanese population of the village is Sunni.

  – Non-Sunni Lebanese majority: More than 50% of the documented Lebanese population of the village is of a single non-Sunni sectarian group.

  – Mixed: No single sectarian group makes up more than 50% of the village population.

We used three data sources for first-stage sampling:

1. UNHCR Syrian registration data

2. Lebanese voter registration data

3. UNRWA registered Palestinian population data in 12 UNRWA camps

In the first stage sample we selected 120 Lebanese localities and then randomly sampled people in each. Our goal is to say something about the drivers of return in the overall population. Hence, in our main analysis, we cluster standard errors by locality, since there are localities in the population of interest beyond those captured in the sample (Abadie et al., 2017).

1.3.1 Syrian population data – UNHCR

Roughly 3/4 Syrians in Lebanon are registered with UNHCR. There are no locality-level population estimates of Syrian population. UNHCR data is presented at the town/village/urban neighborhood level, recorded as of December 2018.

This data only includes areas with greater than 0 Syrians. This is not a problem for our sampling since we would not want to sample towns with a Syrian population of 0.

1.3.2 Lebanese population data – voter registration records

There is no Lebanese census. One commonly used data source to estimate locality-level Lebanese population is voter records, published by the Ministry of Interior and Municipalities. The shortcoming of this metric is that Lebanese do not need to live in their voter registered district and often do not. We do not have estimates of how many do not. The unit of analysis for voter registration data is the electoral district, which is sometimes the same as the town/village, but not always. There is no efficient way to confirm that towns in the UNHCR data and the electoral district data perfectly overlap, but it is likely that the overlap is, on average, fairly close.

1.3.3 Joining the UNHCR and Lebanese data

We can describe our data as the list of villages provided in UNHCR maps (December 2018) that provide the number of Syrians in each town/locality and the corresponding number of Lebanese registered voters who are in that area as per the May 2018 data according to Lebanese voter registration records. The voter records include the number of Lebanese by sect.

1.3.4   UNRWA data

According to UNRWA estimates, about half of Lebanon's 450,000 Palestinians live in the country's 12 UNRWA camps. These UNRWA camp residents appear in UNRWA's Palestinian population statistics, whereas Palestinians outside of camps do not. (See Section 1.5 for further discussion of this point.) The UNRWA data does not include information on sect, but given that the vast majority of Palestinians in Lebanon are Sunni, we code all UNRWA Palestinians as Sunni for purposes of stratification. (See Section 1.5 for further discussion of this choice.)

1.4   Sampling Statistics

We divide our sampling frame into 9 stratification bins according to two dimensions:

Dimension 1:

1. Low prevalence: where the Syrian population accounted for less than 20% of the total population (Syrians + Lebanese + Palestinians).

2. Medium prevalence: where the Syrian population is between 20% and 50% of the total population.

3. High prevalence: where the Syrian population accounted for over 50% of the total population.

Dimension 2:

1. Sunni non-Syrian majority: More than 50% of the non-Syrian (Lebanese + Palestinian) population of the town is Sunni.

2. Non-Sunni non-Syrian majority: More than 50% of the non-Syrian population of the village is of a single non-Sunni sectarian group.

3. Mixed: No one sectarian group makes up more than 50% of the village's non-Syrian population.

Table 1 shows the number of Lebanese localities in each stratification bin.

	sunni maj	nonsunni maj	mixed
low prev	53	212	17
med prev	63	111	11
high prev	28	27	9

Table 1: Number of Localities per Bin

Table 2 shows the corresponding percentage of localities in each bin.

	sunni maj	nonsunni maj	mixed
low prev	0.10	0.40	0.03
med prev	0.12	0.21	0.02
high prev	0.05	0.05	0.02

Table 2: Percentage of Localities per Bin

Table 3 shows the registered Syrian population in each bin.

	sunni maj	nonsunni maj	mixed
low prev	59742	101563	11870
med prev	169708	144525	18332
high prev	229124	131640	27563

Table 3: Registered Syrian Population per Bin

	sunni maj	nonsunni maj	mixed
low prev	0.07	0.11	0.01
med prev	0.19	0.16	0.02
high prev	0.26	0.15	0.03

Table 4: Percentage of Registered Syrian Population per Bin

Table 4 shows the percentage of the registered Syrian population in each bin.

If we had sampled a number of localities per bin in direct proportion to the Syrian population of the strata we would have derived a nearly optimal sample for the country as a whole. This would have corresponded to sampling the number of localities per bin shown in Table 5.

	sunni maj	nonsunni maj	mixed
low prev	10	17	2
med prev	28	24	3
high prev	38	22	5

Table 5: Proportional First-Stage Sample Size

Alternatively, a stratification strategy can select the same number of localities in each stratum which would deliver estimates of nearly the same quality for each locality. That would have meant sampling $150/9 \approx 16.6$ localities in each bin. This would have produced a first-stage sample size across bins like in Table 6.

	sunni maj	nonsunni maj	mixed
low prev	16	16	16
med prev	17	17	17
high prev	17	17	17

Table 6: Equal First-Stage Sample Size

Given these two conflicting approaches and the advantage of each, the distribution of the sample of 150 localities into strata faced the classical dilemma of whether doing it in proportion to the population of the strata or rather selecting the same number of localities in each stratum. Since both are important considerations for our study, and we want to study subnational variation while also making nationwide claims, we followed a first-stage sampling distribution in accordance to Markward's rule (also known as the '50/50 equal/proportional allocation' rule), which is generally considered a good compromise between the two extremes. Given the small number of mixed localities in the pure PPS sample, the 50/50 equal/proportional allocation has an additional desirable feature of avoiding any bins with a very small number of localities. The final sample is shown in Table 7.

	sunni maj	nonsunni maj	mixed
low prev	13	17	9
med prev	23	20	10
high prev	28	19	11

Table 7: First-Stage Sample Size (50/50 Equal/Proportional)

1.5   Note on Including UNRWA Palestinian Population

We used UNRWA data for first-stage sampling, even though we lacked data for non-camp/non-UNRWA Palestinian population for a number of reasons. The Palestinian population data in 12 camps (and therefore in 12 localities out of 1,155 localities) can only impact our first-stage sampling if it shifts any of the 12 localities from one stratification bin to another. So a bias metric is not continuous (e.g., is the Syrian population prevalence correct?) but rather discrete (i.e., is a locality in the wrong bin?). Presumably the Palestinians outside camps are mostly in cities, with large Lebanese populations. And Palestinian populations in small localities are probably small in number. Palestinians in camps make up a large share of the non-Syrian population in those localities, but Palestinians outside camps do not make a large share of the population in those localities.

If we did not use Palestinian data we would dramatically mis-measure the non-Syrian population in camps, and could have localities with camps in the wrong bins. Outside of camps, Palestinians likely only change our estimates of Syrian prevalence or sect populations by 1 percentage point or so, which is unlikely to shift the locality to a different stratification bin. For example, take Mazraa and Bourj el Barajneh as examples of non-camp and camp settings. Mazraa has 7,800 Syrians and 93,000 Lebanese including 78,000 Sunnis. So that puts are prevalence in the low bin at 0.08 and the Sunni population share at 85%. No number of additional non-camp Palestinians could change the prevalence or sect bin for Mazraa. In contrast, Bourj El Barajneh has 18,000 Syrians, 17,600 Lebanese including 15,500 Shia, and UNRWA data includes 17,900 Palestinians, who we will assume are all Sunni. This means that not including the UNRWA data would put the locality in the wrong population prevalence bin and the wrong sect bin. That is, without UNRWA data, we would mistakenly calculate that Syrians make up more than 50% of the locality population, when they actually only make up 33%. And we would call the district majority Shia, when it is actually mixed. These two examples are not general but they demonstrate

that Palestinian population will matter a lot for putting localities in the right bins in localities with camps, and is unlikely to affect bin definition outside of camps.

1.6   Second-stage sampling

Data collectors will collect data from 20 households for each sample draw of a village, since sampling took place with replacement.

In order to capture all Syrians   not just registered refugees   we opted for a random walk strategy rather than sampling from UNHCR's registration database. The random walk strategy aimed to achieve a balance between what is practical in the field and defining a unified approach for all sites. In order to achieve these two goals, we developed and refined the random walk strategy in partnership with the survey firm, pilot tested the strategy, and then revised accordingly based on field observations from the research team and notes from staff at the survey firm.

1.7   Sampling Scenarios for second-stage sampling

The survey firm used the decision tree in Figure 1 to determine the sampling methodology for a given town, based on guidelines below.
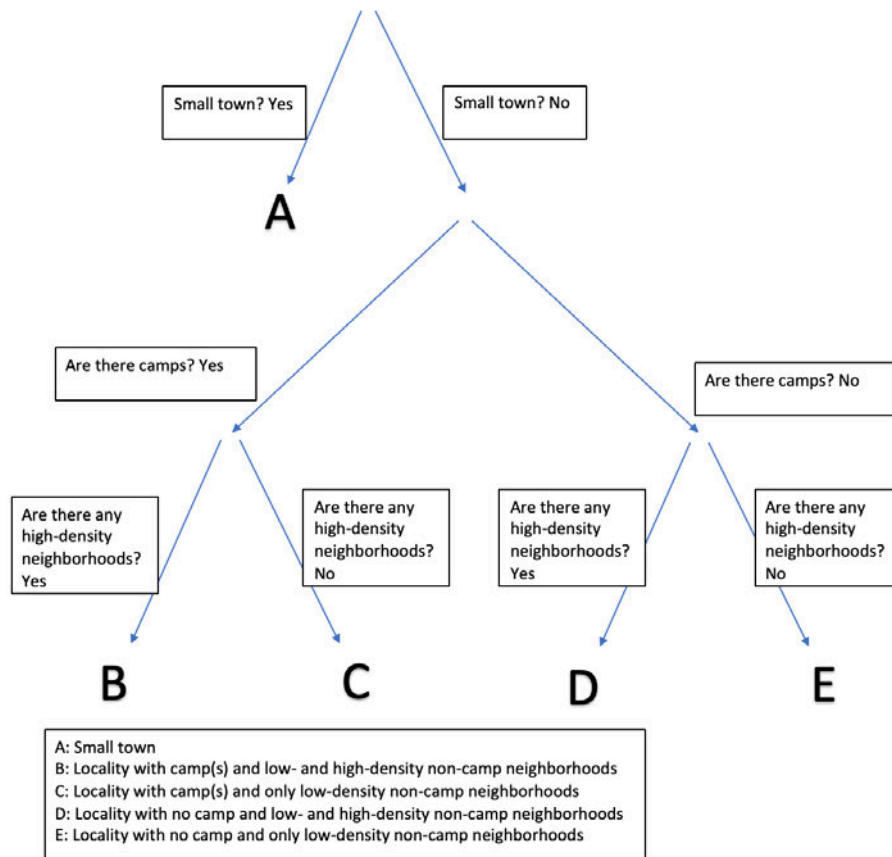
Figure 1: Sampling Scenarios

1.8   Situation A: Small town

Random walk was not necessary in localities with only a couple dozen Syrian households. For very small towns, where the baladiyyeh or another local informant can identify all Syrian households and where the total number of households is not much larger than 20, go to each household. The field staff did this for any household type (apartments, homes, tents etc.). This situation only applied in a small number of cases (maybe 5-10 towns in the sample of more than 90 towns).

In situation A, the field team attempted to interview all Syrians in town. If any household selection is necessary, the field team was trained in this situation to select in an arbitrary fashion, including some households near where the team was initially located and some farther away.

1.9   Neighborhood selection for situations B, C, D, and E

Based on a meeting with a local key informant (like a mukhtar, rais baladiyyeh, or shawish) the survey team would identify neighborhoods where Syrians live. The survey team would arrive with a google map print-out of the locality. The local key informant was asked to draw boundaries around neighborhoods with Syrians, including camps and non-camps. We refer below (as we did in training) to non-camp areas with large numbers of Syrians as 'high-density areas.' We refer to areas with smaller numbers of Syrians spread out through the area 'low-density areas.' Camps are simply called 'camp areas.'

We did not define a strict numeric cut-off for high-density and low-density since most local informants will be making rough estimates. We trained field staff to considered areas to be high density areas when a local key informant could affirm to be true statements like "there are a lot of Syrians" and "there are concentrations (Arabic: tajamoua't) of Syrians." Low-density areas correspond to places where a local key informant might instead characterized the areas as having Syrians "spread across the area" or "not in concentrations," or "there are not a lot of Syrians."

1.10   Rules for Selecting Starting Points for situations B, C, D, and E

Within each neighborhood on the Google map print-out, team leaders defined a number of viable starting points for each neighborhood. The team leader allocated a number of starting points, spread out across each neighborhood/camp.

For each neighborhood, the team leader would randomly select points from those. If there were X enumerators in the field that day, the team leader selected X starting points arbitrarily. See next section for more on this.

Team leaders also gave the enumerators their starting direction (N, NE, E, SE, S, SW, W, NW) for the day. Given that streets limit the direction of movement, the enumerator should proceed along the street that is closest to the starting direction indicated in the enumerator schedule.

1.11   Rules for Distributing Enumerators for situations B, C, D, and E

The team leader would get a sense from the local key informant about what percentage of Syrians live in the different neighborhoods.

Example 1: The municipality (baladiyyeh) tells the team leader that there is one neighborhood where about half of Syrians live. In this case, half of the enumerators were sent to that neighborhood and the others to other neighborhoods.

Example 2: The municipality (baladiyyeh) tells the team leader that about half of Syrians in a locality live in camps, one quarter live in low-density neighborhoods, and one quarter live in one high-density neighborhood. In this case, half of the enumerators were sent to the camps, one quarter of enumerators were sent to low-density neighborhoods, and one quarter were sent to the high-density neighborhood.

1.12   Household selection for situations B, C, D, and E

After beginning walking from the designated point in the designated direction, the enumerators were trained to go to every 3rd building to try to recruit participants. After the enumerator arrived at the 3rd building (house, apartment building, tent, informal shelter, etc.), they would take the following steps.

If an enumerator was in an area with numerous buildings, after arriving to the

third building, they would ask someone nearby, like a shopkeeper or resident, if any Syrians live nearby.

If there were few residences or the enumerator had not received any indication about where Syrians live, they would knock on the door and ask if the household is Syrian. If no one answers or the household was not Syria, they would move to the next building.

They would select one Syrian household per building.

For buildings with more than one residence (e.g., an apartment building) enumerators identified one Syrian family in the building, either by asking people nearby or by knocking on doors. After asking someone if Syrians live inside, if there are multiple options (e.g., there are Syrians living on the first, third, and sixth floors), select one floor arbitrarily, go there and knock on the door. If they no one answers proceed to the other apartments where Syrians live in that building.

1.13   Reaching the edge of a neighborhood

If an enumerator reached the edge of the area designated by the local key informant before recruiting the target number of interviews from that neighborhood, they would return to the starting point, and proceed with household selection moving in the direction opposite of their original starting direction.

1.14   Special considerations for camps: situations B and C

Upon arriving to a camp, the team leader would obtain approval from the shaweesh to collect data in the camp, but make clear that the interviews would be private to make sure that people feel comfortable speaking freely.

Enumerators began at the starting point indicated through the map method above, then proceed in the direction indicated in the daily schedule. Enumerators would sample every 3rd accessible household.

1.15   Household inclusion criteria

Enumerators only surveyed households with Syrian members, including households with at least one Syrian national who is a member of the household by blood or marriage and is at least 18 years of age. This includes Syrian nationals or PRS who live with Lebanese people.

   If a residence included multiple families living together/sharing one dwelling, enumerators asked them about distinct families within the dwelling. From these households select one to be interviewed.

   In some residences, a number of individuals share one dwelling. This is often the case with a groups of young men sharing housing. In such cases, enumerators were instructed to note that one household may be a single person. For example, a number of men who are not related working construction may live together in one apartment. In this case, enumerators selected one person from those who were 18+ years old, present in the dwelling, and were willing to participate.

1.16   Replacement

When a household was ineligible for data collection or refused to participate, data collectors would move to the immediately subsequent accessible dwelling. That is, upon arriving at the 3rd household and asking for Syrians, and being turned away, the enumerator would not count another 3 houses. Instead they would move to the next closest household.

   A household would be considered ineligible for data collection under the following conditions:

1. No member of the household is a Syrian national or Palestinian from Syria

2. An eligible head of household is not available for data collection within any of feasible data collection days and hours

3. They refuse to participate

Moving to the immediately subsequent accessible dwelling entailed:

1. Among single-dwelling buildings or tents, moving to the next dwelling in the direction the enumerator was instructed to walk (i.e., moving in the designated direction until reaching the next dwelling)

2. In multi-dwelling buildings, moving to the next dwelling (e.g., the next apartment, or the next level of multi-home building) with 1 dwelling per floor, moving to the next level of the building (e.g., from the 2nd floor to the 3rd floor)

3. In multi-dwelling buildings, with multiple dwellings per floor, moving to the next apartment of the building (e.g., moving from apartment no. 6 to apartment no. 7).

1.17 Third-stage sampling: Respondent Selection

Enumerators interviewed a head of household — a person with primary decision-making responsibility in the household. An eligible 'head of household' should know the details of the household, its economic situation, and important household decisions. Confirm these facts before accepting a person as a respondent.

If a household has multiple adults who share decision-making responsibility, this classifies as there being multiple heads of household. If a household has more than one head of household, interview whichever head of household is home. If more than one head of household is home, refer to the team leader's schedule, which will inform the enumerator about how to select between more than one head of household.

1.18   Sampling Weights

With the sampling design described here, the probability phvij of selecting individual hvij in household hvi in village hv in stratum h is given by:

The first term on the right-hand side denotes the probability of selecting village hv in the first stage. The first constitutive element of the first term denotes the pps sample size of villages to be drawn in stratum h, that is the product of the stratum share of total Syrians in Lebanon and the total number of villages to be sampled. The second constitutive element of the first term denotes the number of villages in the stratum to sample from. The second term on the right-hand side denotes the probability of selecting household hvi in the second stage. The third term on the right-hand side denotes the probability of selecting individual hvij in the third stage.

- **S:** Nationwide Syrian population estimate

- **V:** Total number of villages to be sampled

- **b:** Total number of strata bins

- v_b Number of villages to be sampled in stratum h

- s_h: Syrian population estimate in stratum h

- m_{hvi}: Total number of Syrian households in village hv

- n_{hv}: Total sample size per village, which is equal to n for all strata and villages

To produce unbiased estimates from the sample, the data from each household *hvi* should be affected by a sampling weight (or raising factor) w_{hvi}, respectively, equal to the inverse of its selection probability (i.e., $w_{hvi} = p_{hvi}^{-1}$).

Within the current study and sampling plan all these parameters have credible empirical estimates from UNHCR records and Lebanese voter registration data.

The number of towns to be drawn in each stratum is the average of the population-proportional number and an even-proportional number:

$$v_b = \frac{\left(\frac{s_h}{S} \times V\right) + \left(\frac{V}{b}\right)}{2} \tag{1}$$

Then we can calculate the probability of given household being sampled as:

$$p_{hvi} = v_b \times \frac{m_{hvi}}{s_h} \times \frac{1}{m_{hvi}} \times n \tag{2}$$

$$= v_b \times \frac{20}{s_h} \tag{3}$$

1.19 Response Rate

Enumerators recorded attempts to reach people and reported response rate to the team leaders. They reported the total number of doors that they knocked on each day (including both doors where no one answered and doors where people answered). They reported the total number of Syrians who refused to be interviewed each day.

Out of 3,882 interview attempts, enumerators conducted 3,003 interviews. Overall success rate was 77%. After someone answered the door and verified that a household member was Syrian, non-response was caused by people declining to participate and people not meeting a key inclusion criterion for our study. That is, in order for people to be recruited we required that they have an active mobile phone so we could stay in touch with them via WhatsApp. Overall, 401/3,882 people refused to participate (10.3%), and 478/3,882 households did not have a phone with WhatsApp (12.3%).

1.20 Replacement

We needed to drop one research site in the Baalbek area from our sample because a shawish refused us entry to the IS. We returned to our original sampling methodology

and drew another town in order to replace this site.

2   Data Collection Quality Assurance

1. Research team conducted regular data checks every 1-3 days, outputting summary statistics for the most recent data, and regularly shared them with the team.

2. Quality checks included the following fields, and example reports for Sep 11 and Sep 12, 2019, are shown in Figure 1 below.

   - Numbers of surveys completed each day.

   - Number of surveys by region.

   - Survey length statistics, flagging any submitted survey that lasted less than 25 minutes.

   - Check if any people required referrals, either to non-urgent medical care, or more urgent problems. For urgent problems we reached out immediately to an NGO.

   - Checked contact numbers for formatting errors, and checked for duplicates that might suggest problematic submissions.

   - Descriptive statistics on how many contact numbers respondents gave us, which will be used in follow-up surveys.

   - Item non-response rate.

   - Return intentions.

   - Checking for logical inconsistencies in survey responses.

   - Enumerators had a question at the end of the survey about who else was present, including one option indicating if authorities had been. If

authorities had been present, we would have stopped data collection in that location.

3. Team members were in Lebanon to directly supervise enumerator training, survey pilot, and the beginning of data collection.

4. The survey firm's senior manager conducted regular quality checks of submitted data and forwarded concerns to us for our attention and discussion on a continual basis.

5. Field team managers, all long-time management at the survey firm, oversaw data collectors in field sites. Every four fieldworkers had one supervisor who escorted them during all data collection and ensured that they followed the sampling technique.

6. Survey firm supervisor sent weekly progress reports to the research team. Progress reports included quality notes itemized by survey ID. Progress reports also included recruitment refusal rates with detailed statistics of recruitment characteristics.

7. Research team built multiple quality checks into the survey instrument and checked them regularly. (Early in data collection these built-in checks allowed us to identify an enumerator submitting problematic data. This enumerator was taken off the project and all her surveys were thoroughly checked for quality and we managed to replace all her problematic surveys with new ones).

8. Continual confirmation of geographic sites of interviews, checking submitted interviews against designated research sites.

3 Survey questions

3.1 Dependent variables

Our outcome of interest is intention to return. The survey included multiple questions about stated short-term and long-term return intentions and questions about preparations for return.

3.1.1 Questions about return intentions

A) Short-term return intentions

- Do you (the respondent) plan to return to Syria in the next 12 months?

- Are other members of your household planning to return in the next 12 months?

  B) Long-term return intentions

- Two years from now, where do you expect to actually be living?

- Do you hope to move back to Syria and live there one day?

- Would you say it is correct that you don't want to return no matter what happens?

3.1.2 Questions about return preparations

- Have you or your immediate family been saving resources in order to prepare for your return to Syria?

- Have you or your immediate family prepared any legal paperwork, such as marriage documents, birth certificates, or proofs of property to prepare for your return to Syria?

- Have you or your immediate family reached out to Lebanese authorities to discuss returning to Syria?

- Have you or your immediate family reached out to UNHCR to discuss returning to Syria?

- Have you or anyone in your immediate family made a scoping trip back to Syria to learn about the situation there?

- Have there been times in the past 12 months when you were planning to return and aborted those plans?

3.2   Independent variables

3.2.1   Push factors from Lebanon

Questions about economic well-being in Lebanon

1. Do you possess the status that allows you to work legally in Lebanon (do you currently possess a work permit or Lebanese residency)?

2. During the past 4 weeks, how many days did you work outside home to make money?)If they did not work, enter 0.)

3. (If worked) On the days that you worked during the past 4 weeks, how many hours did you usually work per day?

4. (If worked) During the past 4 weeks, how much money did you make in total?

5. Can you please tell me which sources your household normally receives income from? By household we mean people who are family members or close relatives and who live under the same roof and share meals with you: Aid or assistance from other organizations (such as the United Nations, other international NGOs, local Lebanese charities, etc.)

6. In the last month, how much money did your household withdraw in total using cards from humanitarian organizations? To clarify, I do not mean using a card to buy from certain shops. I mean using a card to go to and ATM and withdraw cash. (This includes the 260 and other cash aid.) Enumerator: enter the sum of all cash aid used by household members in the last month.

7. In the last month, how much money did your household spend in total using WFP (World Food Program) support, that is, using a card to buy food only from certain shops? (We are referring to the taghziye program.) Enumerator: enter the sum of all food card aid used by household members in the last month.

8. Did you or anyone in your household work in Lebanon before 2011?

9. Does your household income vary from month to month?

10. How does the aid that you're currently receiving compare to the amount of aid that you were receiving a year ago? (Enumerator: Here we are referring to cash and all other aid.)

11. Does this dwelling have the following items that you are able to use?

   - Refrigerator
   - Washing Machine
   - Oven
   - Desktop or Laptop computer at home
   - Car
   - Microwave oven
   - Television

- Internet connection at home (other than through a smartphone, not through a neighbor but owned at home

- Indoor toilet

- Central heating

12. To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize it. Then read all the answer options. In a typical month, what share of your household's monthly expenses and spending needs you are you able to satisfy from household members' income?

13. What is the approximate total value of assets and cash that you possessed when you first came to Lebanon? Note that we're talking about the assets and savings of your household in Lebanon and not anything you left in Syria.

14. To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize it. If you had to live solely on your remaining savings and assets, without any income or debt, approximately how many months of expenses and spending would your savings and assets support you? Note that we're talking about the assets and savings of your household in Lebanon.

15. What was your total household income in the past month? By household we mean people who are family members or close relatives and who lived under the same roof and share meals with you. Enumerator: This does not include income from aid.

Questions about social well-being in Lebanon

1. How would you describe your relationship with Lebanese people?

2. How often do you feel like an outsider in Lebanon?

3. (To enumerator) Did the respondent speak Arabic well?

4. What is the highest level of education you have completed?

5. How well can you read and write?

6. How well do you understand the important issues facing Lebanon?

7. In the last 12 months, how often did you typically discuss major issues facing Lebanon with others?

8. In Lebanon, how difficult or easy would it be for you to do each of the following? Search for a job

9. Which year did you move to Lebanon to stay here until now?

10. Does this town currently have curfews for Syrians?

11. (If no) Has this town had curfews for Syrians in the last 2 years?

12. In your personal experiences, how friendly or hostile would you describe your personal experiences with the Lebanese public in this area (town/neighborhood)? We are asking about the Lebanese general public, not authorities.

13. In your personal experiences, How friendly or hostile would you describe your personal experiences with Lebanese authorities in this area (town/neighborhood)?

14. Have you been detained by Lebanese authorities since arriving to Lebanon? Please note that we will not share this information.

15. Do you think it is hard for you or your family to get housing in Lebanon because you are Syrian? To enumerator: if hard but not because they are Syrians, select "Not hard because we are Syrians"

16. How often are you personally able to travel freely and safely around this area of Lebanon?

17. Are all your household members able to move freely in this town?

Questions about services in Lebanon

1. Have you been sick in the past 6 months to the point of requiring medical treatment? (To enumerator: this does not mean that they actually saw a doctor, just that they needed to be treated. This includes chronic illnesses but not common illnesses like cold)

2. (If yes) Were you able to see a doctor? (To Enumerator: Going to a pharmacy does not count as seeing a doctor)

3. (For each household member) Has (household member) has been sick in the past 6 months to the point of needing medical treatment? (To enumerator: this does not mean that they actually saw a doctor, just that they needed to be treated. This includes chronic illnesses but not common illnesses like cold)

4. (If yes) Was (household member) able to see a doctor? (To Enumerator: Going to a pharmacy does not count as seeing a doctor)

5. In Lebanon, how difficult or easy would it be for you to do each of the following? See a doctor (aside from the cost)

6. Do you think it is hard for your or your family to access healthcare in Lebanon because you are Syrian? (To enumerator: if hard but not because they are Syrians, select "Not hard because we are Syrians")

7. Need school: Whether any children between ages of 6 and 18 never studied OR (did not finish primary school and are above 10 years old) OR (are not currently attending school)

8. Now we would like you to think about all the areas you have lived since moving to Lebanon since you arrived here in (insert year). How many different towns have you lived in in Lebanon (including this place) since you came here in (insert year)? (To measure stability in Lebanon, if always in the same town, the variable gets a value of 3 (most stable). If lived in two towns, the variable gets a value of 2. If lived in more than two towns, the variable gets a value of 1 (least stable))

9. Which year did you start living in in this area (neighborhood or town)? (To code stability in current town, we subtracted answer from 2019 to find number of years in this town then we cut the answers into quantiles)

10. (If age of a child < 18 AND child not currently attending school) Why is (child) not attending school? (Do not read options. Let respondent answer and select all that apply). Any of the following answers would code school as preventive:

    - Person has been out of school for many years and is now at a lower level than his/her peers

    - Financial constraints, such as transportation, obtaining uniforms, textbooks, etc.

    - School is too far

    - There are safety fears or harassment concerns for movement on the way to school

    - Harassment from other students at school

    - Harassment from teachers or school administrators

    - Difficulties adapting to dialect or teaching methods

11. In Lebanon, how difficult or easy would it be for you to do each of the following? Get help with legal problems

12. Does this dwelling have the following items that you are able to use? Running water

Questions about legal conditions in Lebanon

1. How (does respondent possess the status that allows you to work legally in Lebanon (does (respondent) currently possess a work permit or Lebanese residency))? Lebanese residency through family member

2. What is your status with UNHCR? Please note that we will not share this information with anyone (Enumerator: If the respondent says registered or recorded, please ask to see the UNHCR registration file with names of registered individuals). The following coded as 1:

   - Registered (Arabic: Musajjal)

   - Recorded (for people who entered Lebanon after January 2015, Arabic: Mu'arraf)

3. (If this person is a Palestinian from Syria) What is your status with UNRWA? Please note that we will not share this information with anyone. Registered

Questions about networks in Lebanon

1. How many of (original household members from Syria) (excluding yourself) are living in Lebanon now?

2. Please think about the Lebanese people in your phone contacts. With how many of them did you have a conversation either by phone, messenger chat, face-to-face, or text exchange in the last week? Note that this does not include service provides such as the United Nations or NGOs

3. Please think about the Syrians in Lebanon in your phone contacts. With how many of them did you have a conversation either by phone, messenger chat, face-to-face, or text exchange in the last week?

4. Outside of your household, do you have any close Lebanese relatives in this area or elsewhere in the country?

5. In the last 12 months, how often did you share a meal with Lebanese people who are not part of your family? (To enumerator: Those do not need to be friends. They can be people at work or other people.)

3.2.2   Pull factors in Syria

Questions about safety in Syria

1. How would you describe the current risk to civilians' physical safety in (place of origin)?

2. To measure sympathy with opposition, we examine difference between trusting two anti-regime media (Al-Jazeera and Al-Arabiya) and two pro-regime media (Al-Mayadeen and Al-Manar) using the question: How trustworthy would you say each of the following news sources is?

   - Al-Mayadeen
   - Al-Manar
   - Al-Jazeera
   - Al-Arabiya

3. Were there anti-regime protests in (place of origin) in 2011 and 2012?

4. Have you suffered physical or psychological harm because of violence in Syria?

5. How do you expect the safety situation to be in (place of origin) one year from now?

6. Conscription: To examine if someone in household is of/near conscription age, we saw if household includes any males born between 1977 and 2003.

Questions about control in Syria

1. Who mainly controlled (place of origin) in the month before you left?

   - Syrian army
   - Opposition forces such as the FSA
   - Jabhat al-Nusra
   - ISIS
   - Kurdish forces
   - Russian forces
   - Turkish forces
   - It was contested

2. Who mainly controls (place of origin) currently?

   - Syrian army
   - Opposition forces such as the FSA
   - Jabhat al-Nusra
   - ISIS
   - Kurdish forces
   - Russian forces
   - Turkish forces

- It is contested

3. Did ISIS control (place of origin) at all during the conflict?

Questions about economic well-being in Syria

1. How would you describe the current availability of jobs at present in (place of origin) currently?

2. What is the total amount of outstanding debts you currently have in Syria? This includes any debts on unpaid electricity, water, or other bills while you were away

3. Did you or your immediate family own (not rent) the following items in Syria? (ask for each)

   - House (not an apartment)

   - Apartment (other than their house, if they owned a house)

   - Land

4. (If they stated that they own land in previous question) Do you think you would be able to continue as owner and operator of this land if you returned to Syria?

5. (If owned house/apartment/land) Do you or your immediate family have property documents that prove you are the owner?

Questions about services in Syria

1. As far as you know, how many hours per day is there electricity in (place of origin) currently?

2. As far as you know, how many hours per day is there running water in (place of origin) currently?

3. As far as you know, are schools operating in (place of origin) during the school year?

4. As far as you know, are health centers operating in (place of origin) currently?

5. How good do you think that public service provision in (place of origin) (such as health centers, schools, infrastructure) will be one year from now?

Questions about networks in Syria

1. How many of (household members from Syria before leaving) are living in Syria now?

2. Approximately how many Syrian relatives or friends who have lived in Lebanon have gone back to Syria with the goal of staying there?

3. Next, think about your Syrian friends and relatives from (place of origin) who have lived in Lebanon. Approximately how many of them have gone back to (place of origin)?

4. How many people who were in your household in Lebanon at some point since 2011 have gone back to Syria, regardless of where they are now?

3.2.3 Confidence in information

Questions about confidence in information

1. To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about the safety situation in (place of origin), would you say that...: I know enough to be confident I understand the situation

2. To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about employment opportunities in (place of origin), would you say that...: I know enough to be confident I understand the situation

3. To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about the availability of public services (such as health centers, schools, or water) in (place of origin), would you say that...: I know enough to be confident I understand the situation

4. To enumerator: This question is written in colloquial Arabic. Read it loud as it is and do not try to summarize. Then read all the answer options. When thinking of your knowledge about the conscription requirements by the Syrian military, would you say that...: I know enough to be confident I understand the situation

5. Now, please think about the person you communicate with the most who is currently living in Syria. (Respondent did not choose: I don't communicate with anyone in Syria)

6. In general, how often do you communicate with people in (place of origin)?

3.2.4 Mobility

We measured mobility in two ways. First, we measured the distance between the current location of respondents in Lebanon and their place of origin in Syria. Travel routes were calculated using the Google Maps API. We used the R package `mapsapi` and commands `mp_directions()` and `mp_get_routes()`. We calculate travel distance from each survey respondent's town of residence in Lebanon to their hometown

in Syria, via the Beirut Damascus highway and border crossing. Our fieldwork revealed that this was the only legal border crossing open at the time of research and that a majority of Syrians moving back travel via official routes. Second, we used the log of household size. The descriptive statistics for these variables are included in Section 4

3.3 Multiple Imputation

All results for the survey in Lebanon impute missing values using multivariate imputation by chained equations. We specify 10 imputations and use random forest to predict missing values using the `mice()` package in R. We use Rubin's rules (Rubin, 1987) when pooling estimates across imputations.

4 Index construction and descriptives

We measured four key concepts with multiple independent variables using the first principal component.

1. Well-being in Lebanon

2. Expected well-being in Syria

3. Confidence in information

4. Preparation to return

The component variables are mentioned in Section 3. For the first three independent variables, we have many input variables.

Our primary analysis was based on regression models with indices constructed using polychoric PCA unless any variable in the index had too many categories (above 8 categories, in which case we used pearson correlations) (Hainmueller, Hangartner

and Pietrantuono, 2017). Scales for individual questions were reversed as necessary to simplify interpretation (to make sure they were all in the same positive direction). In addition to the indices constructed using the first principal components, we ran mean effects indices (aka z-scores) as a robustness check (Kling, Liebman and Katz, 2007). We constructed those indices by standardizing each variable (demeaning and dividing it by the standard deviation). We then summed the standardized variables and then standardized the sum again.

This section shows the component variables of each of the indices as well as descriptive statistics for these variables. We also show descriptive statistics for the control variables we included in the main regressions.

4.1 Push factors from Lebanon

Economic well-being in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Aid: atm card	0.63	0.58	1.39	0.00	0.00	7.00	0.30%
Aid change from last year	1.49	1.48	0.56	1.00	1.00	3.00	9.16%
Aid: wfp card	1.12	1.02	1.68	0.00	0.00	7.00	0.13%
Assets: months left	0.13	0.15	0.62	0.00	0.00	12.00	0.00%
Assets: value upon arrival	0.62	0.70	2.26	0.00	0.00	16.00	0.27%
Ability to cover expenses	2.78	2.88	1.26	1.00	3.00	5.00	0.17%
Income	1.75	1.90	3.07	0.00	0.00	13.00	0.10%
Work days past 4 weeks	5.82	6.25	9.91	0.00	0.00	28.00	0.00%
Work hours past 4 weeks	2.92	3.13	4.74	0.00	0.00	24.00	0.00%
Able to work legally	0.10	0.11	0.31	0.00	0.00	1.00	0.03%
Income source: aid	0.36	0.33	0.47	0.00	0.00	1.00	0.03%
Stable household income	0.81	0.79	0.71	0.00	1.00	2.00	0.20%
Household income	4.36	4.72	3.50	0.00	5.00	17.00	0.93%
HH worked in Leb. before 2011	0.23	0.24	0.43	0.00	0.00	1.00	0.00%
Own refrigerator	0.80	0.81	0.39	0.00	1.00	1.00	0.00%
Have indoor toilet	0.83	0.83	0.37	0.00	1.00	1.00	0.00%
Have central heating	0.05	0.05	0.22	0.00	0.00	1.00	0.13%
Own washing machine	0.68	0.67	0.47	0.00	1.00	1.00	0.00%
Own oven/stove	0.79	0.79	0.41	0.00	1.00	1.00	0.00%
Own computer	0.02	0.02	0.13	0.00	0.00	1.00	0.00%
Own car	0.02	0.02	0.14	0.00	0.00	1.00	0.00%
Own microwave oven	0.04	0.05	0.21	0.00	0.00	1.00	0.00%
Own television	0.78	0.79	0.41	0.00	1.00	1.00	0.00%
Have internet	0.24	0.25	0.44	0.00	0.00	1.00	0.07%

Table 8: Summary statistics of variables included in constructing the economic well-being in Lebanon index

Social well-being in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Never had curfews	0.65	0.63	0.48	0.00	1.00	1.00	0.93%
Authorities discrimination (higher is less)	2.71	2.73	0.74	1.00	3.00	4.00	15.38%
Ease of mobility	3.16	3.16	1.09	1.00	4.00	4.00	0.13%
Ease of mobility for household	1.62	1.61	0.78	1.00	1.00	3.00	0.53%
Public discrimination (higher is less)	2.97	2.98	0.68	1.00	3.00	4.00	1.60%
Arabic speaking ability	2.87	2.87	0.34	1.00	3.00	3.00	1.27%
Relation with Lebanese	3.66	3.68	0.95	1.00	4.00	5.00	0.33%
Discuss Lebanese politics	1.35	1.38	0.94	1.00	1.00	5.00	0.47%
Ease job search	1.59	1.61	1.02	1.00	1.00	5.00	0.83%
Literacy level	2.10	2.09	0.72	1.00	2.00	3.00	0.00%
Feeling outsider (higher is less)	2.95	2.96	1.29	1.00	3.00	5.00	0.03%
Know Lebanese politics	2.33	2.33	1.24	1.00	2.00	5.00	0.83%
Time in Lebanon	5.52	5.46	2.05	0.00	6.00	9.00	0.07%
No curfew now	0.75	0.73	0.45	0.00	1.00	1.00	0.50%
Housing discrimination (higher is less)	2.48	2.49	0.69	1.00	3.00	3.00	0.33%
Never detained	0.94	0.94	0.23	0.00	1.00	1.00	0.20%

Table 9: Summary statistics of variables included in constructing the social well-being in Lebanon index

Services in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Can access legal services	1.66	1.66	1.07	1.00	1.00	5.00	2.73%
No healthcare discrimination	2.37	2.37	0.73	1.00	3.00	3.00	0.47%
Not sick	0.83	0.83	0.38	0.00	1.00	1.00	0.07%
Received treatment (if sick)	0.94	0.95	0.23	0.00	1.00	1.00	0.00%
Can access doctor	2.80	2.87	1.66	1.00	3.00	5.00	0.70%
No Kids need school	0.59	0.59	0.49	0.00	1.00	1.00	0.03%
Have running water	0.80	0.81	0.39	0.00	1.00	1.00	0.00%
No HH member sick	0.78	0.79	0.41	0.00	1.00	1.00	0.17%
HH members treated if sick	0.91	0.91	0.28	0.00	1.00	1.00	0.03%
School not preventive	0.96	0.96	0.20	0.00	1.00	1.00	0.07%
Period in current town	2.39	2.37	1.16	1.00	2.00	4.00	0.10%
Towns lived in Lebanon	2.64	2.65	0.63	1.00	3.00	3.00	0.10%

Table 10: Summary statistics of variables included in constructing the services in Lebanon index

Legal situation in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Registered with UNHCR/UNRWA (or resident)	0.81	0.80	0.40	0.00	1.00	1.00	0.00%
Legal resident in Lebanon	0.04	0.05	0.21	0.00	0.00	1.00	0.00%

Table 11: Summary statistics of variables included in constructing the legal situation in Lebanon index

Networks in Lebanon

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Syria HH members living in Leb. now	4.58	4.45	3.51	0.00	4.00	15.00	0.00%
Lebanese phone contacts	1.71	1.71	1.11	1.00	1.00	5.00	0.53%
Share meals with Lebanese	1.51	1.53	1.08	1.00	1.00	5.00	0.17%
Syrian phone contacts	3.22	3.20	1.37	1.00	3.00	5.00	0.30%
Lebanese relatives	0.22	0.22	0.55	0.00	0.00	2.00	0.10%

Table 12: Summary statistics of variables included in constructing the networks in Lebanon index

4.2   Pull factors in Syria

Safety in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
HH male at conscription age	0.80	0.81	0.39	0.00	1.00	1.00	0.00%
Exposed to violence	0.28	0.27	0.45	0.00	0.00	1.00	0.03%
Follow anti-regime media more than pro-regime media	0.20	0.21	0.41	0.00	0.00	1.00	0.00%
Hometown had protests	0.67	0.67	0.47	0.00	1.00	1.00	2.40%
Expect hometown to be safe	2.54	2.53	0.86	1.00	3.00	4.00	18.81%
Current safety in hometown	2.05	2.06	0.88	1.00	2.00	4.00	7.13%

Table 13: Summary statistics of variables included in constructing the safety in Syria index. Respondents were coded as following anti-regime media more than pro-regime media if they reported following Al-Jazeera or Al-Arabiya (anti-regime) more than Manar/Mayadeen (pro-regime).

Control in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Contested Now	0.05	0.05	0.22	0.00	0.00	1.00	5.16%
Contested before leaving	0.15	0.15	0.36	0.00	0.00	1.00	3.06%
Controlled by Kurds now	0.13	0.14	0.35	0.00	0.00	1.00	5.16%
Controlled by oppsn/FSA now	0.09	0.09	0.29	0.00	0.00	1.00	5.16%
Controlled by regime now	0.69	0.67	0.47	0.00	1.00	1.00	5.16%
Controlled by Russia now	0.00	0.01	0.08	0.00	0.00	1.00	5.16%
Controlled by Turkey now	0.01	0.02	0.12	0.00	0.00	1.00	5.16%
Controlled by Kurds before leaving	0.02	0.03	0.17	0.00	0.00	1.00	3.06%
Controlled by oppsn/FSA before leaving	0.28	0.29	0.45	0.00	0.00	1.00	3.06%
Controlled by regime before leaving	0.38	0.36	0.48	0.00	0.00	1.00	3.06%
Controlled by Russia before leaving	0.00	0.00	0.06	0.00	0.00	1.00	3.06%
Controlled by Turkey before leaving	0.00	0.00	0.05	0.00	0.00	1.00	3.06%
isis_control	0.37	0.39	0.49	0.00	0.00	1.00	5.13%

Table 14: Summary statistics of variables included in constructing the control in Syria index

Economic well-being in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Debt in Syria	0.31	0.32	0.66	0.00	0.00	2.00	4.96%
Job situation in origin	1.55	1.54	0.70	1.00	1.00	4.00	9.62%
Home ownership docs (1 for some, 2 for everything)	0.58	0.62	0.91	0.00	0.00	2.00	1.90%
Can operate land in future	0.12	0.13	0.33	0.00	0.00	1.00	2.33%
Own house in Syria	0.62	0.64	0.48	0.00	1.00	1.00	0.30%
Own apt in Syria	0.07	0.07	0.25	0.00	0.00	1.00	0.37%
Own land in Syria	0.21	0.21	0.41	0.00	0.00	1.00	0.63%

Table 15: Summary statistics of variables included in constructing the economic well-being in Syria index

Services in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Electricity in origin	2.51	2.51	1.27	1.00	3.00	5.00	15.25%
Health services in origin	0.36	0.36	0.48	0.00	0.00	1.00	19.21%
Expect services to improve in 1 year	2.43	2.42	0.84	1.00	3.00	4.00	17.82%
Schools in origin	0.34	0.34	0.47	0.00	0.00	1.00	19.28%
Running water in origin	2.47	2.46	1.23	1.00	3.00	5.00	14.72%

Table 16: Summary statistics of variables included in constructing the services in Syria index

Services in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Electricity in origin	2.51	2.51	1.27	1.00	3.00	5.00	15.25%
Health services in origin	0.36	0.36	0.48	0.00	0.00	1.00	19.21%
Expect services to improve in 1 year	2.43	2.42	0.84	1.00	3.00	4.00	17.82%
Schools in origin	0.34	0.34	0.47	0.00	0.00	1.00	19.28%
Running water in origin	2.47	2.46	1.23	1.00	3.00	5.00	14.72%

Table 17: Summary statistics of variables included in constructing the services in Syria index

Networks in Syria

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
No. HH members returned to Syria	0.10	0.11	0.65	0.00	0.00	5.00	0.27%
Relatives permanently return to Syria	0.97	1.05	2.99	0.00	0.00	15.00	1.40%
Relatives return to origin	0.53	0.60	2.44	0.00	0.00	15.00	1.17%
Syria HH members living in Syria now	1.86	2.00	3.36	0.00	0.00	15.00	0.10%

Table 18: Summary statistics of variables included in constructing the networks in Syria index

4.3   Confidence in information

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Know Syr. conscription policy	0.28	0.30	0.46	0.00	0.00	1.00	0.23%
Know employment in origin	0.27	0.29	0.46	0.00	0.00	1.00	0.13%
Know safety in origin	0.26	0.27	0.45	0.00	0.00	1.00	0.10%
Know services in origin	0.25	0.27	0.44	0.00	0.00	1.00	0.20%
Communication freq. with origin	2.39	2.45	1.69	1.00	2.00	6.00	0.03%
Communication with someone in Syria	0.61	0.62	0.49	0.00	1.00	1.00	0.03%

Table 19: Summary statistics of variables included in constructing the confidence in information index

4.4 Mobility

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
household size (logged)	1.50	1.48	0.56	0.00	1.61	2.89	0.00%
travel distance (logged)	12.83	12.85	0.52	11.13	12.98	15.63	0.37%

Table 20: Summary statistics of variables to measure mobility

4.5 Preparation to return

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Planned to return but aborted	0.03	0.03	0.17	0.00	0.00	1.00	0.27%
Reached to Leb. authorities about return	0.01	0.01	0.08	0.00	0.00	1.00	0.13%
Prepared docs for return	0.02	0.03	0.17	0.00	0.00	1.00	0.10%
Saved resources for return	0.02	0.02	0.15	0.00	0.00	1.00	0.10%
Conducting scoping trip to Syria	0.03	0.03	0.17	0.00	0.00	1.00	0.10%
Reached to UNHCR about return	0.01	0.01	0.09	0.00	0.00	1.00	0.17%

Table 21: Summary statistics of variables included in constructing the preparation to return index

4.6 Covariates

The following covariates are included in the regressions. When including fixed effects, we also add locality fixed effects in Lebanon and Syria.

Variable	Weighted.Mean	Unweighted.Mean	Std.Dev	Min	Median	Max	Pct.Missing
Household includes elderly	0.04	0.04	0.19	0.00	0.00	1.00	0.00%
Female headed single-parent household	0.11	0.10	0.30	0.00	0.00	1.00	0.03%
High school graduate	0.11	0.12	0.32	0.00	0.00	1.00	0.03%
Hezbollah controlled area	0.16	0.17	0.38	0.00	0.00	1.00	0.00%
Location: Tental settlement	0.35	0.33	0.47	0.00	0.00	1.00	0.00%
Sick required medical treatment	0.30	0.30	0.46	0.00	0.00	1.00	0.23%
Syria origin: urban	0.23	0.23	0.42	0.00	0.00	1.00	0.20%
Household includes toddler	0.45	0.44	0.50	0.00	0.00	1.00	0.00%

Table 22: Summary statistics of variables included as controls in the regressions

5 Scree plots for principal component analysis

The following figures display how eigenvalues change with each additional component for the indices we created using PCA. Throughout the analysis, we used the first principal component.

5.1 Push factors from Lebanon
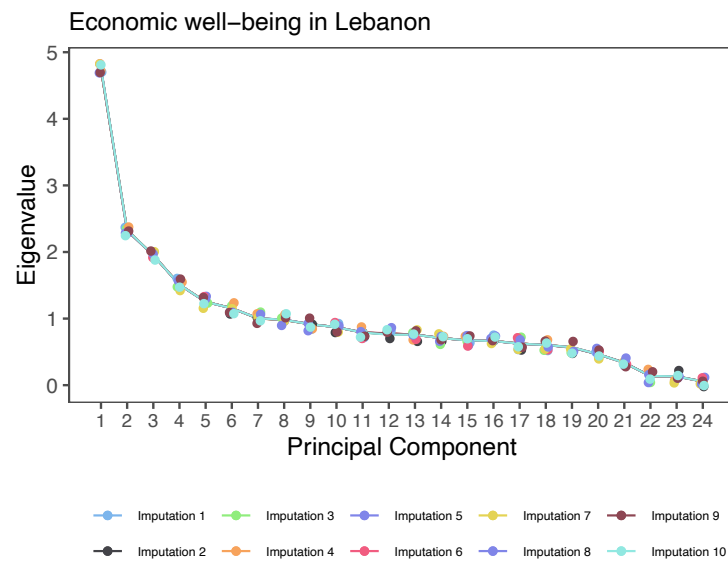
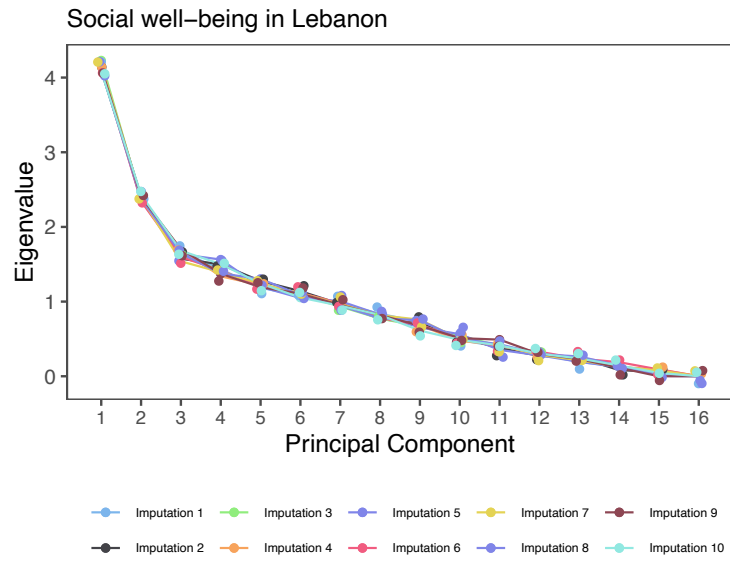Figure 2: Screeplot for the economic well-being in Lebanon index

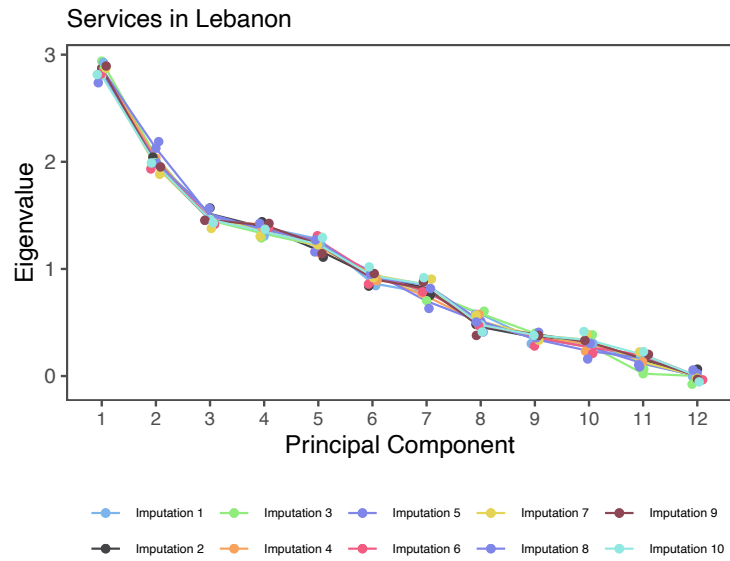Figure 3: Screeplot for the social well-being in Lebanon index
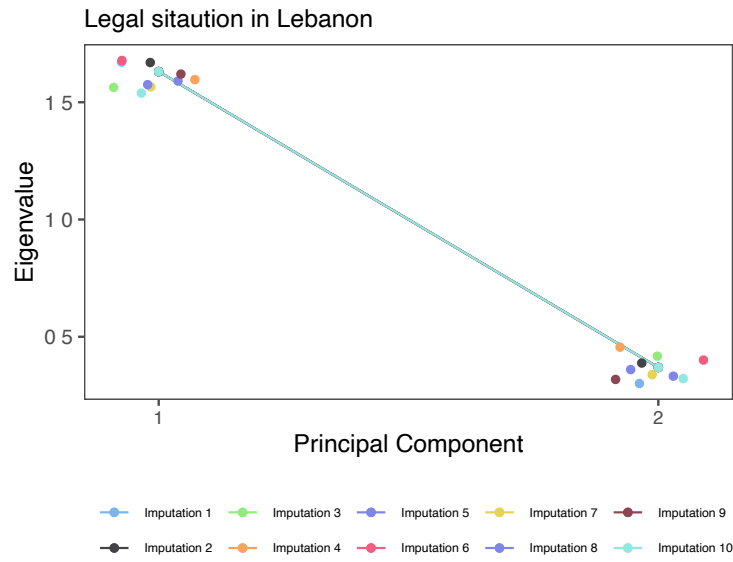

Figure 4: Screeplot for the services in Lebanon index

Figure 5: Screeplot for the legal situation in Lebanon index
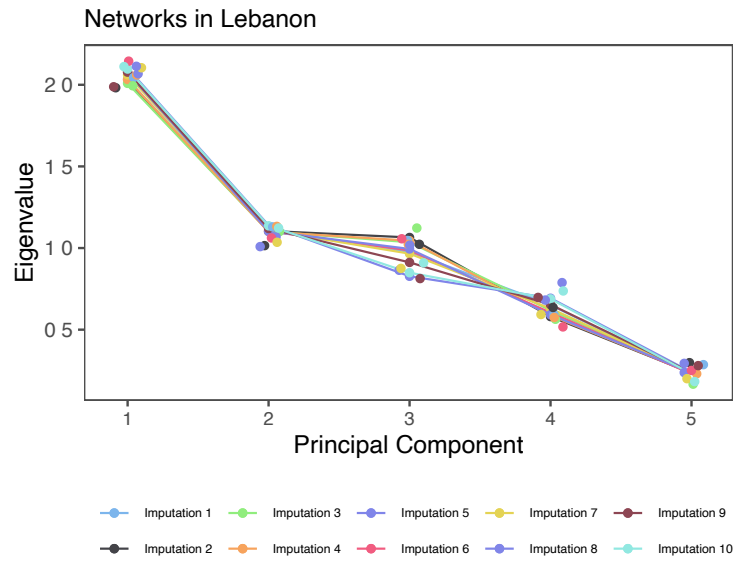
Figure 6: Screeplot for the networks in Lebanon index

5.2   Pull factors in Syria
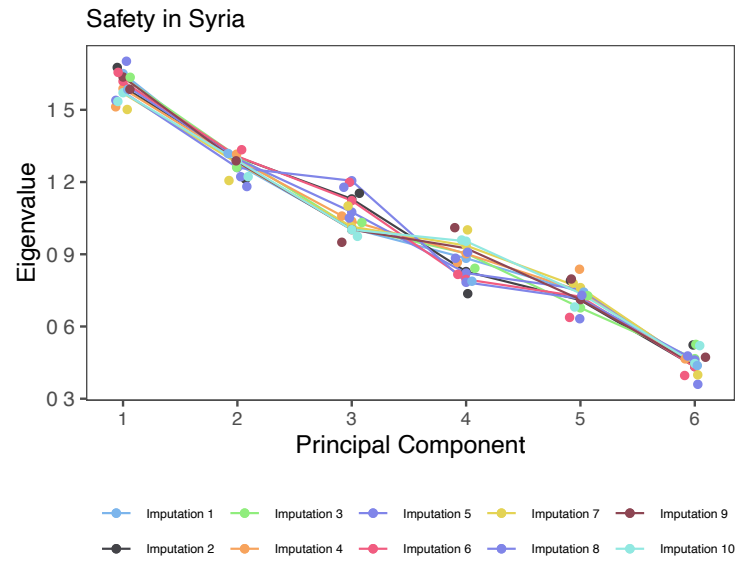
Figure 7: Screeplot for the safety in Syria index
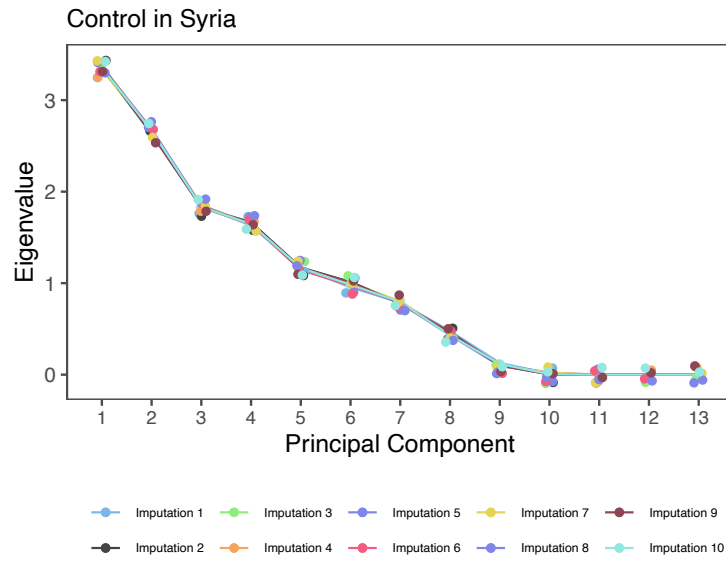
Figure 8: Screeplot for the control in Syria index
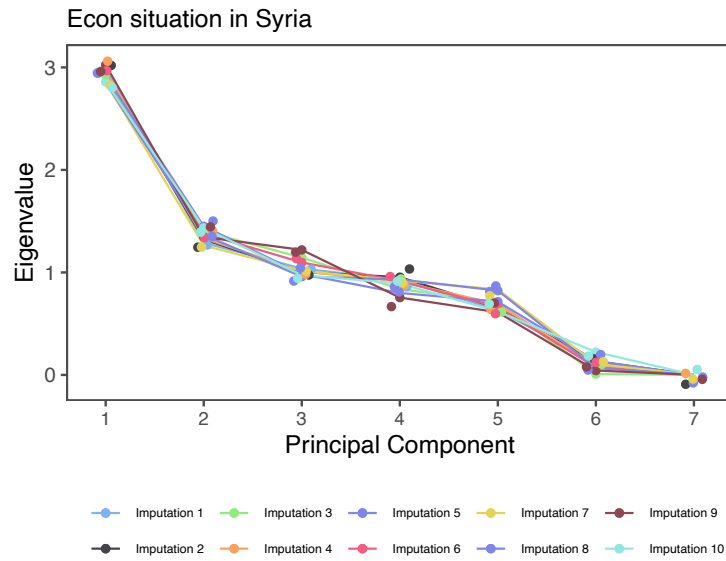
Figure 9: Screeplot for the economic well-being in Syria index
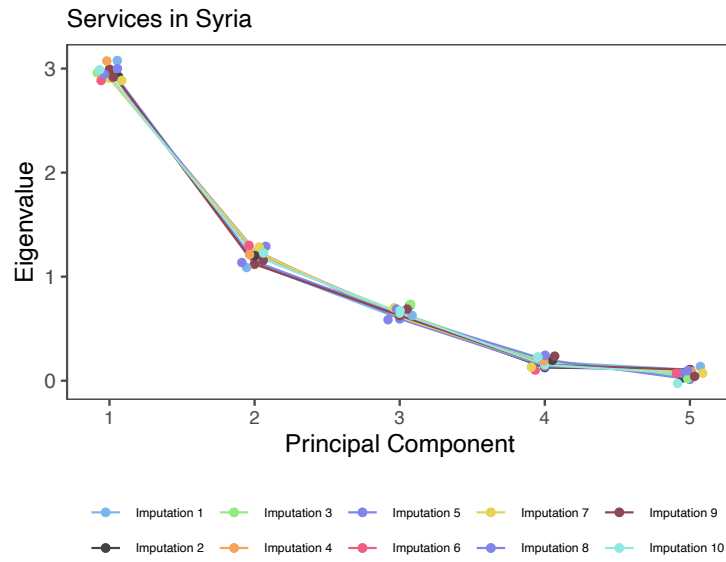
Figure 10: Screeplot for the services in Syria index
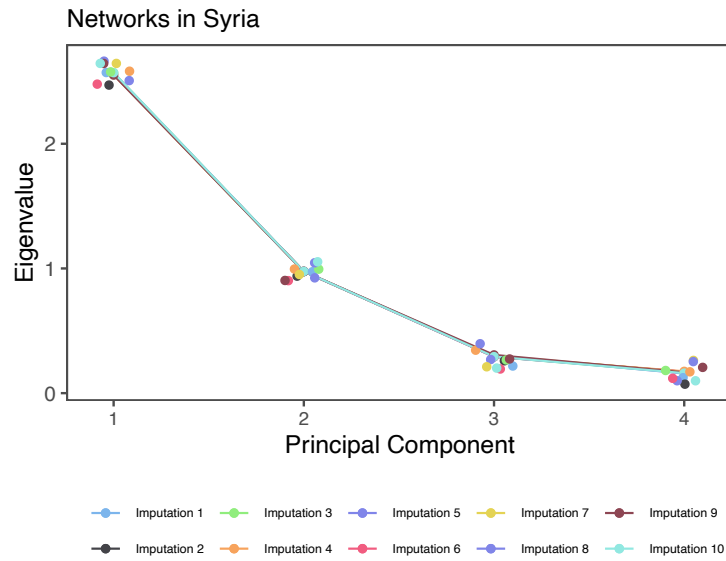
Figure 11: Screeplot for the networks in Syria index

5.3 Confidence in information
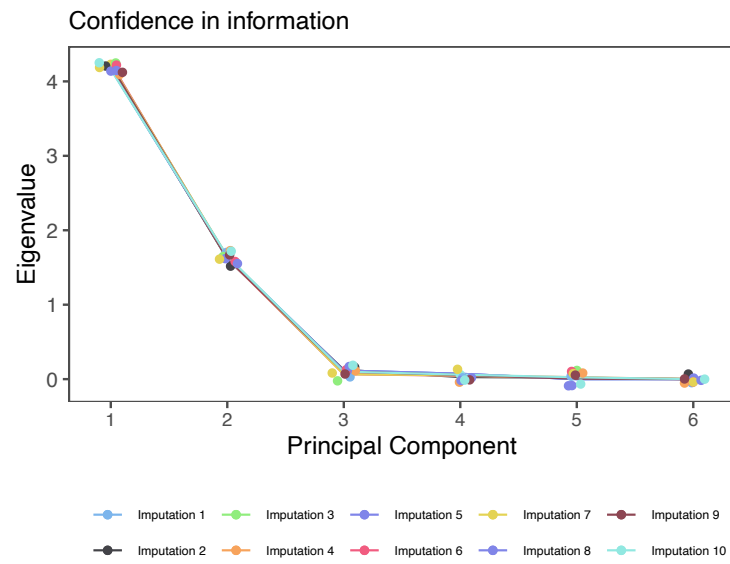
Figure 12: Screeplot for the confidence in information index

5.4 Preparation for return
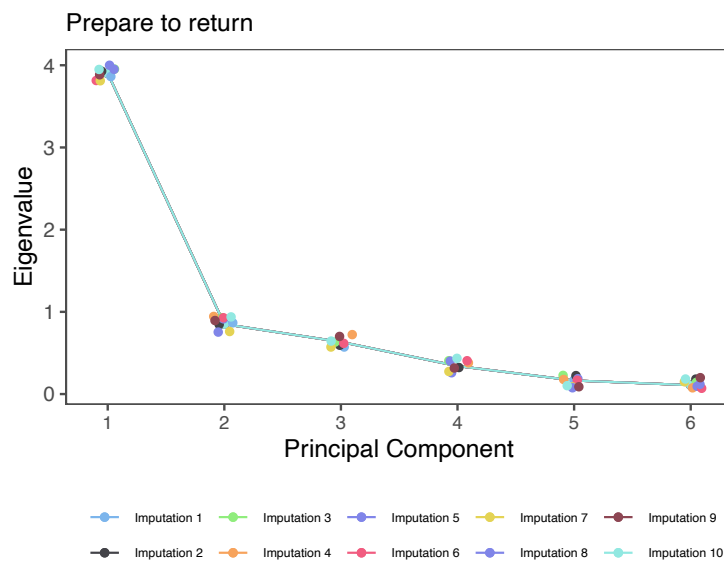
Figure 13: Screeplot for the preparation to return index

6 Deviations from PAP

6.1 Multicollinearity

See Section 7.7 for a detailed discussion and tests of multicollinearity. We indicated in the PAP that we would run one regression model with all indices included on the right-hand side. Upon inspection of our data, however, we discovered that two indices were highly correlated and lead to multicollinearity in our regression analysis. Therefore, we opted for two distinct analyses that avoid the problem of multicollinearity. First, rather than the full model, we run k regressions, one for each of the k indices, regressing return intentions on a single index and a pre-specified vector of control variables in each regression.

We also present results with the pre-specified regression with a minor adjustment

to avoid multicollinearity. As also discussed in Section 7.1, in our regression analyses with all indices, we did not include both safety in Syria and services in Syria because of the high correlation between the two indices. Instead, we ran a regression that included all indices except services in Syria, and then ran a separate regression that included all indices except safety in Syria. We then reported the results of all the coefficients from the first regression and drew the coefficient for services from the second regression. In Section 7.1, we show that results are substantively identical when we do the opposite, drawing all coefficients from the second regression, and adding the coefficient for safety in Syria from the first regression.

6.2   Analysis without locality fixed effects

First, we re-run our main analysis without Lebanese locality fixed effects as a robustness check. This robustness check was not pre-specified, but suggested later by a colleague. This regression can be seen in Figure 19 in Section 7 of the appendix.

Second, after PAP submission we realized that it would be misguided to control for Lebanese locality and Syrian locality when analyzing the role of travel distance from Lebanese locality to Syrian locality. Therefore, in regressions with travel distance on the right-hand side, we do not include locality fixed effects.

6.3   Predictive analysis

The predictive analyses included in appendix section 7.6 were not pre-specified, but were suggested later by a colleague. As discussed in appendix section 7.6, these are run on both the empirical model from the PAP and an alternative extensive model.

6.4   PCA inputs

PCA inputs were pre-specified but required a number of ex post modifications for reasons explained below.

6.4.1   Index 1—economic well being in Lebanon

The PAP mistakenly indicated that a question about someone's former job in Syria would be included in the economic well-being *in Lebanon* index. This was a typo, and it was removed since it is not a dimension of economic well-being *in Lebanon*.

6.4.2   Index 2—social well being in Lebanon

First, upon writing the PAP we intended to calculate the IPL-12 integration score, and use that as an input for PCA. But after pre-field revisions to the questionnaire, some questions were modified and no longer matched IPL-12 inputs. Therefore we modified this index slightly and now use the IPL-12-inspired component variables as inputs, rather than calculating the IPL-12 score and then including it as an input.

Second, household income in Lebanon was mistakenly included in both index 1 and index 2. Therefore, we excluded this question from index 2, since it fits better as an input for index 1. We intended to define index inputs as mutually exclusive. That is, we did not intend to include any variable as an input in multiple indices. A handful of survey questions, however, were mistakenly listed in the PAP for inclusion in two indices. We therefore departed from the PAP in these cases in order to maintain mutually exclusive index inputs.

6.4.3   Index 3—services in Lebanon

A question included in the PAP was subsequently cut from the survey to reduce length: "Have you been forced to move in the last two years, for instance because you were kicked out of your home or your home was deconstructed/demolished?" Therefore, this question is not used an input for Index 3 as was pre-specified.

6.4.4   Index 5.1—Safety

First, we separated regime control from safety conditions in Syria. This was a contingency we detailed in section 5.2 of the PAP. We realized that for theoretical reasons we wanted to directly study territorial control, rather than bundling it with security. Furthermore, we discovered during data analysis that these two concepts did not have a clear relationship, suggesting that they do not deserve inclusion in the same index. Therefore we moved variables about control of territory to a separate index (see Index 5.2 below).

Second, family deaths in Syria was excluded due to flaw in measurement strategy. Our attempt at an indirect question to measure family deaths was to ask how many members the households had before migration and subtract the number of people who were now in Syria, how many were now in Lebanon, and how many were now in a third country. We reasoned that the remainder between the first question and latter three would measured the number of households who had had a household member die. The metric proved to be unreliable. The metric yielded 292 households with deaths, 2500 with no deaths, 212 negatives (uninterpretable), suggesting that the measure is very noisy.

6.4.5   Index 5.2—Regime control

This index was originally bundled under safety (Index 5.1), but for theoretical and empirical reasons we decided to study it independently. Index uses respondents answers to 3 questions: Who controls your place of origin now, who controlled your place of origin before you left, and for how long did ISIS control your place of origin.

6.4.6   Index 10—Information quality

After submitting the PAP we removed the following question from the survey to trim length. We removed this question since it was redundant (and less detailed) than other questions in the information quality index. "How confident are you in your

knowledge about conditions in [Piped place of origin]?" Therefore, this question is not used an input for Index 10 as was pre-specified.

7   Robustness Checks

7.1   All indices regression: switching services for safety

As demonstrated in Section 7.7, two of the indices that serve as our dependent variables are highly correlated. When including the regression analyses with all indices in the paper, we did not include both safety in Syria and services in Syria because of the high correlation between the two indices. Instead, we ran a regression that included all the predictors except services in Syria, and then ran a separate regression that included all the predictors except safety in Syria. We then reported the results of all the coefficients from the regression that included safety (but not services) in Syria. We finally added the single coefficient for services from the second regression.

Here, we do the opposite. We include all the coefficients from the regression that included services (but not safety) in Syria and then add the single coefficient for safety in Syria from the first regression.
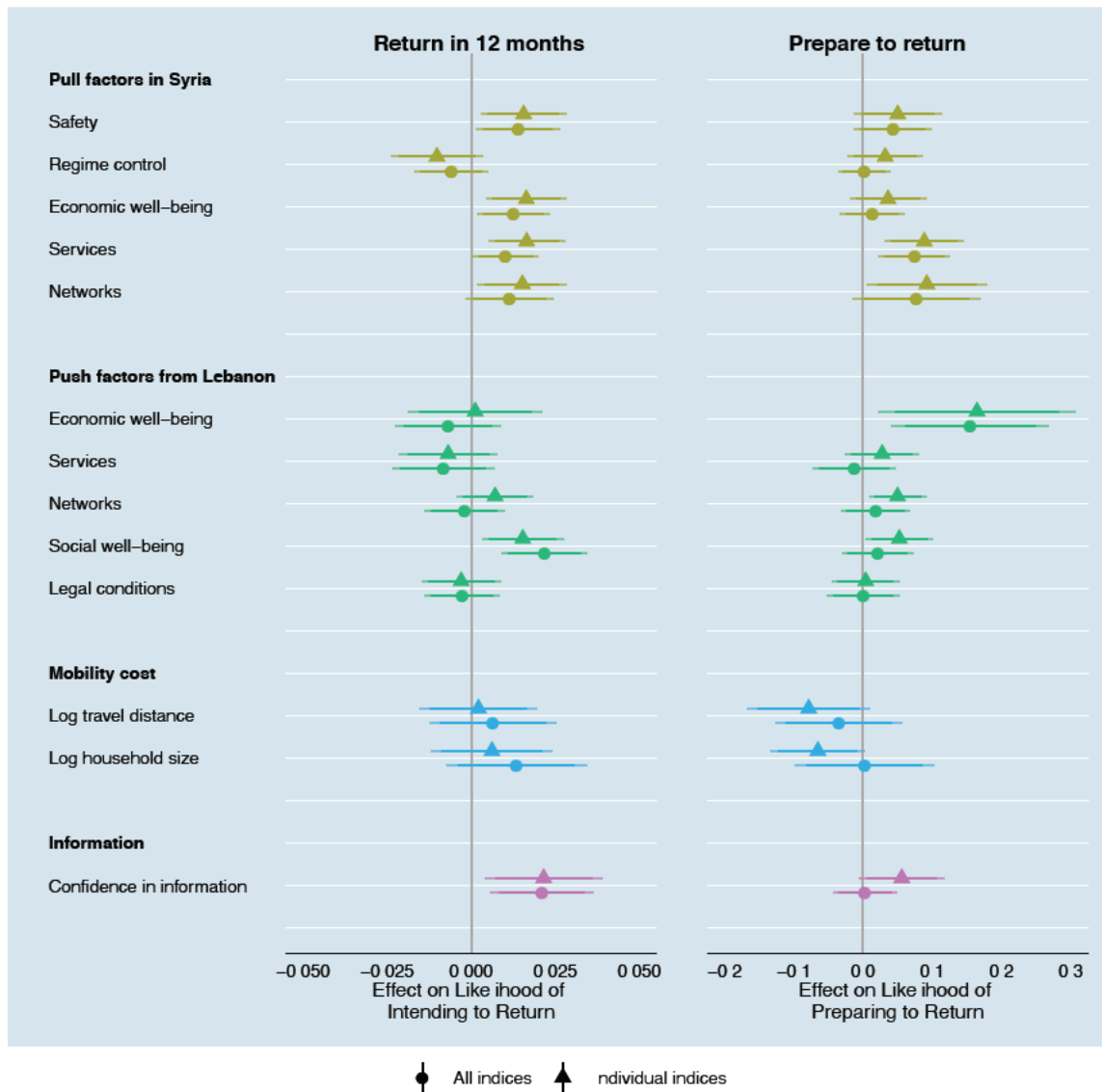
Figure 14: Predictors of return intentions

7.2   Alternative measures of outcomes

In the paper, we showed the predictors of return intentions in the next 12 months. The outcome measure was coded as a binary variable, where 1 meant that the head

of household planned to return in the next 12 months and 0 otherwise. Here, we test the results using different measures of return intentions.

7.2.1   Return in 2 years

In the paper, we presented the results for when the head of household indicates they're plans to return in the next 12 months or ever. We also asked respondents "Two years from now, where do you expect to actually be living?" The following figure shows the predictors of respondents that they expect to live in Syria in two years.
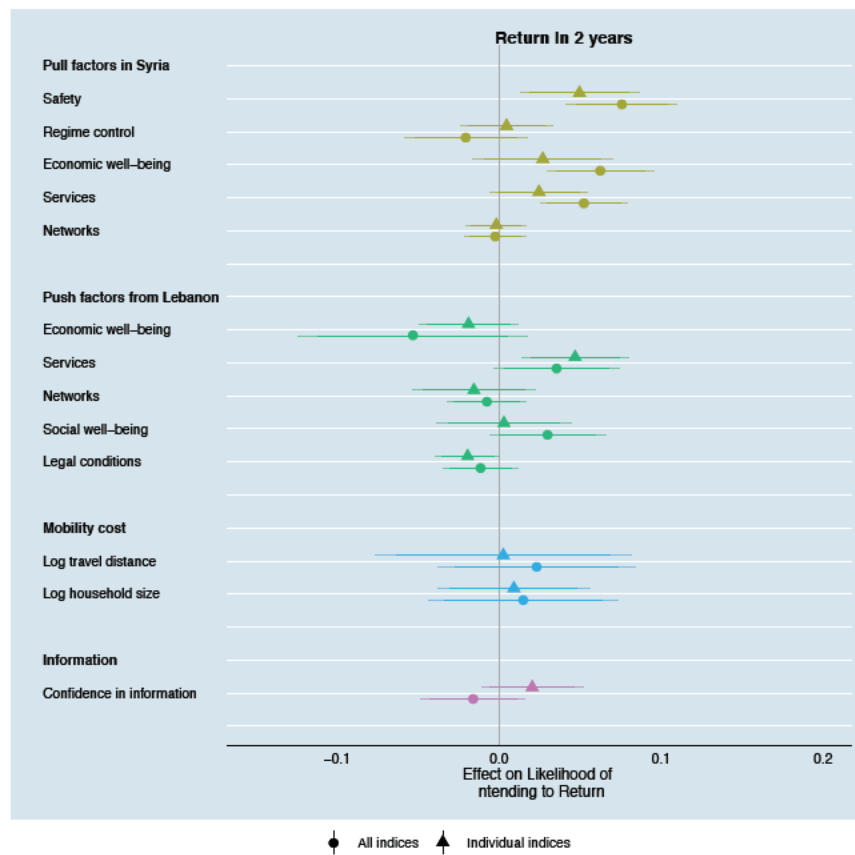

Figure 15: Predictors of return intentions in 2 years

7.2.2 Household members return intentions

In the paper, the main results describe the head of household's return intentions in the next 12 months. In the following figure, the outcome is whether any household member plans to return in the next 12 months.
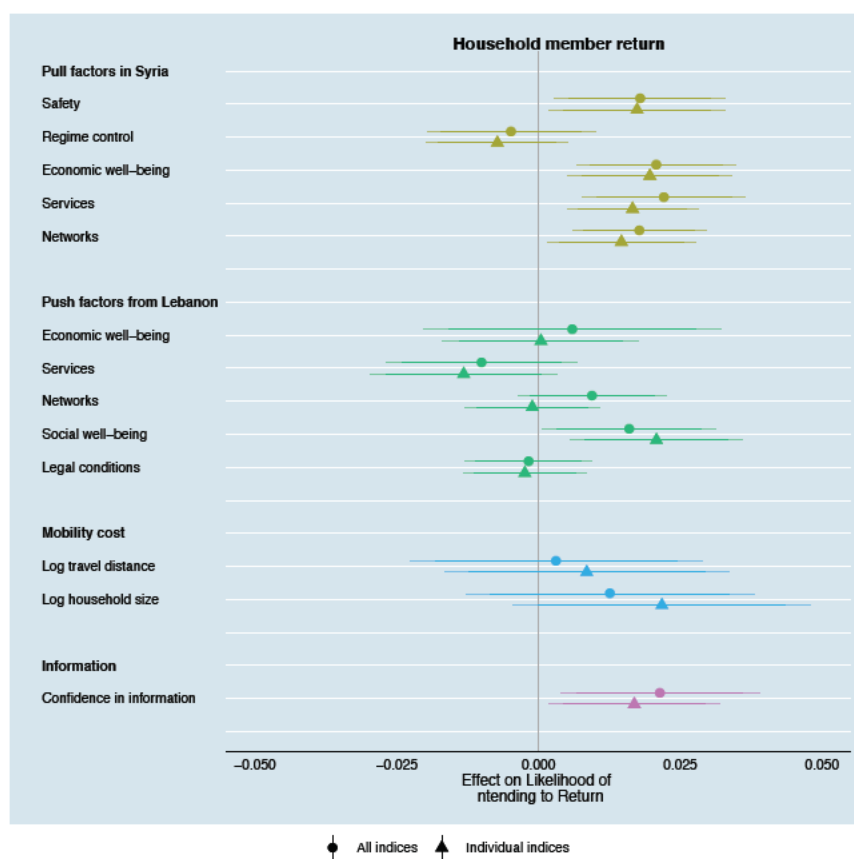

Figure 16: Predictors of any household member's return intentions

7.2.3 Including people uncertain about return plans

In the paper, the main results describe the head of household's return intentions in the next 12 months. There, we coded people who said they plan to return in the next 12 months as 1 and people who were uncertain/indicated they did not want to

return as 0. In the figure below, we code people who are uncertain as 1 (in addition to people who say they want to return in the next 12 months).
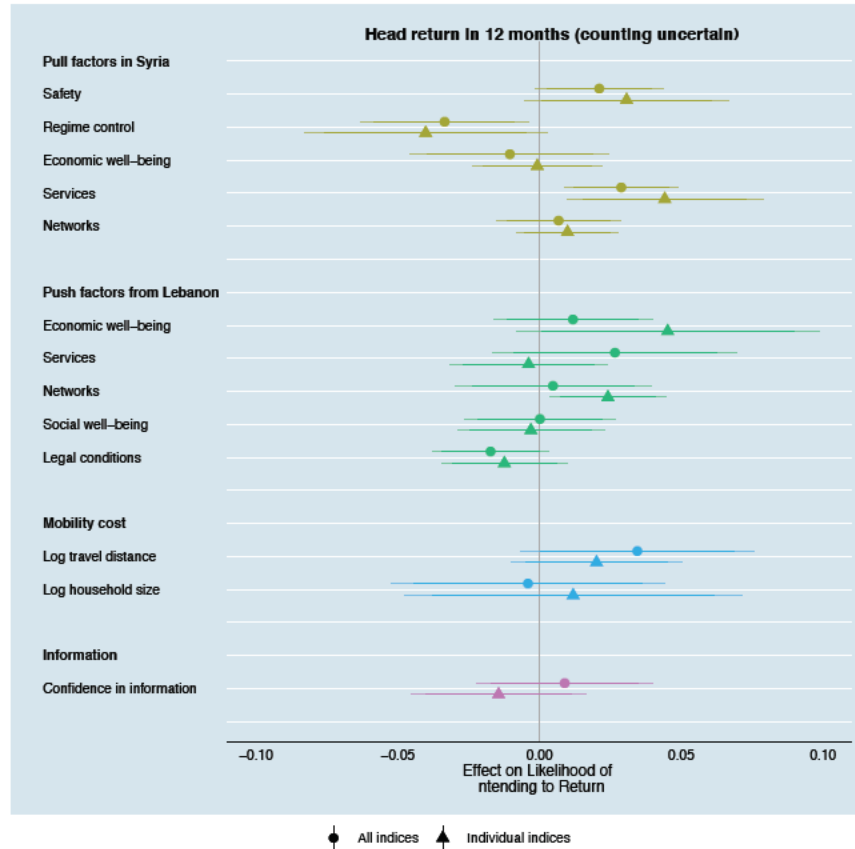

Figure 17: Predictors of return intentions including people uncertain about their intentions

7.3 Additive Indices

To build indices in the paper, we use the first principal component for the predictors under push factors, pull factors, and information as well as for the preparation to return outcome. In the figure below, we present the same results but using mean effects indices (z-scores). We constructed those indices by standardizing each vari-

able (demeaning and dividing it by the standard deviation). We then summed the standardized variables and then standardized the sum.
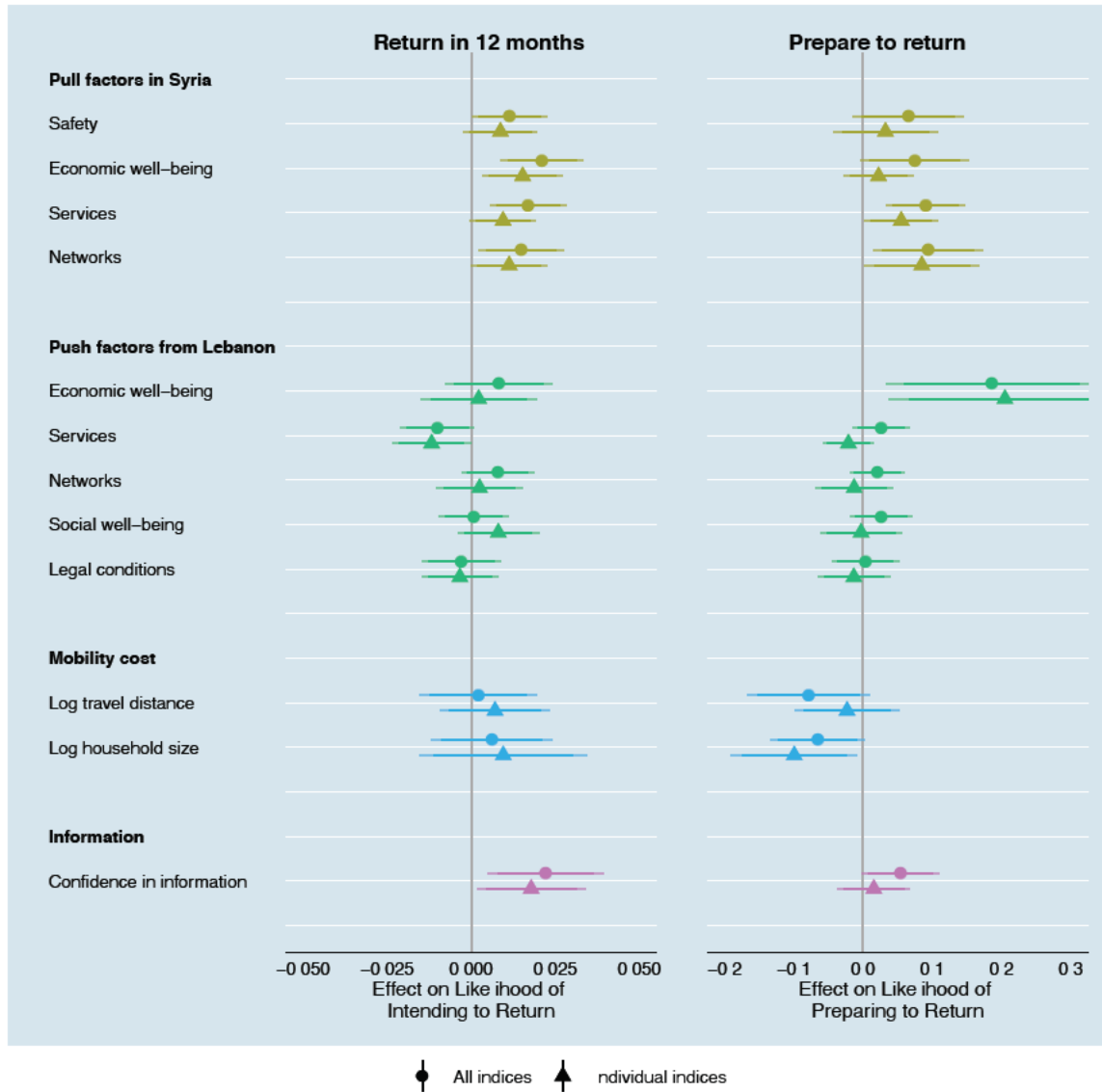

Figure 18: Predictors of return intentions using z-scores to build indices

7.4 Excluding locality fixed effects

We test for the robustness of our results by re-running our models, but without controlling for locality fixed effects.[1] This test addresses the concern that if Syrians in Lebanon face variation in their living conditions between, but not within, localities, controlling for locality fixed effects could prevent us from detecting important relationships between push factors and return intentions. The results, however, do not support this concern. The results shown in Figure 19, are very similar to those from our main model.

The following figure presents the main results for individual indices (where we control for some covariates but include only one index in each regression) without including locality fixed effects in Lebanon. Note that we do not present the all indices regression here (where we control for all indices in addition to covariates) because those regressions do not include locality fixed effects (as these have high correlation with the travel distance predictor)

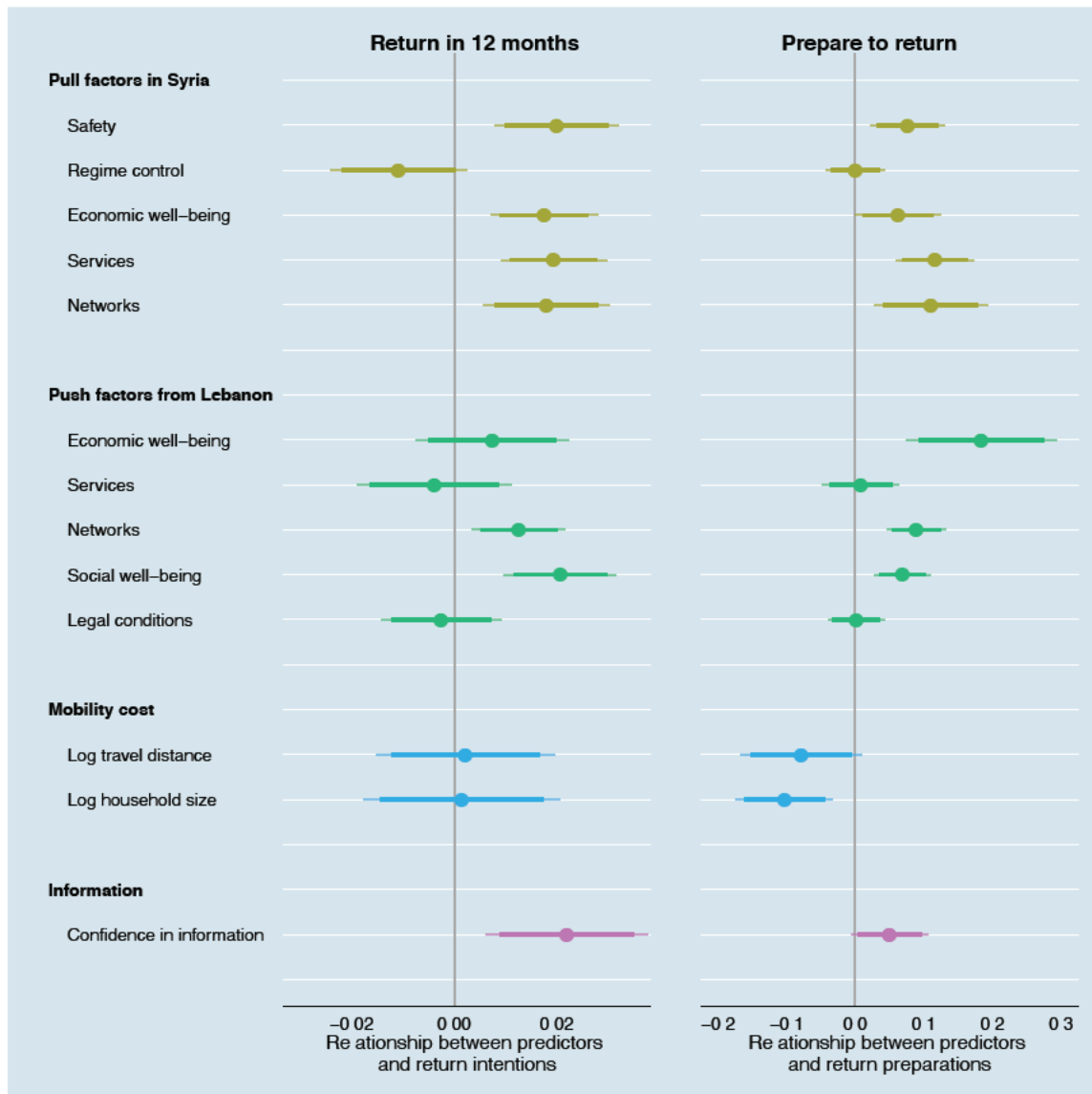[1]This robustness check was not pre-specified.

Figure 19: Predictors of return intentions after excluding locality fixed effects in Lebanon

7.5 Information interaction using Rubin's rules

In the paper, we presented the figure that includes interacting information confidence with conditions in Syria. To obtain the confidence intervals, we used the 97.5th (95th) and the 2.5th (5th) percentiles of coefficients across all bootstraps and imputations. An alternative way of obtaining confidence interval is by finding the empirical variance of coefficients across bootstraps within each imputation then finding the pooled variance across all imputations using the rules of Rubin (1987). Here, we present the results from this second approach.
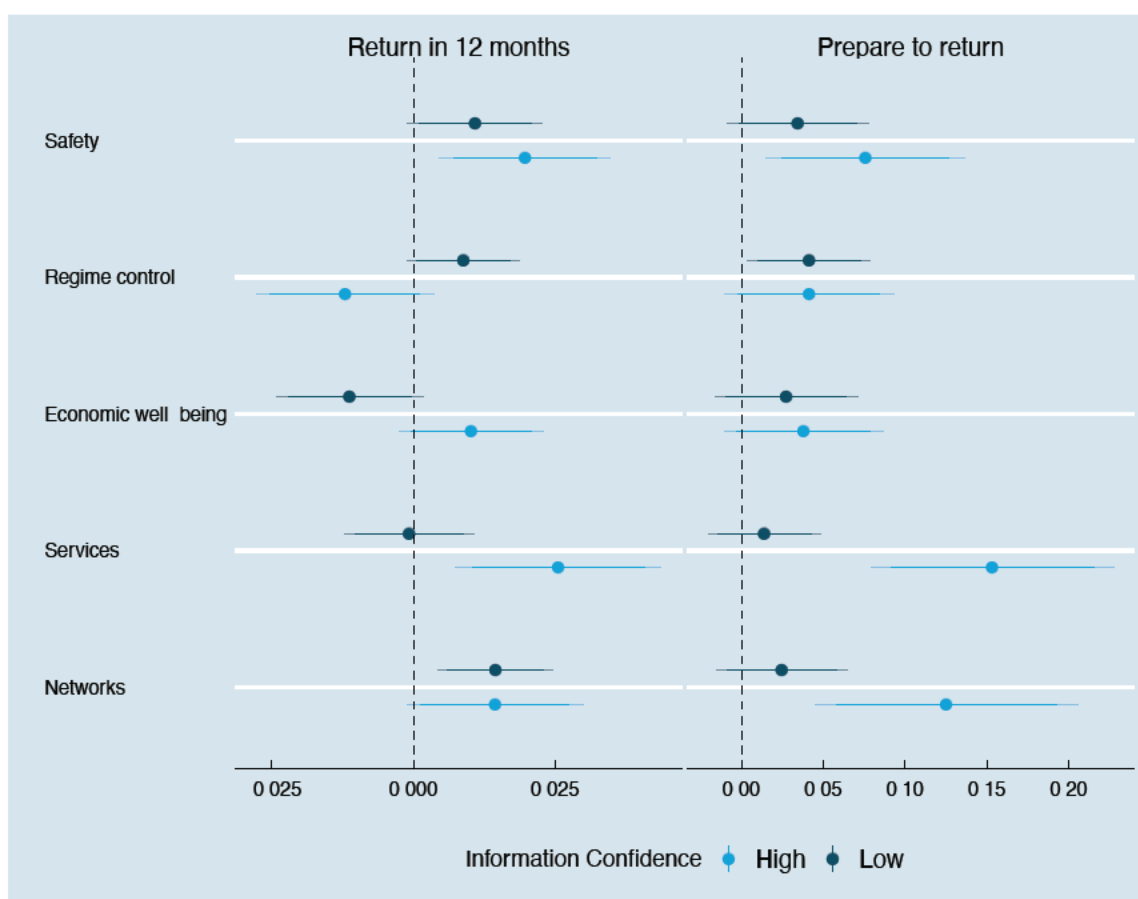

Figure 20: Interactive effects of information on intentions and preparations to return

7.6 Testing the models' predictive power

Below we present prediction plots, including OLS (same models as in the PAP) and lasso, with AUC results for ROC and PR. PR is often as a better performance metric than ROC for predicting rare outcomes. This is due to the fact that the ROC allows for relatively "good" performance by predicting all zeros, which is not the case with PR plots.

Looking at the results in Figures 21 24, we see that the trends are consistent with our main findings but the differences across models are small. Looking at the ROC plot, we witness a ∼2.5 percentage point increase (∼5%) comparing the push model to the pull model, and a ∼6 percentage point increase (∼9%) moving from push to full (i.e., push + pull).[2] Looking at the PR curves, we see a ∼1.5 percentage point increase (∼17%) comparing the push model to the pull model, and a ∼2.5 percentage point increase (∼29%) moving from the push model to the full model (i.e., push + pull). The gains in terms of PR AUC are large in percentage terms, although not absolute terms.

The results suggest a few key takeaways about the predictive power of the models presented in the paper. First, the Syria model is a better predictor than the Lebanon model, aligning with our main results Second, the Syria+Lebanon model is the best predictor, suggesting that push factors are relevant, just less than pull Third, the gains across models are large in percentage terms, but not in absolute terms. Lastly, overall predictive power is low and prediction is hard in our case, possibly due to studying a rare outcome, but also highlighting that understanding the aggregate drivers of return will not necessarily allow policymakers to make reliable predictions about whether an individual household will return or not.

Another consideration for testing the predictive power of push and pull factors would include expanding the sets of inputs variables. The models run in 21 24 use

---

[2]We indicate that these differences are approximate since their precise magnitudes will vary across different simulations.

the index inputs in the PAP as predictors. These indices, however, were intended for marginal effects estimation of coherent concepts, not aggregate prediction. Therefore, we also run the prediction exercise with an extensive set of push/pull predictors this time using all the survey data we have about push and pull factors, not just the index inputs we defined in the PAP.

Figures 25 28 show results from analysis with all survey questions about push factors and pull factors. First, we see that absolute predictive performance is better with the more extensive predictor set. Second, we see that the relative performance of pull factors over push factors is similar to what we observe with the limited set. Results still align with our paper main takeaways. Also, we still see small to moderate gains in predictive power when comparing push and pull models, similar to what we saw in the Figures 21 24.
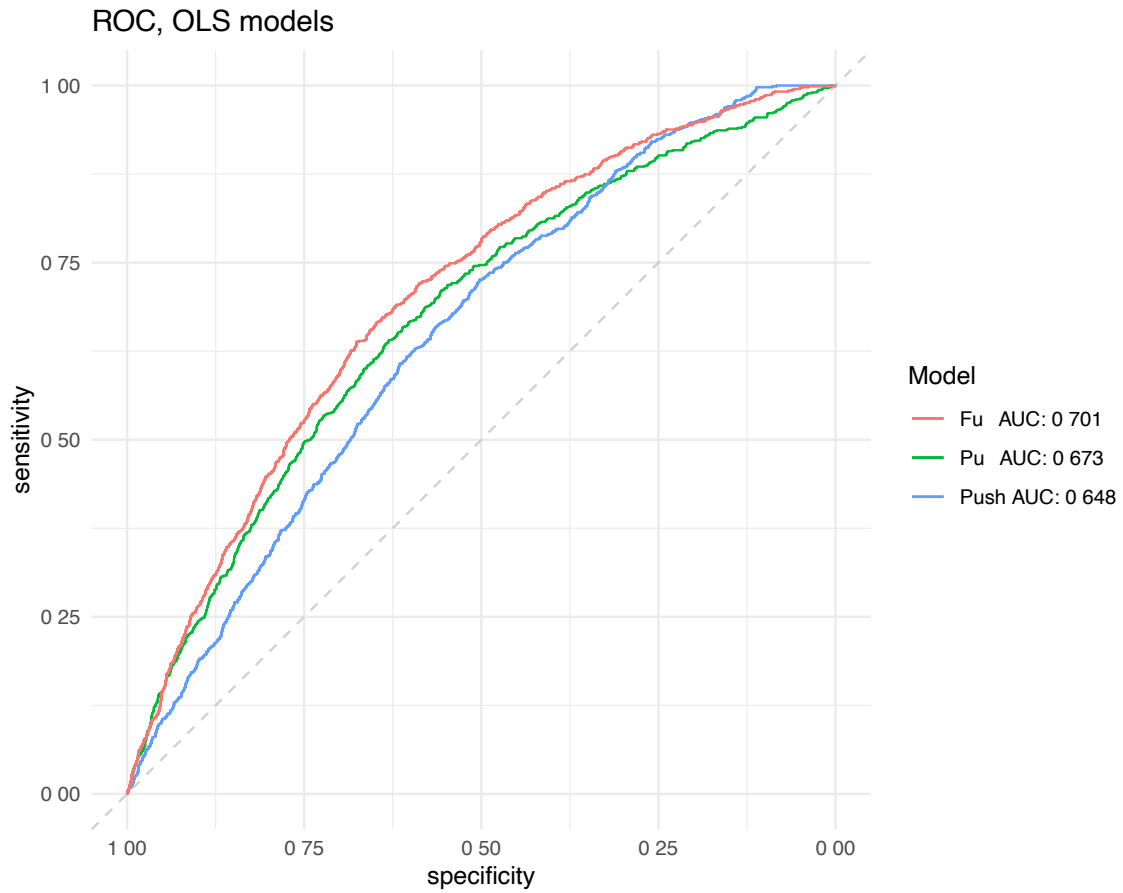
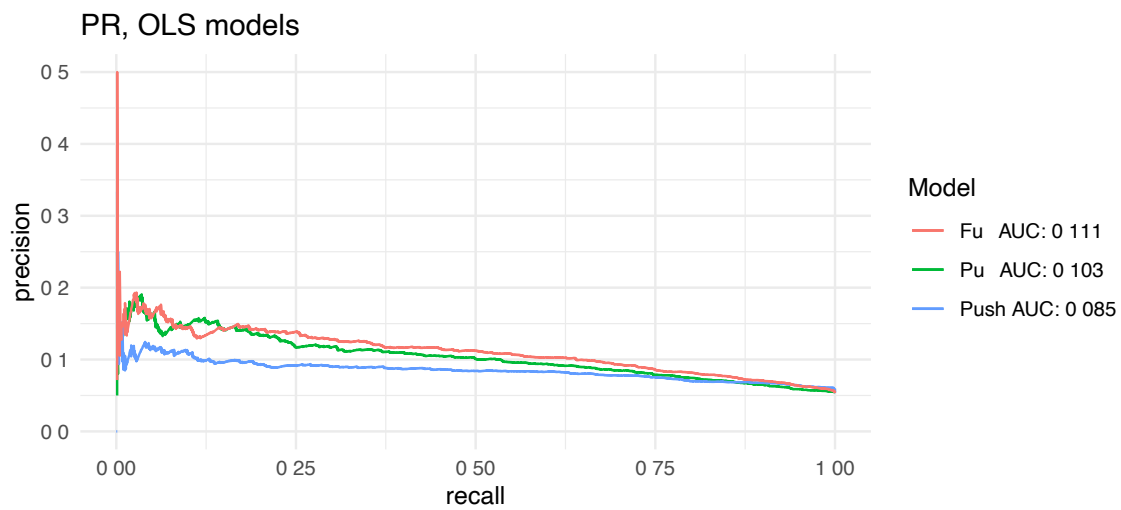Figure 21: ROC OLS models, pre-specified indices as predictors

Figure 22: Principal-response curve   OLS models, pre-specified indices as predictors
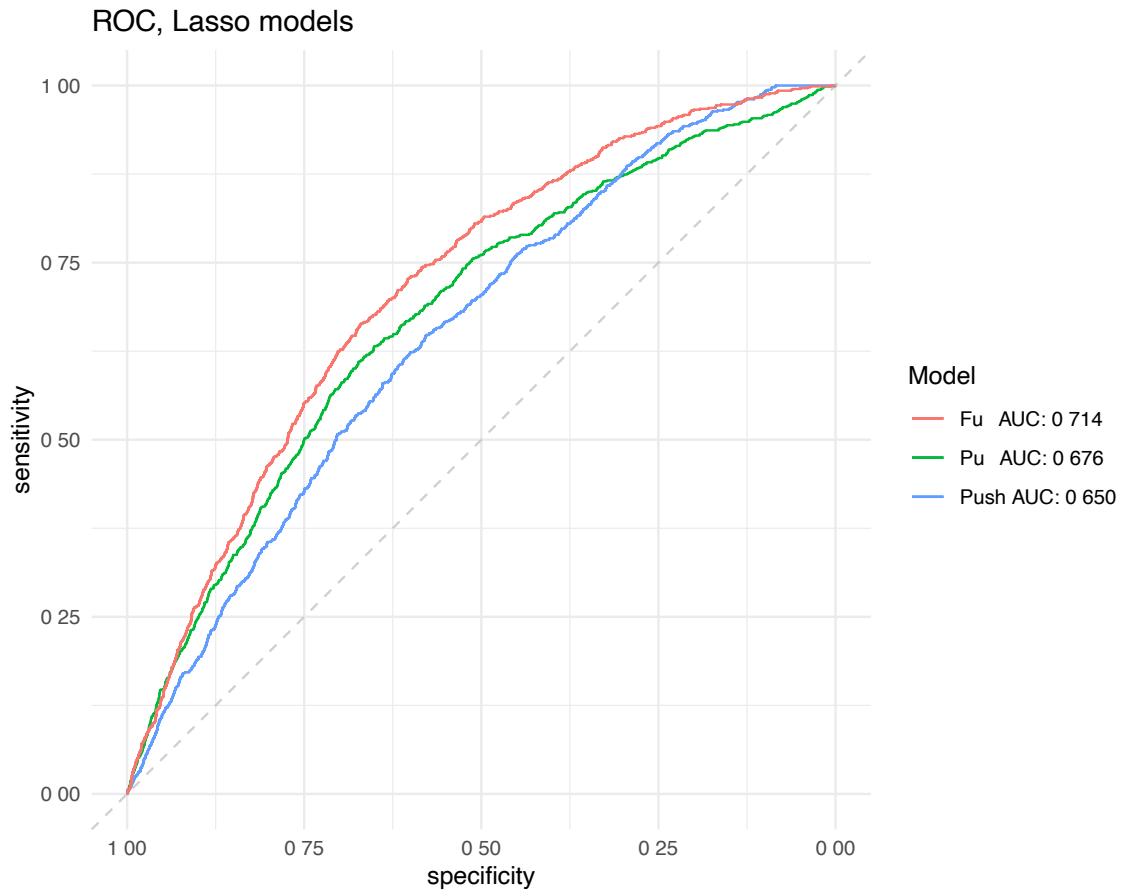
Figure 23: ROC   Lasso models, pre-specified indices as predictors

Figure 24: Principal-response curve   Lasso models, pre-specified indices as predictors

Figure 25: ROC   OLS models, extensive predictor set

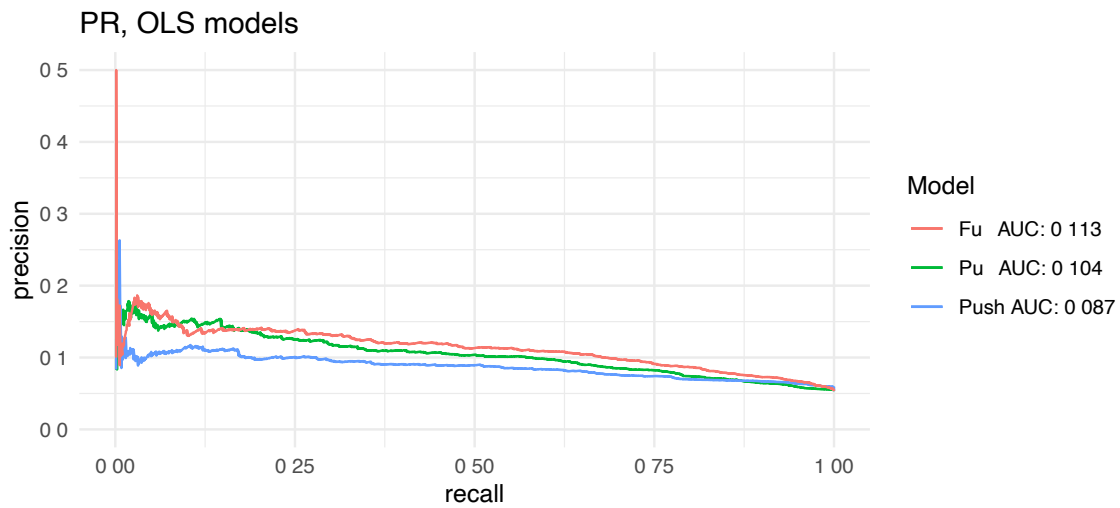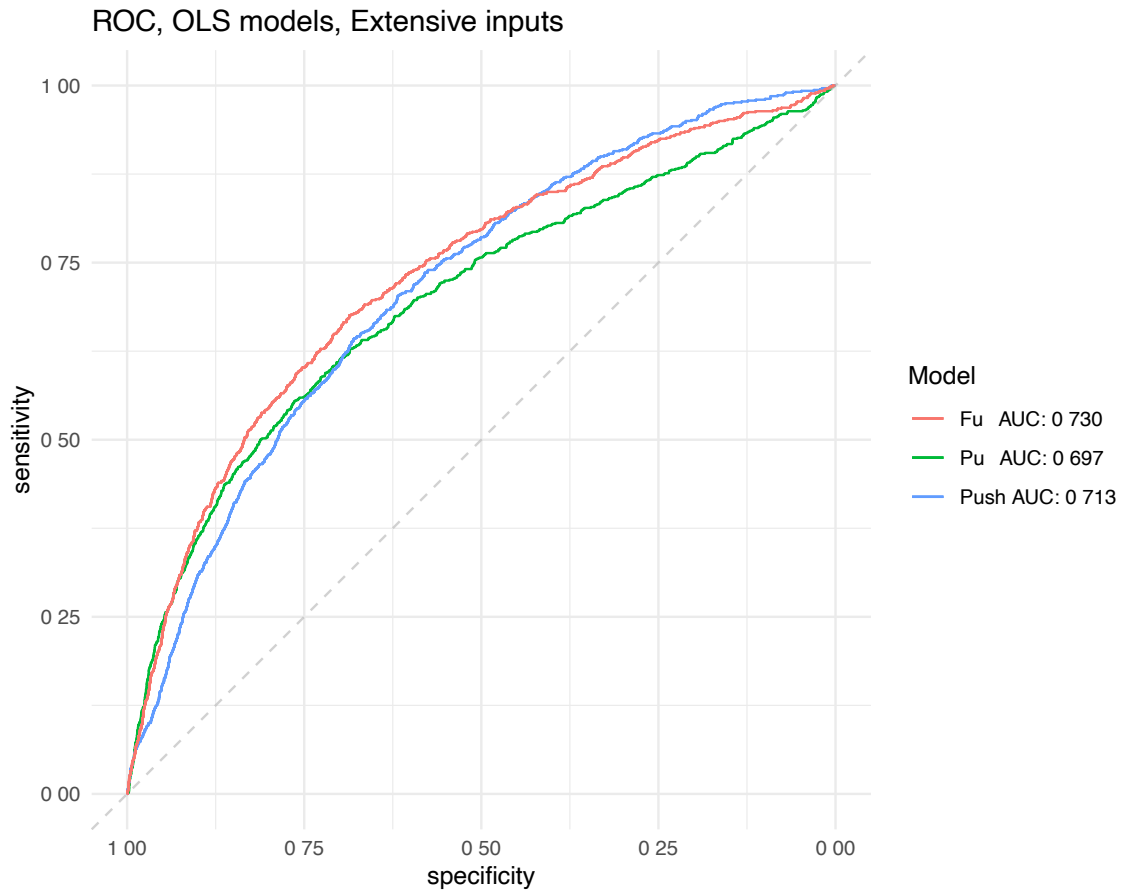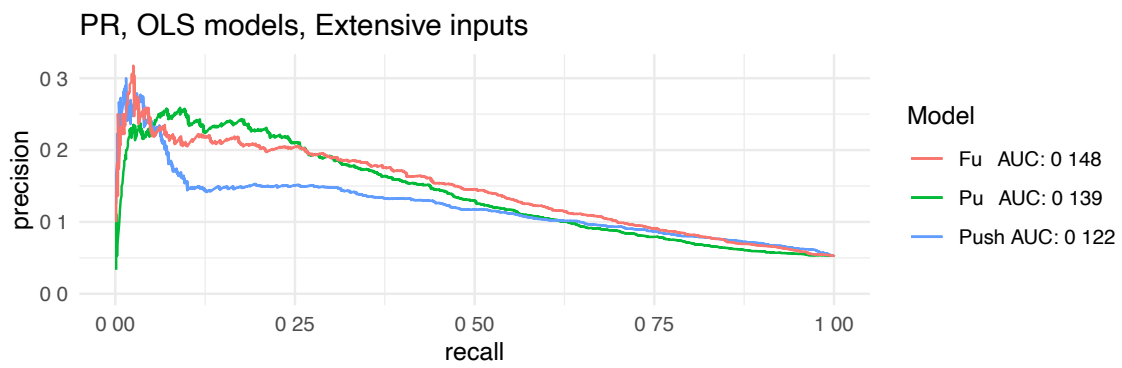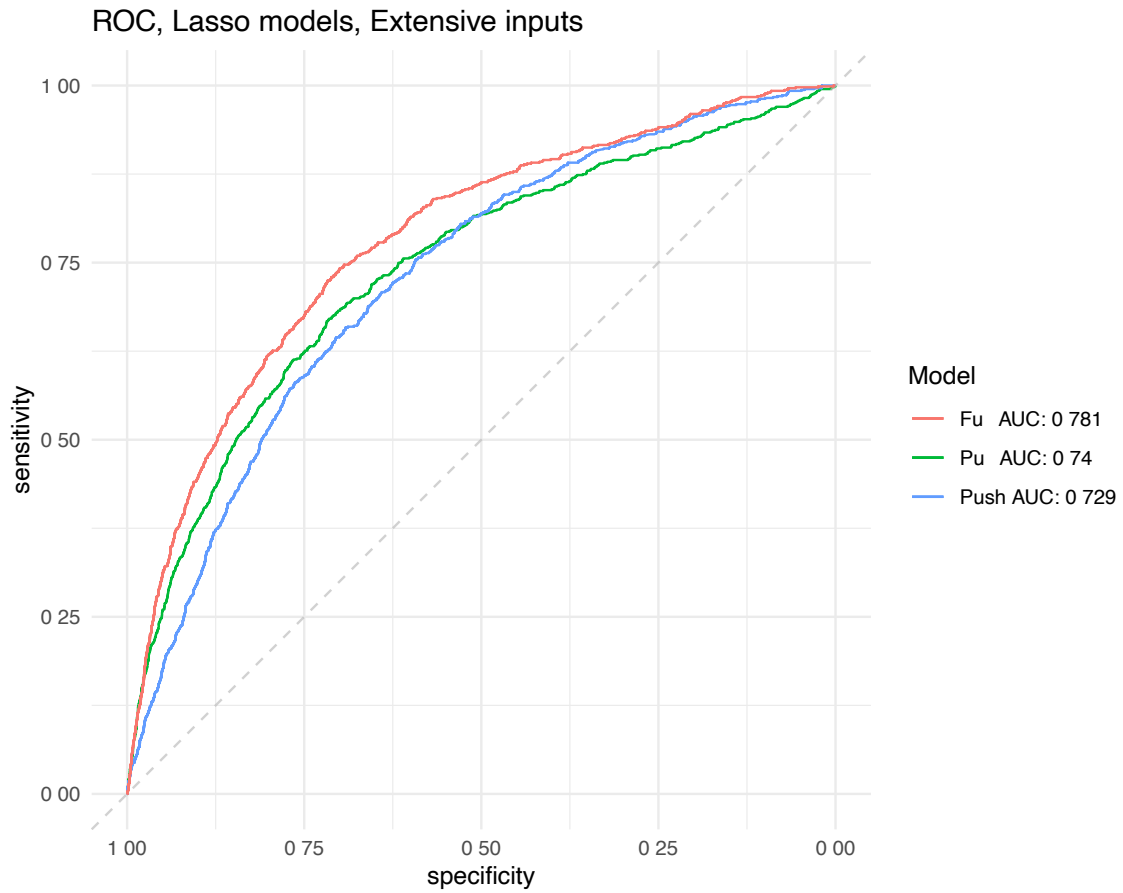Figure 26: Principal-response curve   OLS models, extensive predictor set

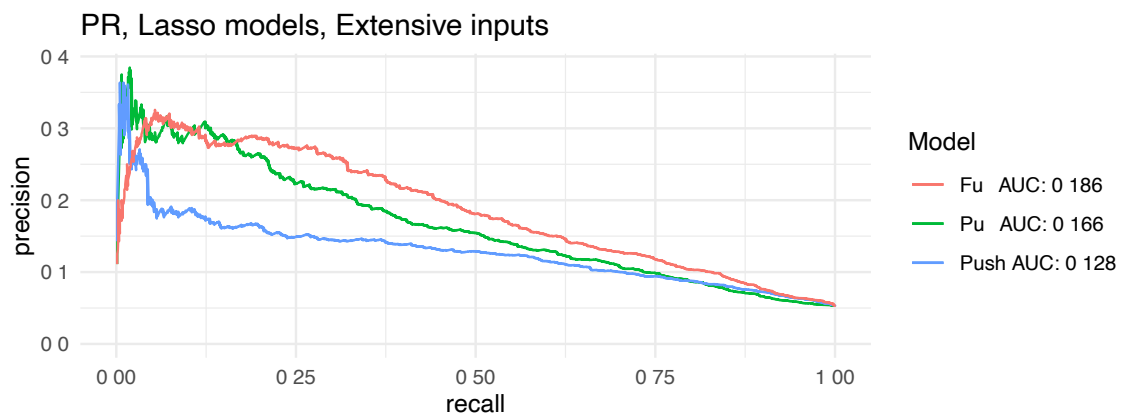Figure 27: ROC   Lasso models, extensive predictor set

Figure 28: Principal-response curve   Lasso models, extensive predictor set

7.7   Multicollinearity

In the pre-analysis plan, we intended to run regressions including all indices as predictors in the same regression model. In the main paper, we present regressions

with each predictor (and controls) individually. Therefore, we present results from the pre-specified analysis here. Overall, results align with the findings in the main paper, although with some differences that are likely driven by multicollinearity.

We made the choice to present an alternative regression after tests of model performance identified multicollinearity in our regression predictors. The first simple test of model performance involved examining the simple pairwise correlations between our predictors. We find that the pairwise correlation between the security index and services index in Syria is high at 0.57.

Of course, simple correlation is not multicollinearity. Therefore, second, we tested for an association between predictors conditional on the other variables in the model. The variance inflation factor is a measure to analyze the magnitude of multicollinearity of model terms. Using the `performance()` package in R, we run all provided tests for multicollinearity. Across multiple tests we find evidence of multicollinearity.

We find that in the pre-specified models we identify very high variance inflation factors for a number of indices, most notably services in Syria and regime control. Based on the `performance()` package output, the Farrar Glauber Chi-Square test, Theil's Method, and the Condition Number test indicate the presence of multicollinearity.

In Figure 29, we present the results from the pre-specified regression, which includes all indices as predictors in the same regression model. Looking at results from this all-indices model and comparing them to the results in the main paper, overall the results align with our main findings. The important difference from single-index results is that two Syria indices (safety and services) get smaller point estimates and are no longer statistically significant. This difference between the individual-index models and the all-indices model aligns with the evidence of correlation and multicollinearity above. That is, the variables that exhibit evidence of multicollinearity exhibit different point estimates when we include them in the same model, suggested that this difference is driven by multicollinearity. Beyond the differences due to mul-

ticollinearity, the results from the pre-specified model align with our findings in the main paper.
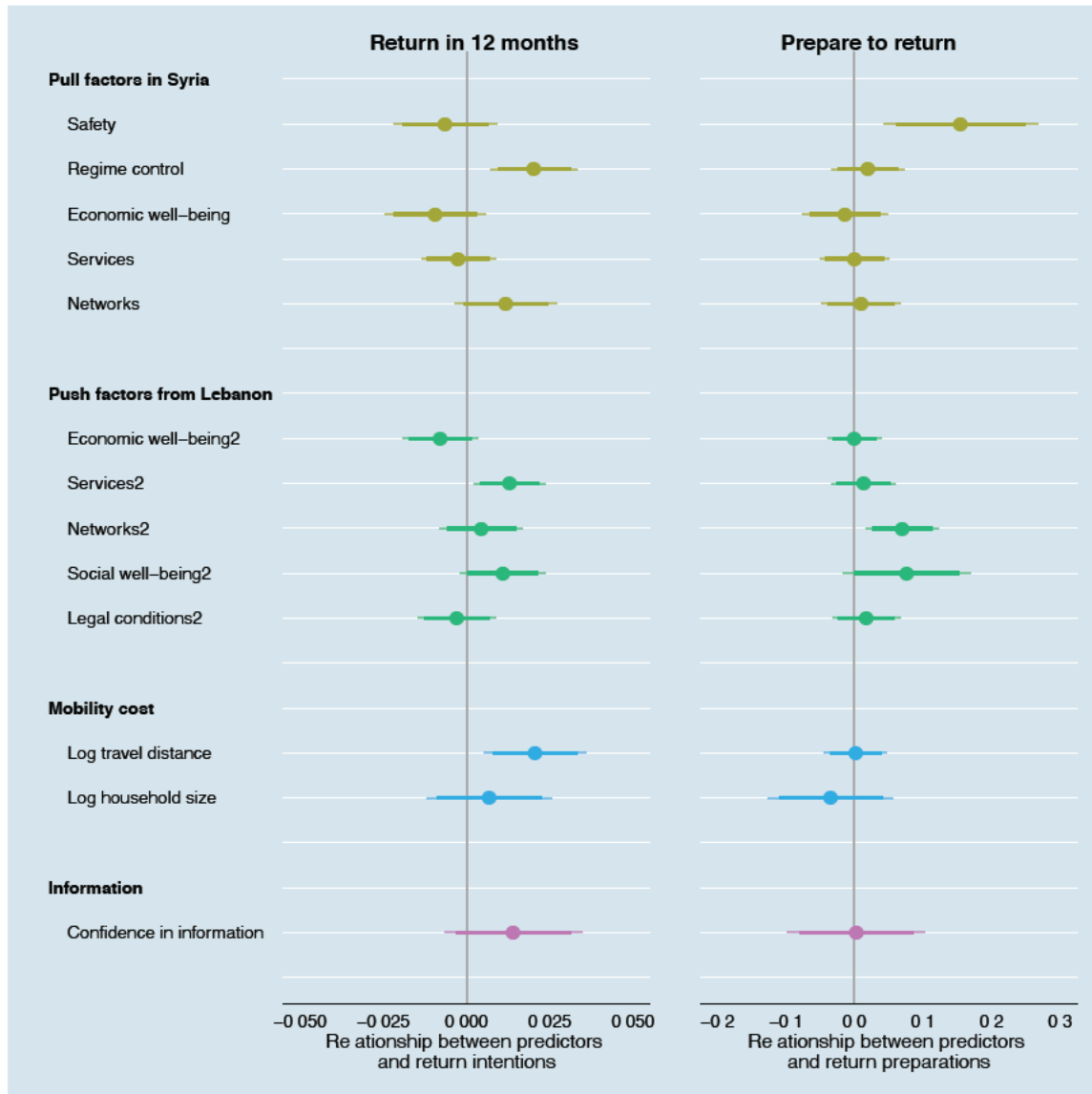

Figure 29: All-indices regression

8 Ethical Considerations

The ethical imperative to do no harm is especially pressing in research with refugees, given their extreme vulnerability (Masterson and Mourad, 2019). As with much research in dynamic and unpredictable conflict settings, the consequences of researchers' choices in such settings cannot be wholly anticipated, and the political sensitivity of particular issues can change rapidly and in ways that make it more challenging to know what is safe to collect and publish (Knott, 2019).

We designed this project to reduce potential harm, maximize policy relevance, and increase opportunities for direct benefits to research participants. The achieve the first two goals, the authors drew on exploratory fieldwork, interviews with international and local humanitarian actors, and a research planning workshop with the humanitarian community in Beirut during which we discussed our research plan and questionnaire to minimize potential harm and ensure that the design can provide the humanitarian community with required evidence to fulfill their needs. To address the first and the third goals, the research team partnered with NGOs to provide protection training to enumerators and established a referral mechanism through which research participants in need of humanitarian services were connected to available resources.

Below we begin by discussing consent and then discuss compensation. Then we discuss potential benefit(s) to be gained by the participants and/or by society as a result of this study. Then we discuss potential risks and responses in designing our research. We approach each risk through a standard risk assessment approach, involving a consideration of the magnitude of potential consequences (levels of impacts) and the likelihood (levels of probability) that such consequences occur.

8.1   Consent

Obtaining meaningful informed consent was an essential step in our strategy for protecting participants. Relying on each participant's sense of their own risks is an essential complement to our risk-benefit assessment, discussed in detail below, given that potential participants certainly understood details of their own risk profile that we could not know a priori.

Upon arriving at potential participants' households, before beginning a survey interview, enumerators read the consent script, explaining the goals and structure of the study. Enumerators, who were native Arabic speakers, read the consent form out loud to respondents in their native language. We trained enumerators to offer clarification of any complex concepts to ensure meaningful consent for participation in the research. The consent form was also printed and provided to respondents. It had a contact number and email for IRB as well as a WhatsApp number for a hotline used by the research team. This number was offered to respondents to answer any concerns or questions respondents had.

The research team obtained verbal consent rather than signed consent. There were two reasons for not asking participants to sign anything. First, a large share of Syrian refugees are illiterate or semi-literate. We deemed asking people to sign a form that they cannot read as likely to cause anxiety and mistrust. Second, even among Syrians who can read, asking refugees with precarious legal status to sign formal documents was likely to create anxiety and fear about what exactly the document states and how it might be used.

Only after obtaining consent from the respondent did the enumerator begin the survey interview.

8.2   Compensation

Survey interviews took about 30-40 minutes to complete and we provided survey respondents with $10 cash compensation for their time. We decided to compensate respondents following extensive interviews with humanitarian actors, as many humanitarian actors suggested that it is only fair to compensate participants for their time following a somewhat long survey. Cash is a common means of compensation for research participation for Syrians in Lebanon. The survey company had used cash compensation many times in research with Syrians in Lebanon and did not foresee likely security, ethical, or practical challenges with cash compensation. We are not aware of this project's cash compensation causing any reportable incidents. Cash compensation was appropriate in this context with widespread access to markets and stores for all Syrians in Lebanon. Furthermore, $10 is an appropriate amount, corresponding to about half of what many Syrians in Lebanon made for a day's work as a day laborer at the time of the study (ACTED 2014, Lehmann and Masterson 2020, p. 49, and our survey data).

8.3   Potential benefits

The importance of this research extends to the humanitarian realm. Facing declining aid flows and the potential for significant movements to return, the humanitarian community is beginning to think seriously about how to engage in the return process.

However, during over a dozen meetings with INGOs, NGOs, and CSOs in Lebanon and Jordan, many complained about the knowledge gap related to the decision making of refugees when it comes to return. This has constrained the ability of the humanitarian community to design effective responses to assist the safe and voluntary return of refugees. In our research project, we will work closely with humanitarian actors in order to meet their urgent need for empirical data about the dynamics and challenges around return.

To strengthen the project's public good contribution, we closely consulted with humanitarian actors at all stages of the project. At the outset, we worked with humanitarian actors to conduct focus groups and meetings with refugees in Jordan and Lebanon. We also consulted with humanitarian actors closely on the development of the questionnaire. Prior to data collection, we conducted a workshop with humanitarian actors in Beirut, Lebanon (in March 2019) to present the whole project and get their feedback on specific aspects, including the questionnaire. We then revised the questionnaire based on their comments (for instance, we removed questions that directly measured political attitudes, which they suggested were too sensitive and inappropriate for the context). In June 2020, we conducted several online workshops with humanitarian actors in Lebanon to share our results and answer any questions about information that humanitarian actors required.

8.4   Referrals

When we presented the research design and questionnaire to humanitarian actors in March 2019, they suggested that we develop a referral strategy for refugees who require or ask for help. The goal of referrals is to help facilitate refugees' access to services by either 1- putting individuals in need of services directly in contact with the service providers or 2- enable people to seek assistance and support them in receiving assistance. Using standard operating procedures of humanitarian actors in Lebanon, we developed specific guidelines for the enumerators to use in order to refer respondents who require assistance or who ask for assistance. Enumerators and team leaders who participated in data collection received special training from a humanitarian actor in Lebanon on the goals of referrals, when to refer respondents, and how to refer respondents. In addition to the training, enumerators received the following referral guidelines document:

8.4.1   Referral guide

The following subsection is a copy of the referral guide that enumerators received and were trained on.

Purpose   As this study concerns a particularly vulnerable population, we want to support our respondents and the humanitarian community by referring people in extreme cases who require urgent assistance or face threatening situations. We do so through developing a referral process that connects this population with specialized agencies that can assist them. This document concerns all those involved in the data collection, including the enumerators and the research team. It uses the existing standard operating procedures for humanitarian actors in Lebanon to develop a referral guideline that utilizes existing mechanisms in Lebanon. The document explains how to refer respondents who request action to facilitate their access to services or those who face immediate threat, danger, and lifesaving situation to the specialized agencies.

Referral involves passing the background information and needs of individuals to the relevant service providers. It does not include providing general information on available services or telling respondents how to independently access services, although we include these steps for people that require referrals. Referral occurs after respondents request access to services and then consent to the provision of their information to service providers. Upon obtaining consent, enumerators will fill out a short form, which will be added to the end of the survey instrument. Finally, the enumerators will provide respondents with our contact information in case respondents want to follow up with us regarding the referral process. We will forward the cases that require referrals to the specialized agencies and then follow up with them to check on whether referred individuals have received assistance. Throughout the process, all those involved should maintain confidentiality and refrain from sharing any information regarding the people requiring referrals outside of the

referral process.

Who gets referred? Many Syrian refugees in Lebanon are vulnerable and require assistance. At the same time, service providers are often overwhelmed and work beyond their capacity. To balance these two issues, referrals should be conducted only in two cases as recommended to us by humanitarian actors in Lebanon:

1. When respondents directly ask for referral to an organization.

2. When enumerators observe a threatening, dangerous, or lifesaving situation.

Only in these cases should enumerators refer individuals.

The referral process

1. Observe: The enumerators should only provide people with the option of referrals in one of two conditions:

   (a) if the respondents explicitly request to be referred.

   (b) if the enumerators observe a threatening or dangerous situation.

2. Consent/assent: At the end of the survey, there will be a question on whether the respondent needs to be referred. If the respondent satisfies one of the two conditions above, the enumerators should select the option to indicate that the person needs referral.

   Respondents will be asked to consent to be referred. The enumerators will read a prompt to obtain consent from the respondents. In the consent form, the enumerators will provide a phone number for the research team.

   The enumerators should **not collect the respondents' information without obtaining consent**. If the respondents do not consent to having their information shared with the service providers, the enumerators should only

provide them with information on how to contact these services as outlined in step 2 above.

3. Fill out information: Once consent is obtained, the tablet will ask for the required information. This information will consist of the following:

- Demographic information: name, address, phone number, UNHCR case number, sex, age

- If the respondent is a child (less than 18 years old): Name of primary caregiver (usually a parent), relationship of primary caregiver to child, address, phone number, and whether the caregiver is informed of referral (if not, then explain why)

- Specific need, which consists of the following categories:

  - **Child protection**: Includes children with specific needs, victim of/ at risk of abuse/neglect/violence, taking care of siblings alone, head of household, separated/unaccompanied, engaged in the worst form of child labor e.g. street-based work, exploitative work, physically dangerous work, etc. at risk or victims of child marriage.

  - **Safe spaces and psychological support** (GBV prevention and response services): It can include victim/ at risk of physical and/or psychological violence, abuse or neglect, exploitation, early marriage etc.

  - **Legal**: It can include family members arrested / detained / issued with departure order, in need of mediation with the landlord/ employer / service provider or in need of individual support on legal issues (i.e. residency, birth or marriage registration).

  - **Persons with specific needs**: It can include elderly (> 65 years old) unable to care for self, single parents caring for dependents ($<$18,

including older people), person with a disability lacking a caregiver and/or unable to care for self and individuals facing specific protection risks (e.g. risk of removal, harassment by community members/authorities, other).

– **Basic assistance/food assistance**: It can include lost PIN for cash assistance card, lost card, mistreatment by bank staff, card malfunctioning, concerns related to exclusion from cash/food assistance, in need of Core Relief Items/Non-Food Items (e.g. family tent, jerry can, sanitary material for females).

– **Health**: It can include individuals in need of specialized mental health services and individuals in need of hospital care (not otherwise covered by UNHCR/NEXtCARE, which deals with emergency situations).

– **Education**: It can include a child not attending school or at risk of dropping out from school, or community learning spaces, rejected enrollment of child by public school. It also includes children and youth who have special needs in need for learning support or referral to specialized services.

– **Shelter**: It can include new arrivals/homeless/eviction cases with no shelter, bad shelter conditions in informal settlements, residential and non-residential structures, heavy flooding/inundation in informal settlements and destroyed shelters due to fire or natural hazards.

4. Decide on referral priority: At the end of the form, the enumerators will decide on the referral priority. There are two priority types here:

- Fast track, which requires following up to be within 24 hours. This is only reserved for cases when there is an immediate threat, danger, or lifesaving situation.

- Normal, which requires following up within 7 working days.

5. Provide with our information: Inform the person being referred that if they face any problem accessing the services they can come back to referring agency staff or volunteers. If the beneficiary/individual mentions that he/she is not able to access a specific service try to provide information (if available) on other relevant, nearby services.

6. Research team sends the referrals: The research team collaborated with a humanitarian actor in Lebanon to refer the cases. The humanitarian actor monitored and reported the cases on referral to the research team.

General guidelines Any sensitive and identifying information collected about any individual should only be shared on a need-to-know basis with as few individuals as possible and only for the purpose of providing services to the person concerned, based on their informed consent (see below).

Throughout the process, everyone involved in the referral process needs to respect the following principles:

1. Confidentiality: It is essential to protect information gathered in relation to respondents and ensure that it is made available to service providers alone. This requires that collecting, storing, and sharing this information is conducted safely and ensuring data protection. Respondents decide if, how, when, and to whom information about their cases is disclosed. Upon obtaining consent, those involved in the referral process (including enumerators) should avoid disclosing this information to anyone not directly involved in this process. They should never discuss this information with family, friends, and colleagues.

2. Consent: Referrals should take place only after the individual has given informed consent. Individuals have the right to limit information they want to

share and with whom it will be shared. Informed consent means making an informed choice freely and voluntarily. Informed consent occurs when the person understands the consequences of the choice, and freely chooses to accept the consequences, and is based on equal power relations. Obtaining informed consent means that before any information is shared with others, or any referral is made, in order to be able to make an informed decision, the individual should be given honest and complete information about possible referrals, their implication, and of any risks or implications of sharing information about her/his situation and of any limits to confidentiality.

When a case involves a child, consent would need to be obtained from both the child and the parent or caregiver, unless it is inappropriate to involve the childs caregiver (such as in cases of parental abuse). If children are too young to give informed consent, the enumerators need to ask for informed assent, which means the expressed willingness of the child to participate in services.

3. Respect the individual: Your role as staff, partner or volunteer is to provide information about services available, in order for them to make a free and informed choice. **Under no circumstances should you give counseling or put pressure to the individual to access one or other services.** Respect their decision-making capacities and preferences. **You are not supposed to express your opinion, pass judgment or blame the individual.**

4. Do not make promises or create expectations: Only share information if, based on service mapping, the services exist/are available. The available services will show enumerators if the referring agency cannot guarantee access to the services or the results or quality of the service, and this needs to be explained to beneficiaries. However, it should also be explained that in case of any problem accessing the services, individuals requesting referrals can come back to the referring agency (in this case research team) and can ask for additional

support.

5. Safety and security: Organizations' staff must take actions to ensure the physical and emotional safety of individuals who have experienced or are at risk of violence, abuse, exploitation or neglect. The physical safety of the individual should be prioritized above all other actions or referrals that may be available. Safety and security considerations should also be taken into account when presenting referral options to an individual, to the extent that enumerators can reasonably be expected to be aware of relevant risks.

8.4.2   Summary figure

In addition to the above document, the Figure 30 was included in the enumerator guide:

When and how to do a referral?

A. The respondent requests to be referred.

OR

B. I witness a life-threatening situation.

Can I ask for consent?

Yes

No

At the end of the survey, I inform the respondent about the possibility to be referred and its implications.

I ask for consent.

Does the respondent agree to be referred?

Yes

No

I thank the respondent and end the referral.

Is the person to be referred over 18?

No

Can I ask the caregiver to give consent?

Yes

Yes

No   (child abuse e.g.)

I fill out the basic information form on the tablet.

I hand out the sheet of paper with research team's contact to the respondent.

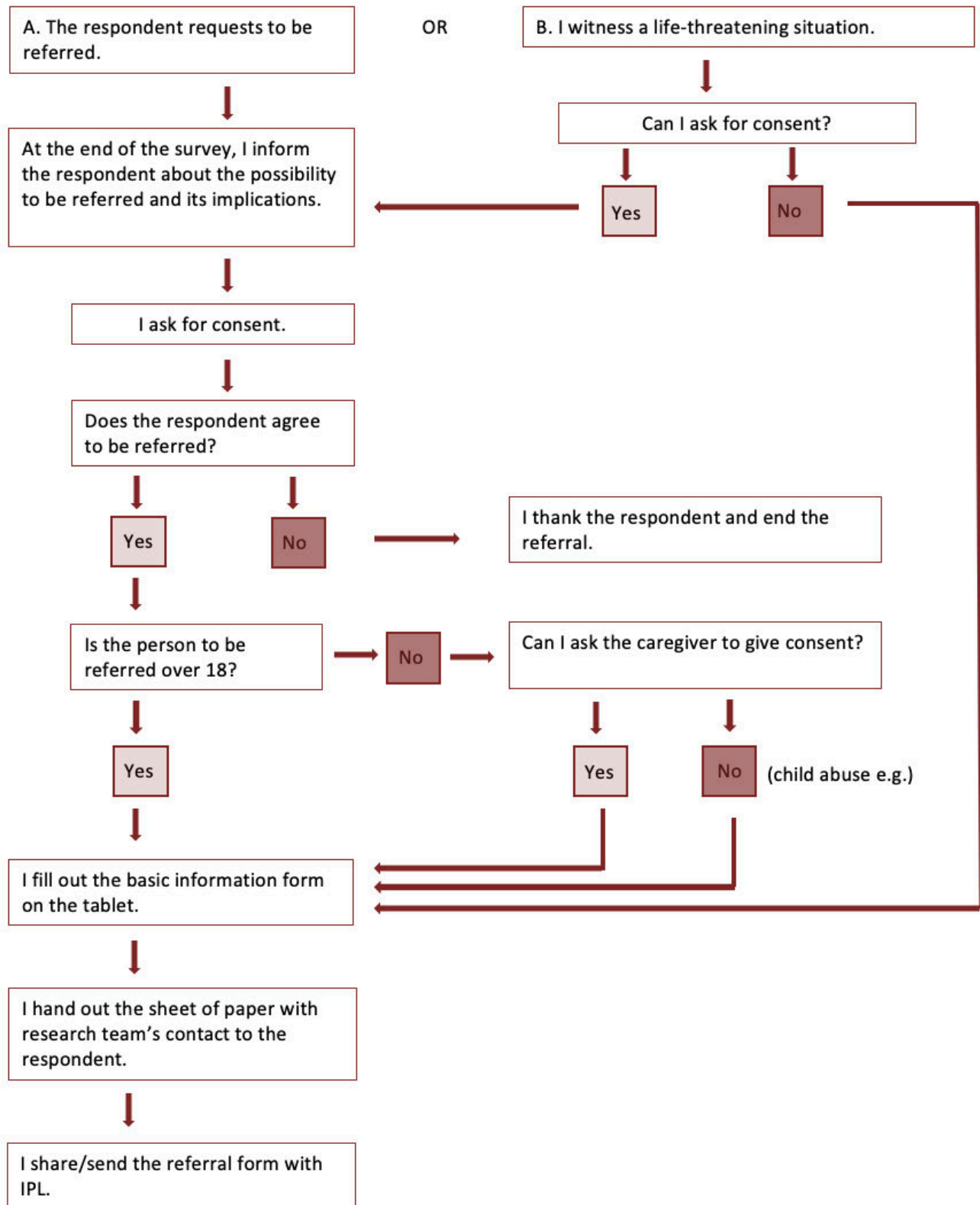I share/send the referral form with IPL.

Figure 30: Referral guide summary

8.4.3 Referral report

Humanitarian organizations in Lebanon collaborate on a centralized referral platform called Referral Information Management System (RIMS), which was developed by the Danish Refugee Council (DRC). This use this in addition to regular referral methods (by reaching out directly to the responsible organization). As RIMS was not available for use by researchers (only humanitarian actors could use it), the research team collaborated with a local humanitarian organization, Amel, to conduct the referrals. Following referrals, Amel sent the research team the following report (The language was edited for clarity):

Following the request of the researchers, Amel received the list of 314 cases who needs to be referred for different sector but mainly to health sector (health problem to one or more of the family member), therefore Amel worked on referring them within one month of receiving the data through RIMS and regular referral methods (email or phone); most of the referred case were through the regular referral methods (70%).

In order to follow up on the cases, five months after the referral, Amel contacted all the list to follow up on their situation:

No follow up	Followed up/need more support	Follow up successful	Phone number changed	Total
119	3	105	87	314
38%	1%	33%	28%	100%

The cases that were not followed up initially and who needed additional follow up (first two columns) were referred again to a different association.

Challenges: the main challenge that Amel faced while referring the cases were 1) Not all NGOs were available through RIMS. 2) Most of the NGOs have limited resources to support and limited equipment and medicines 3) There is a huge gap in the capacity to cover long treatments and expensive surgeries and medical operation and expensive medicines.

Recommendations: this collaboration between the researchers and associations is very important and very useful. This is especially the case as researchers usually are able to reach different groups that are not covered by the NGOs that is why, as Amel we are recommending this for all researchers. It would be very helpful if we try to find a platform or a unified form that connect the researchers and the associations to follow up on the cases the researcher identify during their work. And also, the identification of beneficiaries with need of referrals should become a main outcome of any research or needs assessment done by the associations. And with the current situation in Lebanon, it should be applied to the researcher and groups and initiatives participating in the response to the Beirut Blast (distribution, constructions, etc.).

9   Jordan survey

In this section, we present the set of questions we used to construct each index from the Jordan survey data. As with the Lebanon indices, we constructed these indices by extracting the first components from PCA of the input variables. Some of the questions differ from the wording used in Lebanon in order to fit the Jordan context. Also, due to space constraints in the Jordan survey, the survey did not contain the full set of questions used in Lebanon.

9.1   Safety in Syria

1. How would you describe the risk to civilians physical safety (such as fighting, kidnapping, IEDs, crimes) in your place of origin?

2. What were the main reasons for you to leave your home country Syria? (Violence/bombardment selected)

3. Head of household could be conscripted (age is between 18 and 42 and is male)

9.2   Economic well-being in Syria

1. How would you describe the availability of jobs at present in your place of origin?

9.3   Services in Syria

1. As far as you know, how many hours per day is there electricity in your place of origin?

2. As far as you know, how many hours per day is there running water in your place of origin?

3. As far as you know, are schools operating in your place of origin?

4. As far as you know, are health centers operating in your place of origin?

9.4   Networks in Syria

1. Approximately how many of your relatives or friends in Jordan have gone back to Syria?

2. Approximately how many of your friends or relatives in Jordan have gone back to your place of origin in Syria?

9.5   Information quality

1. Know about the following at home:

   (a) Safety/security news

   (b) Status of infrastructure in a particular location

   (c) News about friends/family in Syria

2. How confident are you in your knowledge about conditions in your place of origin?

9.6 Economic well-being in Jordan

1. Is any member of your household currently working?

2. Are your currently working?

3. Have you received assistance from [government organizations, location organizations, NGOs, UN in the last two months]?

4. If you think back about the situation one year ago in terms of access to assistance, has the situation improved, stayed the same, or deteriorated?

5. Do you or any of your household have a work permit?

6. Have you received food vouchers during the last month?

9.7 Services in Jordan

1. On 24 January [2019] the Prime Minister's office signed a decree informing that Syrian refugees in MOH hospitals and health centers will be requested to pay directly 80% of the applicable 'foreigners rate,' whereas from November 2014, they were treated like Jordanians who did not have health insurance and were able to access health services at subsidized rates. Has it impacted your ability to access health services?

2. Number of school aged children out of school

9.8 Networks in Jordan

1. In the last 12 months, how often did you share a meal with Jordanians who are not part of your family?

2. Please think about the Jordanians in your phone contacts. With how many of them did you have a conversation either by phone, messenger chat, face-to-face, or text exchange in the last week?

3. Please think about the Syrians in Jordan in your phone contacts. With how many of them did you have a conversation either by phone, messenger chat, face-to-face, or text exchange in the last week?

9.9   Social well-being in Jordan

1. How connected do you feel with Jordanian society?

2. How often do you feel like an outsider in Jordan?

3. What is your level of education?

4. When did the first member of your family arrive to Jordan?

5. Does anyone in your family face verbal or physical harassment, meaning verbal or other actions meant to annoy, threaten, intimidate, or make someone feel scared for their safety, in the area around your house?

6. Would you describe the relations with your neighbors as mostly positive/neither positive nor negative/mostly negative?

7. Have you received help from your neighbors?

9.10   Legal situation in Jordan

1. Do all your family members have a valid registration with UNHCR?

2. Do you have a government service card, currently called "MOI Card"?

9.11   Outcome question

- Do you think it will ever be possible to return to your place of origin in Syria?

References

Abadie, Alberto, Susan Athey, Guido W Imbens and Jeffrey Wooldridge. 2017. When should you adjust standard errors for clustering? Technical report National Bureau of Economic Research.

ACTED. 2014. "Labour Market Assessment in Beirut and Mount Lebanon.".

Hainmueller, Jens, Dominik Hangartner and Giuseppe Pietrantuono. 2017. "Catalyst or crown: Does naturalization promote the long-term social integration of immigrants?" *American Political Science Review* 111(2):256 276.

Kling, Jeffrey, Jeffrey Liebman and Lawrence Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83 119.

Knott, Eleanor. 2019. "Beyond the field: ethics after fieldwork in politically dynamic contexts." *Perspectives on Politics* 17(1):140 153.

Lehmann, M Christian and Daniel T R Masterson. 2020. "Supplementary Materials: Does Aid Reduce Anti-refugee Violence? Evidence from Syrian Refugees in Lebanon." *American Political Science Review* 114(4).

Masterson, Daniel and Lama Mourad. 2019. "The Ethical Challenges of Field Research in the Syrian Refugee Crisis." *APSA MENA Politics* 2(1).

Rubin, Donald B. 1987. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.