Predicting Bilateral Refugee Flows: Evaluating the Gravity Model and Ethnocultural Linkages for Migration Policy

Candidate number: XGRX0

Word Count: 7,948

Supervisor: Dr Moritz Marbach

Dissertation submitted in part-fulfilment of the

Bachelor's degree in Politics and International Relations.

Department of Political Science

University College London

May 27, 2024

Abstract

Robust quantitative methods for forecasting refugee movement are essential for policy-makers to adequately prepare for refugee crises and avoid resource allocation inefficiency, poor integration and infrastructural deficiencies. This paper investigates to what extent an "ageing superstar in regional science" (Poot et al. 2016) - the gravity model - can predict bilateral refugee flows. Drawing on intriguing scholarship on ethnocultural linkages and refugee flow networks, this dissertation further asks if making an ethnoculturally sensitive gravity model strengthens predictive performance. Using a dyadic panel dataset from 1989 to 2013, I train two models - one without and one with ethnocultural sensitivity. The latter slightly outperforms the former in out-of-sample prediction. I then test whether the chosen model can accurately predict refugee flow magnitude in three historical cases: the Rohingya refugee crisis, the Rwandan genocide, and the Yugoslav wars. I find moderate-to-strong correlations (between 0.6 and 0.9) between actual and predicted flows, statistically significant at all conventional levels. Finally, I test the model's ability to predict quasi-real-time refugee flows from Ukraine in 2023 to determine readiness for policy application. The results indicate that it is not quite ready, but nonetheless useful, suggesting that a model-mixing approach could be a promising avenue for further research to harness the real power of the gravity model. Even imperfect forecasting offers advantages over a complete absence of foresight.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr Moritz Marbach, for his unwavering guidance, support, and encouragement throughout this dissertation journey. His expert insights, constructive feedback, and critical analysis have been invaluable in shaping the direction and quality of this work. I would also like to thank Dr John-Paul Salter for his informative lectures, seminars and discussions - all of which were instrumental in preparing me for this dissertation journey.

I would also like to thank my family for their unconditional love and support, especially during those challenging moments when I needed it the most. Their belief in me has been a constant source of motivation and inspiration, and the hugs and plates of fruit mid-writing always put a smile on my face.

Finally, I am grateful for the friendship and encouragement of my closest friends and colleagues who have provided me with a welcome distraction from the demands of this project and have helped me to maintain perspective throughout.

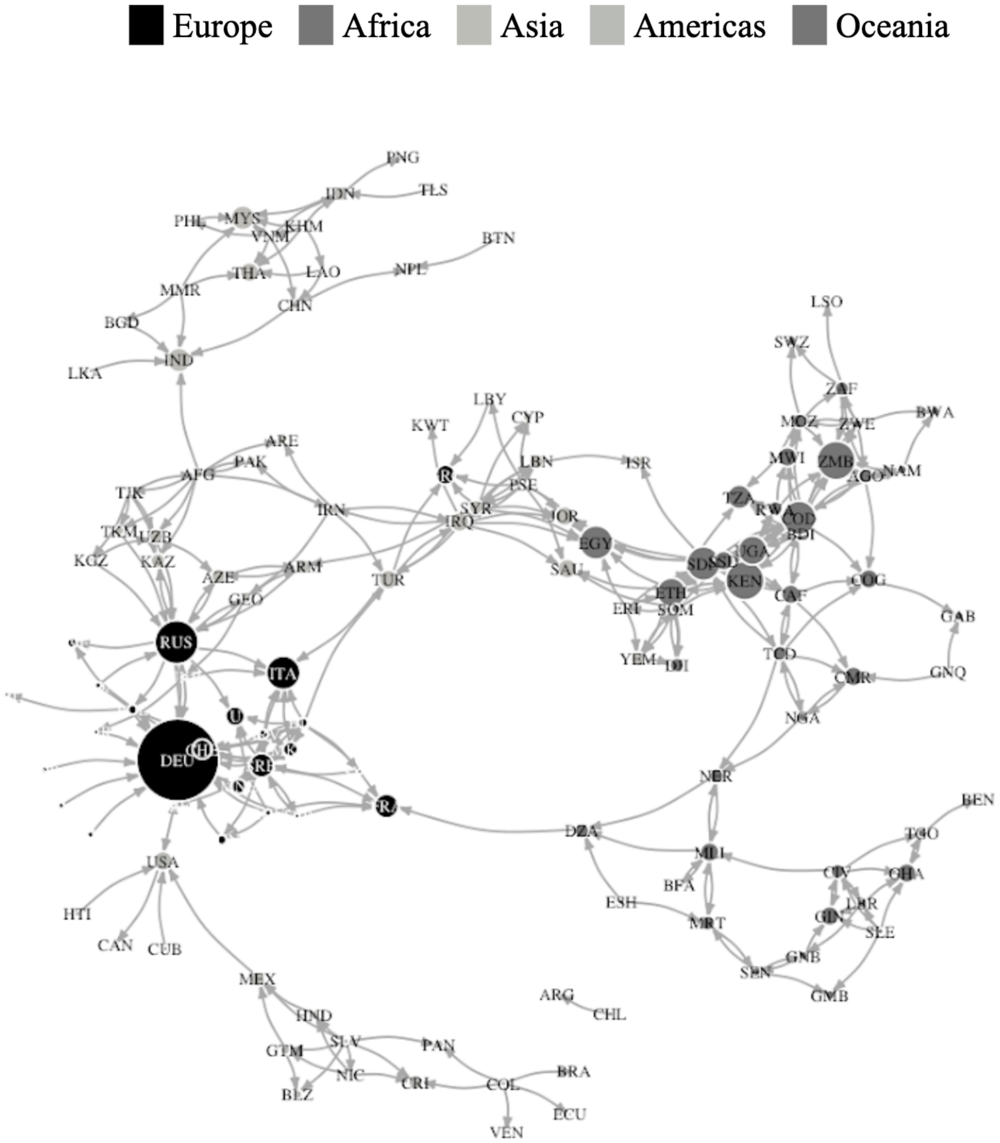
Contents

Chapter 1

Introduction

At the intersection of rising globalisation (Borghesi and Vercelli 2003; Perraton 2019), resource insecurity (Belsey-Priebe et al. 2021; Piesse 2020; Milner 2021), and conflict in the modern world (OHCHR 2021; Roser et al. 2021), forced displacement is an unavoidable consequence. As of 2022, over 100 million people, including 32.5 million refugees, have been forcibly displaced (Concern Worldwide 2023), with the number projected to increase throughout the century (Guo et al. 2020). The detriments of the global refugee crisis on world systems warrant an evidence-based, effective response infrastructure (Shultz et al. 2020, 128), which must include appropriate strategies for measuring and forecasting transnational bilateral refugee flows (Pellandra and Henningsen 2022). This requires understanding the determinants of refugee destination choice. Geographical distance is commonly used as a determinant of refugee flows (e.g. Iqbal (2007)), but the uneven and increasing global diffusion of refugees to countries far beyond their immediate neighbours evidences more factors at play (see Figure 1.1).

Figure 1.1: Transnational Network of Refugee Movements Across Diasporic Ties



Constructed using the Ethncity of Refugees dataset (Rüegger and Bohnet 2018)

Otherwise, it would be counterintuitive for Germany to receive 27% of EU refugees, mainly from Syria and Afghanistan, between 2011 and 2015 (Rüegger and Bohnet 2018, 65). Scholars have attempted to understand the determinants of refugee flows, but there is a need for stronger quantitative study (Rüegger and Bohnet 2018; Stein 1981, 320).

Gravity modelling (GM) has shown impressive predictive power in forecasting regular migration between and within countries (see Mayda 2010; Garcia et al. 2014; Crozet 2004; Beine et al. 2011), and could be applied to the study of refugee movement by integrating the distinct factors that drive it. This model would allow for analysis of the predictive power of factors such as ethnocultural linkages on asylum destination choice, drawing on promising results from Rüegger and Bohnet (2018). Thus, this paper explores the reliability of GM in predicting refugee flight patterns, asking three questions:

1. What is the predictive power of GM as it relates to refugee movement?
2. Do considerations of cultural and ethnic linkages between origin and destination countries improve the capacity of GM to predict refugee flight patterns?
3. To what extent can stakeholders and policymakers rely on GM to inform asylum and refugee management policy?

The paper proceeds as follows: Chapter 2 sets the theoretical framework for refugee flight patterns and GMs. Chapter 3 details my dyadic panel dataset, containing 7,500 unique dyad-years (91 origin countries, 107 destination countries), from 1989 to 2013. It also details the variable selection, the training of two gravity models, and the comparison of their performance. The chosen model is then applied to predict refugee flows in three historic crises, as well as from Ukraine in 2023. Chapter 4 presents and discusses the findings. While the gravity model is reliable for explaining past flows, it requires a model-mixing approach to improve its predictive power for current flows before being useful for policymakers.

Chapter 2

Theoretical Framework

2.1 Background

As the cornerstone of international refugee law (Marshall 2010, 61), the *1951 Convention relating to the Status of Refugees* and its 1967 Protocol define a refugee as someone who cannot return to their country of origin due to well-founded fear of persecution on the grounds of race, religion, nationality, membership of a particular social group, or political opinion (United Nations 1951). Given the proliferation of conflict worldwide, the Convention is as important as ever (Kirişci 2021). Nevertheless, it has been criticised for its inability to protect new refugees, such as those fleeing ethnic violence in Bosnia or Kosovo, and those displaced by factors outside its definition, such as environmental refugees (House 2005, 1; Marshall 2010). Scholars further critique its focus on persecution, arguing that present-day refugees - categorised by (Fitzpatrick 1996, 229-230) into activists, targets and victims - are often driven out by violence lacking a persecutory focus. To accommodate, this paper takes a maximalist definition of a refugee as someone who left their country of origin because of conflict, persecution, or events seriously disturbing public order (Marshall 2010, 63; Rüegger and Bohnet 2018, 68). I note the distinction between refugees and asylum seekers as the former being asylum seekers with an accepted claim (Offe 2011, 166), however as this paper is only concerned with destination choice, I henceforth will refer to both as refugees.

Studying the movement of refugees is made necessary by the numerous detriments to humanitarian systems and state actors of having inadequate forecasting methodology. Being ill-prepared and unable to anticipate arrivals on time is how refugee crises become humanitarian crises (Miller and Chtouris 2017). These determinants are most evident in refugee camps - "temporary facilities built to provide immediate protection and assistance to people forced to flee" (UNHCR 2013). They are crucial in providing protection and assistance but are vulnerable to crises born of inadequate infrastructure. Examples include a fire that destroyed the overcrowded Moria refugee camp on the Greek island of Lesvos (Markham 2022), as well as public health crises in the Cox's Bazar network of camps housing Rohingya refugees during COVID-19 (Gaffar 2018; Guglielmi et al. 2020), and lack of safety and respect for human rights in the Dadaab, Dagahaley, Ifo and Hagadera camps in Kenya (Verdirame and Harrell-Bond 2005). Understanding what drives asylum destination choice can help anticipate flight patterns and prevent overcrowding in the future.

2.2 The Gravity Model

As a classic example of “social physics”, pioneered by John Q. Stewart (1950), wherein tools from mathematical physics are borrowed to explain human behaviour, the gravity model takes inspiration from Newton’s 1687 law of gravity (Anderson 2011). The law holds that two objects exert a force of attraction on each other that is directly proportional to the product of their masses, and inversely proportional to the square of the distance separating them (ibid). In computational social science, the Gravity Model applies this concept to spatial interactions such as international trade, information exchange, and migration. The magnitude of these flows is proportional to the size of the locations involved (usually measured by population or GDP), and inversely proportional to the distance between them, acting as a proxy for transportation costs (Poot et al. 2016; Ramos 2016). The model takes the following form:

$$m_{i,j,t} = G\frac{P_{i,t}^{\alpha}P_{j,t}^{\beta}}{D_{i,j}^{\gamma}} \tag{2.1}$$

Where refugee flows from origin country i to destination country j in period t ($m_{i,j,t}$) are proportional to the product of the ‘size’ of both countries, measured by the population of the origin and destination country at a given time t ($P_{i,t}$ and $P_{j,t}$ respectively), and the geodesic distance that separates them ($D_{i,j}$). G is the gravitational scaling constant. To estimate parameters α, β and γ , the most common approach is to apply logarithmic transformation and estimate using the following form (Poot et al. 2016):

$$ln\, m_{i,j,t} = G + \alpha\, ln\, P_{i,t} + \beta\, ln\, P_{j,t} - \gamma\, ln\, D_{i,j} + \varepsilon_{i,j,t} \tag{2.2}$$

While mostly used in economics (Ramos 2016), this model has featured significantly in the early works of Ravenstein (1885, 1889, in Poot et al. 2016) who identifies gravity-like properties of migration into the UK, and Zipf (1946) who applies a gravity approach to analyse U.S. intercity migration. It continues to do so today; for example, Karemera et al. (2000) use it to analyse migration into North America, identifying origin countries' population and destination countries' income as significant drivers. Cohen et al. (2008) and Kim and Cohen (2010) use it to deduce that geo-demographic characteristics impact destination choice more than socioeconomic/historic factors. Afifi and Warner (2008) and Backhaus et al. (2015) expand the GM to show that migration is also influenced by environmental factors such as overall environmental degradation and climate. Poprawe (2015) does so too, instead to show the effect of political parameters. Conversely, Beyer et al. (2022) criticise the GM for only being able to capture cross-sectional (between-country) variance, and failing to capture basic temporal dynamics according to standard validation techniques. To assess it against such criticism, the GM stands to benefit from further study, making use of gaps in the literature such as the relative lack of studies applying it to refugee migration. Given its intuitive consistency with migration theories, the ease of estimation in its simplest form, and its goodness of fit in most applications, I put forward the following hypothesis:

H_1: The GM has strong predictive power in predicting the direction and magnitude of refugee flows.

2.3 Determinants of Refugee Flight Patterns

Fairchild (1927, in Kunz 1973) suggests that measurable exogenous characteristics determine the pathways of forced migration, prompting discussions on the *push* factors of refugee flight patterns. Push factors can be understood as origin country phenomena which catalyse a refugee into flight. Scholars hold that political violence is the main source of refugee outflows, with secondary effects such as economic hardship also contributing to the push effect (Adhikari 2012). Several studies look at violent conflict and forced migration using quantitative methods, finding human rights violations to be a significant predictor (Hakovirta 1993; Apodaca 1998). Critiquing such studies for selection bias in using only countries that generate refugee outflows, Schmeidl's (1997) pooled time series analysis of 109 countries (with both zero and non-zero outflows) between 1971 and 1990 examines forced migration as a macro-structural problem and finds that measures of generalised violence are stronger predictors than measures of human rights violations or economic spillover effects. Moore and Shellman (2004) and Davenport et al. (2003) expand on these findings, the former fitting a multivariate model on a global sample of 40 countries, and the latter fitting a fixed effects least squares model on a pooled cross-sectional time-series data set of 129 countries between 1964 and 1989. In Moore and Shellman's study (2004), government and dissident threat are the primary determinants of refugee outflows, while Davenport, Moore and Poe's study (2003) finds that refugees flee when their integrity - liberty, physical person, or lives - is threatened.

Conversely, *pull* factors are characteristics of destination countries which attract refugees, the most common of which is geographical proximity. According to Neumayer (2004), this is based on an evaluation of the net benefit of all possible asylum destinations, making refugees utility maximisers. Some may push against this notion, arguing that in most situations forced migrants are pushed into flight by violent threat, implying that evaluation is linked to survival instead of utility, which would make geographical proximity the main determinant. While valid, this is a false dichotomy as Neumayer's 2004 framework allows for geographical proximity to be integrated as a source of benefit. Nevertheless, the salience of geographical proximity is not misguided, with many scholars studying its relationship to asylum destination choice. For instance, Iqbal (2007) studies forced migration in Africa from 1992 to 2001 via a gravity-like model controlling for armed conflict, regime type and population, finding evidence of the strong influence of distance on forced migrant flows. Anecdotal evidence also confirms the causal link between distance and refugee flows. For example, Rwandan refugees could not cross Lake Kivu, which influenced flight patterns during the Great Lakes refugee crisis (Rüegger and Bohnet 2018, 67; British Refugee Council 2022). In 2013, UNHCR reported that 80% of its global caseload was concentrated within their region of origin (ibid).

Beyond geographical proximity, pull factors include socioeconomic integration prospects and policy friendliness. For example, poor initial labour market conditions have a negative impact on refugees' earnings (Aksoy et al. 2020), and employment bans can significantly affect their integration into the labour force (Marbach 2018). Citizenship options also play a role in destination choice, as refugees value the opportunity to resolve statelessness (Tucker 2018). However, the vector of pull factors is complex, and no single factor prevails as the most influential (Borselli and van Meijl 2020). Some may criticise the framing of the aforementioned deterministic theories as too macro-structural, representing refugees as a phenomenon (Bleiker et al. 2013) instead of autonomous agents whose decisions are informed by micro-level factors (Borselli and van Meijl 2020). As it is outside the scope of the current paper to meaningfully study micro-level determinants in the gravity model framework, this paper concerns itself more with macro-structural factors. However, as mass movement dynamics are an aggregate of individual decisions, studying *macro*-trends can still reveal *micro*-level trends with the aid of qualitative studies that build on macro-patterns identified.

2.4 Ethnocultural Linkages

Refugee movement is influenced by push and pull factors, but it is also important to consider network dynamics beyond individual countries of origin and destination (Barthel and Neumayer 2014), in order to understand "social drivers or larger processes" which underpin flight patterns (EASO 2016, 16). Migration systems theory (MST) views community-level feedback loops as key drivers of destination choice (Mabogunje 1970). Migration systems are "constantly evolving, self-modifying, and perpetuating mechanisms" that emerge from sustained socio-cultural linkages between states (Zlotnik, 1992, in EASO 2016, 16). Network theory is useful in explaining how refugees facilitate their flight by leveraging migrant systems to lower the costs and risks related to migration (Massey et al. 1993, 448–450). A higher share of past asylum seekers and long-term residents from a particular country of origin lowers the costs of migration for others wishing to settle in this destination country, making it more attractive to a refugee (Neumayer 2004). To this end, understanding the role of networks and linkages in refugees' asylum destination choices stands to significantly improve the predictive performance of the GM discussed in Section 2.2.

Building on the intriguing findings of Rüegger and Bohnet (2018, 65) that refugees flee to countries with "ethnic kin populations and a history of accepting other co-ethnic refugees", this paper seeks to understand whether GM performance is positively impacted by making it sensitive to ethnocultural ties. Building on the work of Cederman et al. (2010), I align myself with Rüegger and Bohnet (2018, 66) in assuming that "ethnic group membership can be based on different markers with varying relevance in different political scenarios, such as a common language, religion or physical features". Such dimensions of ethnicity can be understood as parameters of migration systems, as they form networks which refugees use to decrease transport, knowledge, and assimilation costs (Barthel and Neumayer 2014, 1; Moore and Shellman 2004; 2007; Newland 1993; Schmeidl 1997). Moreover, Böcker and Havinga (1997) find that the presence of an asylum community in the country of asylum, language links and colonial ties have the strongest predicting effect on migration systems in the EU.

This can be observed empirically with the Rohingya Muslim minority fleeing Myanmar due to ethno-religious violence (Ullah 2011). The Rohingya diffused to proximate Muslim-majority countries such as Bangladesh (ibid), as well as more distant countries like Indonesia and Malaysia, where the promise of Muslim solidarity generated hope amidst the inadequate protection framework in Southeast Asia (Missbach and Stange 2021). While some may argue that this is religious, not ethnic, kinship, it is difficult to separate religion from ethnicity, as they are often closely intertwined (Cederman et al. 2010). To this end, I put forward the following hypotheses:

H_2: ethnocultural variables improve the predictive power of the gravity model as it relates to refugee flows.

H_3: an ethnicity-conscious gravity model is a reliable tool for policymakers to predict refugee flows in specific crisis scenarios.

2.5 Contributions

This paper expands on existing literature in three ways. Firstly, GMs are yet to be meaningfully applied to the study of refugee movement using global data, as existing applications (e.g.Iqbal 2007) use regional data. As detailed in section 3.1, this paper uses a global dataset to train GMs and fills this gap. Secondly, scholars tend to view the relationship between danger/threat/violation and forced migration as much more obvious and less complex than that between economic factors and voluntary migration (see Massey et al. 1993). Thus, there remains a gap in understanding the statistical aetiology of refugee flows, with most studies separately treating push and pull factors. This paper considers both push and pull factors in training the gravity models, thus filling this gap. Finally, among the already scarce literature using comparative and quantitative methods to study refugee movement (BenEzer and Zetter 2014), Rüegger and Bohnet (2018, 66) identify a gap in quantitative literature studying the effect of ethnocultural pull factors on destination choice. Moreover, there are no studies yet applying the GM framework to this question - a gap that this paper seeks to fill as well.

Chapter 3

Methodology

3.1 Data

I construct my own dyadic panel dataset to investigate my hypotheses. Such a dataset is appropriate, as a static dataset would not be able to capture temporal variation. To train the GM, the dataset must include a measure of yearly bilateral refugee flows by origin i and destination j, population of i and j, as well as bilateral distance. Moreover, the dataset should include covariates capturing economic attractiveness, policy friendliness, political situation, and diasporic ethnocultural linkages, given the theoretical framework set out in Section 2. Table 1 provides the data sources (Gilardi et al. 2017; Conte et al. 2022; Mebelli et al. 2023; Teorell et al. 2023; de Haas et al. 2015; Rüegger and Bohnet 2018; Vogt et al. 2023), variables, limitations and justifications for inclusion into the dataset:

Table 1: Summary of dataset components

Data Source	Variable (***mnemonic***)	Limitations	Justification
UNHCR (1989 - 2013, 177 countries)	**Refugee Flows** (*tot_coml*)	UNHCR only records non-zero dyadic flows, resulting in about 7% of refugees being uncategorised as the data provide no information about their country of origin. Thus, some unreported values are not true zeroes (Marbach, 2018:1).	I use the imputed data by Gilardi and colleagues (2022) - a reputable peer-reviewed academic collaboration.
CEPII Gravity Database (1948 - 2021, 196 countries) (Conte, Cotterlaz and Mayer, 2022)	**Distance** (*distw_harmonic*) **Contiguity binary** (*contig*) **Population** (*pop_orig, pop_dest*)	Many different metrics available, possible sensitivity to measurement strategy.	Take the harmonic mean of the population-weighted average distance between the most populated cities of each country in a dyad, in km, recommended by the authors for consistency with the GM framework.
	Aggregate GDP (*gdp_orig, gdp_dest*) **Per Capita GDP** (*gdpcap_orig, gdpcap_dest*)	Some may argue that GDP 'flatly fails' in measuring economic well-being, as it is designed to measure productivity and economic growth (Kumar, 2019).	It is a standard measure of the economy, and using GDP is more suited to my interest in having a larger sample size.
	Common Official Language Binary (*comlang_off*) **Common Spoken Language Binary** (*comlang_ethno*) **Common Religion Binary** (*comrelig*) **Common Colonial Past Binary** (*comcol*)	Binary variables may fail to capture varying degrees of similarity which may inform refugee flow magnitude.	Lack of better data available, as well as consistency with my definition of ethnicity as well as this database existing for gravity modelling, it is the best source available.
Quality of Government dataset (compiles 100+ data sources and contains 2100+ variables) (Teorell et al. 2023)	**State Fragility Index** (*cspf_sfi*) (Marshall and Elzinga-Marshall, 2017) **Political Terror Index** (*gd_ptsa*) (Gibney et al., 2020) **Conflict Intensity** (*bti_ci*) (Donner et al., 2022)	Conflict intensity, for example, is only available post-2005, requiring imputation which may create bias.	There are reliable imputation methods available, so I still include these covariates given that ethnic, social and religious conflicts are the main driver of refugee outflows (Rüegger and Bohnet, 2015: 68).

DEMIG POLICY dataset (tracks 6,000 migration policy changes in 45 states between 1945 and 2014) (de Haas, Natter and Vezzoli, 2015)	**Yearly change in migration policy restrictiveness** (*change_restrict*)	The concept of policy friendliness is subjective and hard to measure.	This measure takes a maximalist definition which includes a broad range of state efforts to alter refugee flows. This measure is still the best available data, as it is coded to allow meaningful comparison between states.
Ethnicity of Refugees (ER) dataset (Rüegger and Bohnet, 2015), **Transnational Ethnic Kin** (TEK) Dataset (Vogt et al., 2015)	**Diasporic ethnocultural linkages** (ethnic_link) (constructed manually, see below paragraph for details)	Using two different datasets may be risky as some ethnic groups may be coded differently, yielding false negatives.	Both datasets are produced by the International Conflict Research (ICR) group at ETH Zürich, intended to be used together.

As no such variable exists, I detail the construction of my *ethnic_link* variable, which measures whether refugees from i were following their ethnic kin into j. I use the Ethnicity of Refugees (ER) dataset by Rüegger and Bohnet (2018), which records the three largest ethnic groups of refugee stocks globally between 1975 and 2021 based on UNHCR data. An extra step is required to ascertain whether the refugees have ethnocultural kin in their receiving countries. To do this, I use the Transnational Ethnic Kin (TEK) dataset, which records all politically-relevant ethnic groups living in at least two countries (Vogt et al. 2023). I code a binary variable - 1 if at least one of the ethnic groups of the refugee flow matches an ethnic group in the destination country, and 0 otherwise. There is a missingness problem given that the ER dataset only records ethnic groups with 2000+ refugees, however in the absence of better data I proceed with my analysis.

To maximise sample size, I use a Classification and Regression Trees (CART) imputation method to fill in missing values. CART is a multiple-imputation technique wherein the conditional distribution of a univariate outcome (i.e. a variable of interest) is computed from multiple predictors (Burgette and Reiter 2010). It constructs a decision tree on the available observed data, which is used to impute missing values by following the decision rules in the tree (Breiman 2017; Hastie et al. 2009). This method has advantages over other imputation methods, as it can handle missing data in both continuous and categorical variables, can handle interactions and non-linear relationships between covariates, and is easy to implement. As data imputation may introduce model dependence that skews the final results, I replicate my analysis using simple arithmetic mean imputation, and compare between methods to check for model sensitivity. The final dataset comprises 7,500 unique dyad-years (91 origin countries, 107 destination countries), from 1989 to 2013. 15.8% of the data is imputed using CART imputation.

3.2 Gravity Model Specification

3.2.1 Model Equations

I specify two GMs: one which does not include ethnocultural variables - the baseline GM - and one which expands the baseline GM to include ethnocultural variables - an ethnoculturally-sensitive GM. Both models take the same dependent variable - a dyadic measure of refugee flows, and the same baseline covariates, such as distance, population, metrics of economic performance, conflict and policy-friendliness.

The model specifications are:

Model 1: Baseline GM

$$ln\, m_{i,j,t} = G + \alpha\, ln\, P_{i,t} + \beta\, ln\, P_{j,t} - \gamma\, ln\, D_{i,j} + \mathbf{X}_{i,t}\theta + \mathbf{Y}_{j,t}\phi + \mathbf{Z}_{i,j,t}\eta + \varepsilon_{i,j,t} \tag{3.1}$$

Model 2: Ethnicity-conscious GM

$$ln\, m_{i,j,t} = G + \alpha\, ln\, P_{i,t} + \beta\, ln\, P_{j,t} - \gamma\, ln\, D_{i,j} + \mathbf{X}_{i,t}\theta + \mathbf{Y}_{j,t}\phi + \mathbf{Z}_{i,j,t}\eta + \mathbf{Q}_{i,j,t}\lambda + \varepsilon'_{i,j,t} \tag{3.2}$$

Where $m_{i,j,t}$is the dependent variable: refugee flows from origin country i to destination country j in a given year t. The model equations are derived from equation (2.2), by expanding the error term $\varepsilon_{i,j,t}$, assumed to be normally distributed with homoskedastic variance, which captures cross-sectional and time variation unexplained by distance and population. The expansion breaks the error term up into three covariate vectors ($\mathbf{X_{i,t}}$,$\mathbf{Y}_{j,t}$,$\mathbf{Z}_{i,j,t}$). $\mathbf{X_{i,t}}$ holds covariates capturing variation within the origin country, $\mathbf{Y}_{j,t}$ does the same but for destination countries, and $\mathbf{Z}_{i,j,t}$ captures dyadic covariates not related to ethnicity and culture. Model 2 sees a fourth covariate vector added - $\mathbf{Q}_{i,j,t}$ - which captures the ethnocultural dyadic variables. θ, ϕ, η an λ capture the coefficients associated with each covariate. Table i (see Appendix A) summarises which covariates have been log-transformed.

3.2.2 Variable Selection

To 'clean up' my baseline GM prior to comparing model performance, I employ Bayesian Model Averaging (BMA) - a widely used method for variable selection (Wang 2018, iii). BMA deals with the uncertainty covariate inclusion in a model - a critical issue in forecasting exercises (Steel 2011). Take a linear regression model of the following form:

$$y = \alpha_\gamma + \mathbf{X}_\gamma \beta_\gamma + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \tag{3.3}$$

Where y is the dependent variable, α_γ is the constant, β_γ the coefficients and ε the error term. Uncertainty emerges as to which elements $\mathbf{X}_\gamma \in \{X\}$ to include in the model, often leading to the use of a single linear model which includes all available covariates - an inefficient approach (Feldkircher and Zeugner 2022). As an alternative, BMA estimates models for combinations of $\{X\}$, such that if $\{X\}$ contains k covariates, 2^k models are estimated and grouped in a model space $\{M\}$, each with a unique matrix of covariates (ibid). Drawing on Bayes' Theorem, each model is assigned a weight representing the probability that it is the best in terms of predictive power - the posterior model probability (PMP), based on the available data:

$$P\left(M_\gamma \mid y, X\right) = \frac{P\left(y \mid M_\gamma, X\right) P\left(M_\gamma\right)}{P(y \mid X)} = \frac{P\left(y \mid M_\gamma, X\right) P\left(M_\gamma\right)}{\sum_{s=1}^{2^k} P\left(y \mid M_s, X\right) P\left(M_s\right)} \tag{3.4}$$

Where $M_\gamma \in \{M\}$, $P\left(M_\gamma \mid y, X\right)$ is the PMP, $P\left(y \mid M_\gamma, X\right)$ is the marginal likelihood of model M_γ (i.e. the probability of our data being observed given the model M_γ), and $P(M_\gamma)$ is the prior probability of model M_γ being true before looking at the data, and $P(y \mid X)$ is the integrated likelihood which is a constant term across all models contained in $\{M\}$. By a process of renormalisation, Feldkircher and Zeugner (2022) detail how equation (3.4) can be used to compute posterior inclusion probabilities (PIPs) for any individual covariate θ.

$$P(\theta \mid y, X) = \sum_{\gamma=1}^{2^k} P\left(\theta \mid M_\gamma, y, X\right) P\left(M_\gamma \mid X, y\right) \tag{3.5}$$

I set the PIP exclusion threshold of 0.5 (ibid) and use it to determine if any covariates need to be excluded from the final baseline GM. Some may argue that a frequentist method such as a Least Absolute Shrinkage and Selection Operator (LASSO) approach is computationally simpler and sufficient for selecting a subset of predictors that maximise goodness of fit. However, I opt for BMA because it is more comprehensive while accounting for model uncertainty, and to favour simpler models in a way that LASSO does, I use the Bayesian Information Criterion (BIC) for selecting a prior distribution (ibid). To check for sensitivity to prior distribution selection methods, I replicate my BMA results using the Zellner-Siow prior.

Moreover, BMA may generate misleading results in estimating the PIPs of covariates when there is high multicollinearity (Wang 2018, iii). To check for this, Figure *i* (see Appendix B) shows the full correlation matrix of the data, indicating weak correlations among most covariates. Exceptions include GDP and population of origin countries, political terror and state fragility, and spoken and official languages, which are logically related (e.g., GDP is a component of state fragility). Fortunately, Figure *i* also shows that ethnic and non-ethnic covariates are not correlated, which is multicollinearity that would mainly bias my conclusions.

3.2.3 Limitations

Firstly, while working with the natural log of variables is a common approach in gravity modelling to "reduce the skewness of the data and to mitigate the influence of large values" (Hatton, 2009, in Neumayer 2004, 13), it leads to omitted zeros, which can bias the sample (Turkoglu and Chadefaux 2019). To address this, I add 1 to all variables undergoing a natural log-transformation, and subtract 1 after exponentiating the results. This ensures that there are no omitted zeros. Secondly, GM estimates are vulnerable to standard error clustering when some variables apply to only one country in a dyad-year (Redding and Venables 2004; Rose and Wincoop 2001). To address this, I add country fixed-effects for origin and destination countries, as recommended by Feenstra (2015). Finally, GMs fail to capture temporal dynamics, as argued by Beyer et al. (2022) Beyer et al. (2022). This is a valid critique which prompts mixing GMs with other techniques better adapted for capturing temporal variation, such as autoregressive models (Bijak et al. 2019). As model-mixing is outside of the scope of this paper, I instead add lagged versions of the dependent variable (refugee flows) as well as some covariates (see Table i in Appendix A) to capture time effects, making my models dynamic.

3.3 Evaluating Model Performance

3.3.1 Out-of-Sample Performance

To assess the power of a model, its generalisability is best tested through out-of-sample predictions using a test set (Hawkins et al. 2003). The validation set approach, where data is split into training and test sets, has been criticised for yielding unreliable mean squared error (MSE) estimates across different permutations (James et al. 2021). *Leave-one-out cross-validation* (LOOCV) has been suggested as a remedy, whereby a single observation is used as the test set and the rest as the training set, iteratively computing the MSE and obtaining the CV estimate of the test error as the average MSE across all iterations (ibid.). While LOOCV eliminates bias, it is computationally heavy in large-N datasets. Moreover, due to a bias-variance trade-off, LOOCV approximately eliminates bias in the test error estimates at the expense of variance, because each iteration uses almost identical training data which makes the final estimates highly correlated.

As an improvement to LOOCV, I use *k-fold cross validation* (kFCV), wherein the data is divided into k folds - or strata - and each one is varied as the test set with the rest being used to train the model (ibid). While the usual values for k are either 5 or 10, I fold through time in the interest of comparing model performance on future data. Thus, each fold takes one year as the test set and all previous years as the training set. For example, the first fold will use all observations for 1989 to train both gravity models, and then compute a CV estimate using all observations for 1990. This process is iterated 24 times (i.e. $k = 24$) for the 25 years of data. Some may argue that I should reduce my k value for computational simplicity, and to avoid having a low-performing first fold yielding an inaccurately high error estimate. However, I am interested in seeing how well my models predict refugee flows using available data, even if it is only available for one year prior to the needed prediction, as is the case with the Ukraine refugee crisis.

Thus, I proceed with kFCV to calculate my model performance statistics, for which I take the RMSE (representing the square of the variance of the models' residuals) and the adjusted R^2 (how much of the variance in y is explained by the model - a measure of model fit) (Alexander et al. 2015). As I am comparing two models with different numbers of covariates, I take adjusted R^2 over regular R^2 to account for the change in model correctness resulting from the higher number of covariates.

3.3.2 Testing Policy Applicability

After using kFCV to choose the better gravity model, I am interested in simulating how the chosen model would perform in a policy case that warrants estimating the magnitude of refugee flows between specific countries - an essential resource for policymaking (Abel and Sander 2014). I first compare predicted versus actual flows (both before and after reversing log-transforms) in three historical mass refugee outflows: the Rohingya refugee crisis (Guglielmi et al. 2020), the 1991-1999 Yugoslav Wars (Radović 2005), and the 1994 Rwandan Genocide (Fair and Parks 2001). I run each simulation by identifying the relevant dyad-years for each historical case (see Table *ii* in Appendix C) and partitioning the data such that the model is tested on those dyad-years and trained on the remainder. In this spirit, I also test my model using quasi-real-time data relating to refugee outflows from Ukraine as a result of the Russian invasion. As not all parameters have available data for 2022 and 2023, I estimate values such as conflict intensity using Ukraine in 2014-2015 as a reference point, given the conflict at the time. This data is then joined to the input data detailed above.

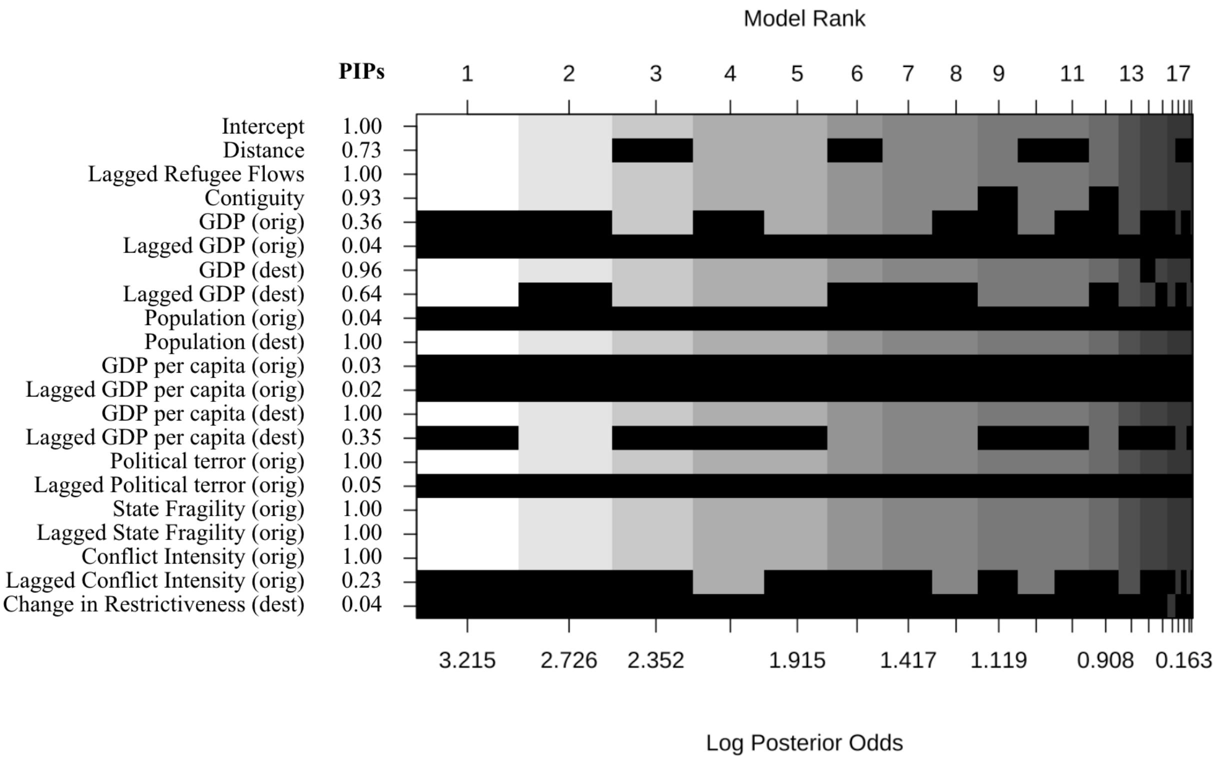
Chapter 4

Findings

4.1 Chosen Variables with BMA

Figure 4.1 shows the BMA results:

Figure 4.1: Covariate inclusion based on BMA



Having set a PIP threshold of 0.5, the results clearly indicate which variables should be used to train the baseline gravity model, and which ones can be excluded at no detriment to the model's performance. Based on the results, I exclude the following non-gravity variables (and their lagged versions) from the baseline gravity equation: GDP (origin), GDP per capita (origin) and Policy Restrictiveness (destination). For GDP per capita (destination), Political Terror (origin) and Conflict intensity (origin), the contemporaneous values have a high enough PIP but the lagged versions do not, so I only remove their lags. Thus, the final baseline model is as follows:

$$\begin{aligned}
&\ln\left(m_{i,j,t}\right) = G + \alpha \ln\left(P_{i,t}\right) + \beta \ln\left(P_{j,t}\right) - \gamma \ln\left(D_{i,j}\right) + \\
&\omega \ln\left(m_{i,j,t(ag)}\right) + \lambda\left(X_i\right) + \mu\left(X_t\right) + \eta_1\left(\text{ contig }_{i,j}\right) + \\
&\varphi_1 \ln\left(GDP_{j,t}\right) + \varphi_2 \ln\left(GDP_{j,t(lag)}\right) + \\
&\varphi_3 \ln\left(GDP \text{ per capita }_{j,t}\right) + \theta_1 \ln\left(\text{political terror }_{i,t}\right) + \\
&\theta_2 \ln\left(\text{state fragility }_{i,t}\right) + \theta_3 \ln\left(\text{state fragility }_{i,t}(\text{lag})\right) + \\
&\theta_4 \ln\left(\text{ conflict intensity }_{i,t}\right) + \varepsilon_{i,j,t}
\end{aligned} \tag{4.1}$$

4.2 k-Fold Cross-Validation

This section deals with the results of the kFCV conducted using the baseline GM in equation (4.1) and the ethnicity-conscious GM constructed by adding the ethnocultural variables. Table 2 shows the time-series kFCV results for both gravity models:

Table 2: kFCV results

k = 24	Average RMSE - $CV_{(k), RMSE}$	Average R_2 - $CV_{(k), R^2}$
Baseline gravity model	1.664438	0.746152
Ethno-cultural conscious gravity model	1.663257	0.751582

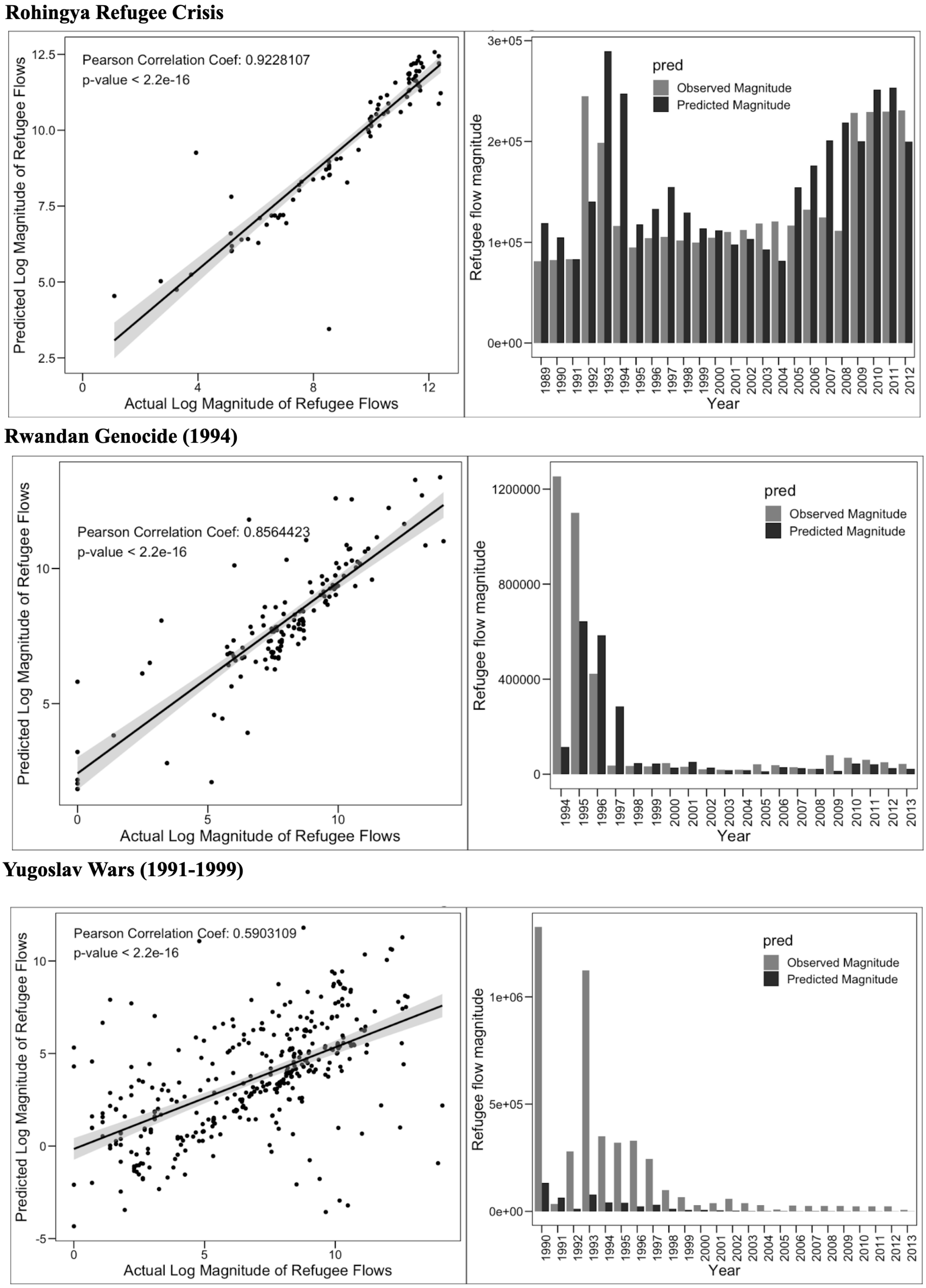
According to guidance by Tropsha, Gramatica, and Gombar (2013, in Alexander et al. 2015), a strong R^2 is equal to or larger than 0.60. Moreover, it is difficult to determine what a good RMSE value is, as it is a measure of how wrong on average a prediction is, binding the range of the RMSE estimates to the range of the dependent variable (ibid). As such, I take a good RMSE as one which is below 10% of the maximum dependent variable value.

With that said, both models exhibit high average adjusted R^2 values. On average, both models explain roughly 75% of the variance in the log-magnitude of flows out-of-sample, with the baseline model having a slightly lower value (74.6%) than the model containing the ethnocultural variables (75.2%). This suggests that the model tends to generalise correctly to new data, evidencing strong out-of-sample performance, especially in congruence with the work of Suleimenova et al. (2017) who find similar model fit (75%) for a generalised simulation development approach they propose to estimate refugee destinations. Secondly, looking at the average RMSE across folds, I also find very similar results for both models, with the model containing the ethnocultural variables slightly outperforming the baseline model. Thus, on average, the predicted log-magnitude is over-predicted by 1.66 units. Given that the log-magnitude of observed flows ranges from 0 to 15, the average RMSE represents an error of roughly 11%, which could make the model substantially less reliable than what is suggested by the R^2. Based on these results I choose the ethnoculturally-sensitive GM to run case-specific simulations and further assess forecasting performance.

4.3 Case Simulations

Figure 4.2 shows the results of the case simulations:

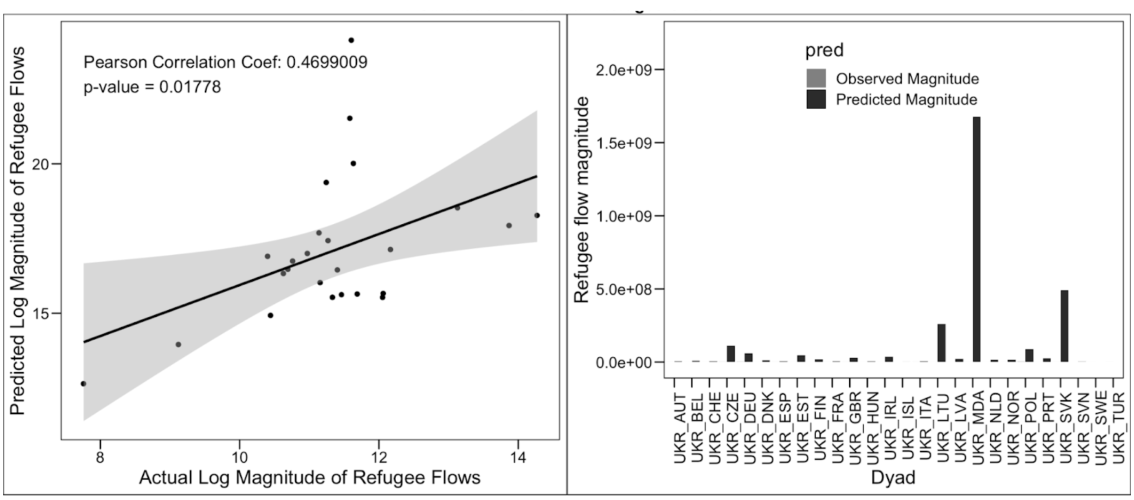
Figure 4.2: Case simulations



I find strong (> 0.5) Pearson correlations, statistically significant at all conventional levels given the p-values ($< 2.2 \times 10^{-16}$), in all three case simulations. Specifically, the model performs surprisingly well in the case of the Rohingya refugee crisis: with a correlation coefficient of 0.92 between the predicted and actual magnitude of refugee flows. Visually represented in the right-hand Rohingya graph, the model was able to capture time variation in total flows per year, as evidenced by the parallel spikes and drops between 1992 and 1994 and around 2008 for example. In the case of the Rwandan genocide, the model also captured the spike in refugees in the immediate aftermath of the crisis and showed a strong (0.86) correlation as well. A moderately strong correlation (0.59) was present in the Yugoslav Wars simulation, with a weak capturing of the time variation and a large difference in absolute flow magnitudes.

Finally, after constructing the input dataset, I run the same simulation on Ukraine refugee outflows into a variety of destinations (see Table *ii* in Appendix C), the results of which are shown in Figure 4.3.

Figure 4.3: Results of real-time Ukraine refugee flow forecasts (as of April 2023)



The results show that the model does not perform as well in this simulation. The correlation is not strong (< 0.5), while statistically significant at the 95% confidence level given the p-value of 0.02. The plot has been adjusted so that the right-hand graph has the dyads on the x-axis instead of the years, as the estimates are only for 2023. They show that the model vastly overpredicts the actual number of Ukrainian refugees, especially in the case of Moldova.

4.4 Robustness

To ensure the reliability of my results, I conduct two robustness checks. Firstly, as a robustness check against sensitivity to prior selection methods in the BMA, I replicate the BMA results using the Zellner-Siow prior instead of the Bayesian Information Criterion to determine the prior model distribution. As seen in Figure *ii* (see Appendix D), the results are very similar, yielding identical conclusions as the BIC method regarding which covariates to include in the baseline model. Thus, the results appear not method-dependent.

Secondly, as a robustness check on the performance assessment, I replicate the results of the kFCV using data imputed using a simple arithmetic mean approach instead of the CART method, the results of which can be seen in Table *iii* (see Appendix E). While the RMSE is generally the same, there is an almost 10 percentage point difference in the R^2 estimates. This difference could come from a variety of reasons - for instance, the mean-imputation resulting in some NAs remaining post-imputation if there are no single values for a specific dyad. As a result, these dyads must be removed to allow the gravity model to run, as at its foundation it is a generalised linear model, yielding a smaller dataset (n = 6,625). Moreover, unlike the multiple imputation method used, mean-imputation cannot meaningfully differentiate between continuous and categorical variables, meaning that categorical variables may be imputed with meaningless estimates. This evidence shows that there could be some degree of imputation method dependence that warrants further research into the sources of bias.

4.5 Discussion

The results support all three of my proposed hypotheses, with some nuance. H_1 posited that the GM can predict forced migrant flows with strong predictive power, later defined to mean an $R^2 > 0.60$ and RMSE below 10% of the maximum dependent variable value. The approximate average R^2 of 0.77 for both the baseline and the ethnoculturally-sensitive GMs partially supports this hypothesis. However, the average RMSE comprises about 11% of the maximum dependent variable value, indicating a substantial presence of prediction error that challenges the model's generalisability. This invokes the question of whether the RMSE or the R^2 reveals more about the model's performance. On one hand, scholars prefer RMSE over R^2 to assess predictive power, because "the value of a model is generally in its overall accuracy and precision and not how successfully it explains the variation in a particular data set" (Alexander et al. 2015, 1318). On the other hand, the RMSE is also criticised as its values can range between zero and +infinity, with a single value "not saying much about the performance of the regression with respect to the distribution of the ground truth elements." (Chicco et al. 2021, 1). Due to the interpretability of the R^2 and the ability of RMSE to capture prediction accuracy and precision, I use both as recommended by Alexander et al. (2015). This means that evidence to support H_1 is ambiguous, warranting further research into estimates' sensitivity to performance statistics.

H_2 theorised that "ethnocultural variables improve the predictive power of the gravity model on refugee flows". My results also support this, as the average RMSE of the ethnoculturally-sensitive GM is slightly higher than for the baseline model. However, the difference is minuscule, prompting further study into whether the difference in average RMSE has occurred due to specific features of the data itself, or out of a true effect of ethno-cultural variables. As more data is hard to compile, it is outside of the scope of this paper to investigate this.

H_3 proposed that "an ethnicity-conscious gravity model is a reliable tool for policymakers in predicting refugee flows in specific crisis scenarios". The results here are mixed. On one end, the model performed well in the three historical cases I simulated, doing especially well in the Rohingya refugee crisis simulation, providing some evidence to support H_3. This could be explained by the direct ethnic targeting of the Rohingya making ethnocultural linkages more salient than with victims of the Yugoslav wars for instance, who were fleeing general violence.

However, it fell short in two ways. First, while the model was able to somewhat simulate the patterns of refugee flight in the specified scenarios, there was a sizeable error in the absolute magnitude of flows, especially in the Yugoslav Wars. Second, its performance worsened in the Ukraine simulation, providing evidence to reject H_3. This may have occurred for a variety of reasons, including the lack of reliable, peer-reviewed data for more recent refugee movements (Rüegger and Bohnet 2018) which creates a gap between 2013 and 2022. Moreover, one could question the use of a global dataset as opposed to targeting specific regions, given that data is more readily available on the regional level, especially in Europe, and could thus produce more reliable, albeit geographically bound, estimates. While valid, Section 2 discussed the lack of comparative and quantitative research on refugee movement using global datasets, while highlighting that most studies of this nature focus on European data. As such, it is important to use all data available to expand knowledge on patterns that span beyond regions.

Overall, the results give mixed evidence of H_1 and H_3 especially, meaning that further research is still required to uncover the power of GM in the context of forced migration policymaking. A possible route to investigate is model mixing. For example, gravity models could be combined with exponential smoothing and autoregressive integrated moving average (ARIMA) models. Due to their ability to capture time variation, these approaches have been used to forecast refugee movements by Mebelli et al. (2023). This could help improve policymakers' and humanitarian actors' real-time response to ongoing refugee crises, like the one in Ukraine.

Chapter 5

Conclusion

This paper examined the reliability of the gravity model (GM) in predicting the magnitude and direction of bilateral refugee flows, as well as the potential improvement of its predictive power with ethnocultural sensitivity. I constructed a global dataset from various peer-reviewed databases, informed by a thorough review of existing qualitative and quantitative literature on the push and pull factors of forced migration.

I hypothesed that GMs would exhibit strong predictive power, and that adding ethnocultural linkage covariates would strengthen prediction accuracy. I also hypothesized that an ethnoculturally-sensitive model would reliably predict refugee movement in the cases of the Rohingya refugee crisis, the Yugoslav Wars and the Rwandan Genocide. To simulate a policy use-case, I tested the model against quasi-real-time Ukraine refugee statistics published monthly by the UNHCR.

The results revealed mixed evidence. On one hand, the model exhibited surprisingly strong performance in the Rohingya simulation via its ability to capture time variation in total yearly flow magnitudes, as well as a remarkably high average R^2, demonstrating promising potential. On the other hand, poor performance in the Ukraine simulation, as well as large prediction error in the non-log transformed estimates, suggested that the model is not quite ready for policy applications.

Ultimately, the gravity model is a useful framework, particularly powerful in explaining historical flows. With further research into model-mixing approaches, testing other statistical learning methods, and exploring different covariates, the GM stands to be a powerful tool for policymakers and humanitarian agents working to improve the world's capacity to deal with refugee crises.

Appendix A

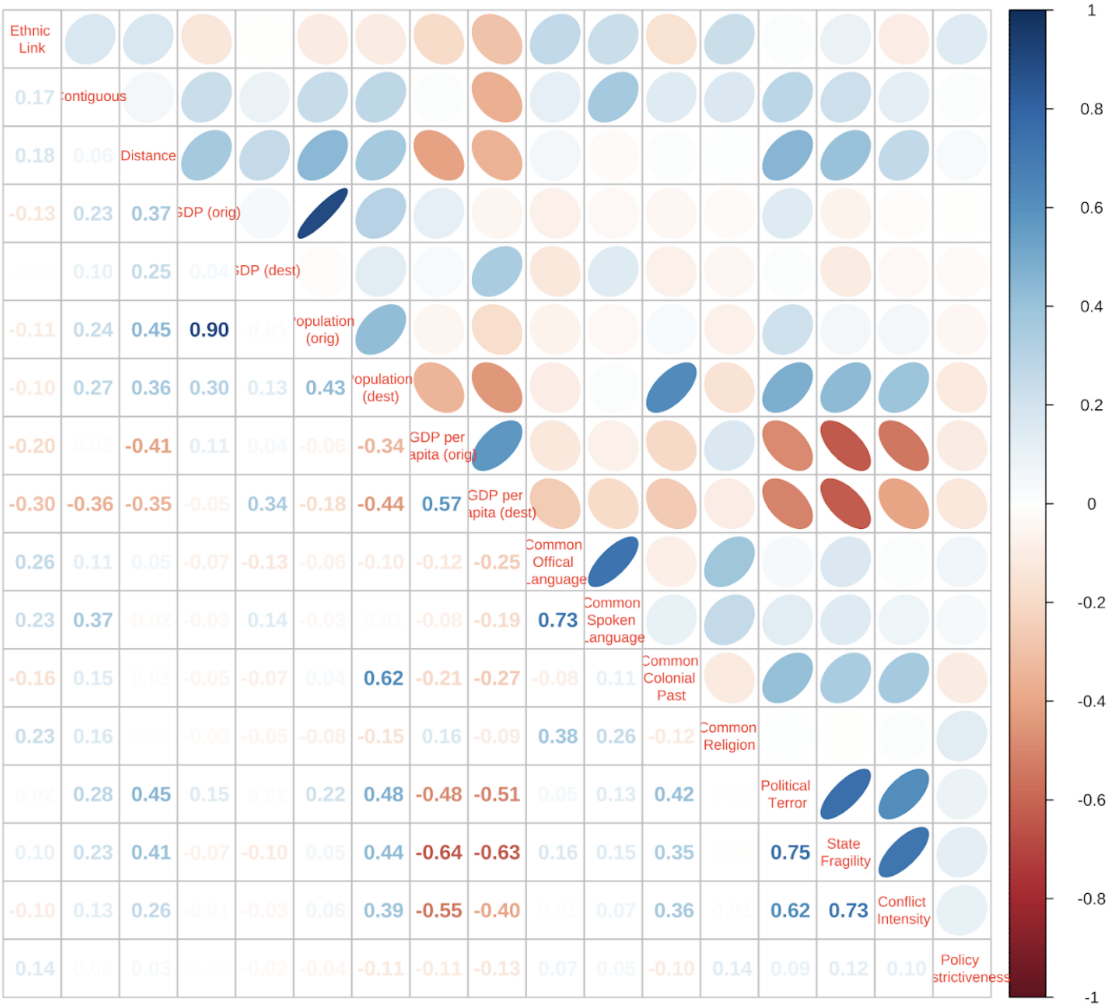
Summary of covariate logs and lags

Table i: Covariates available for each dyad-year.

Vector	Variation captured	Variables (units)	Logged?	Lagged?
$\mathbf{X}_{i,t}$	Within country of origin	Population (number of people)	✓	
		Gross Domestic Product (GDP) (USD, 2017 PPP)	✓	✓
		GDP per capita (USD, 2017 PPP)	✓	✓
		State fragility (index score)	✓	✓
		Political terror (index score)	✓	✓
		Conflict intensity (index score)	✓	✓
		Dummy variable for country of origin.		
$\mathbf{X}_{j,t}$	Within country of asylum	Population (number of people)	✓	
		Gross Domestic Product (GDP) (USD, 2017 PPP)	✓	✓
		GDP per capita (USD, 2017 PPP)	✓	✓
		Change in migration policy friendliness (index)		
		Dummy variable for country of destination.		
$\mathbf{Y}_{i,j,}$	Non-ethnic dyadic factors	Distance (km)	✓	
		Contiguity (binary)		
$\mathbf{Z}_{i,j,t}$	Ethno-cultural dyadic variables	Refugee-specific ethnic linkage		
		Common official language		
		Common spoken language		
		Common religion		
		Common colonial past.		

Appendix B

Pairwise correlations of covariates



Appendix C

Which dyad-years are included in each case simulation?

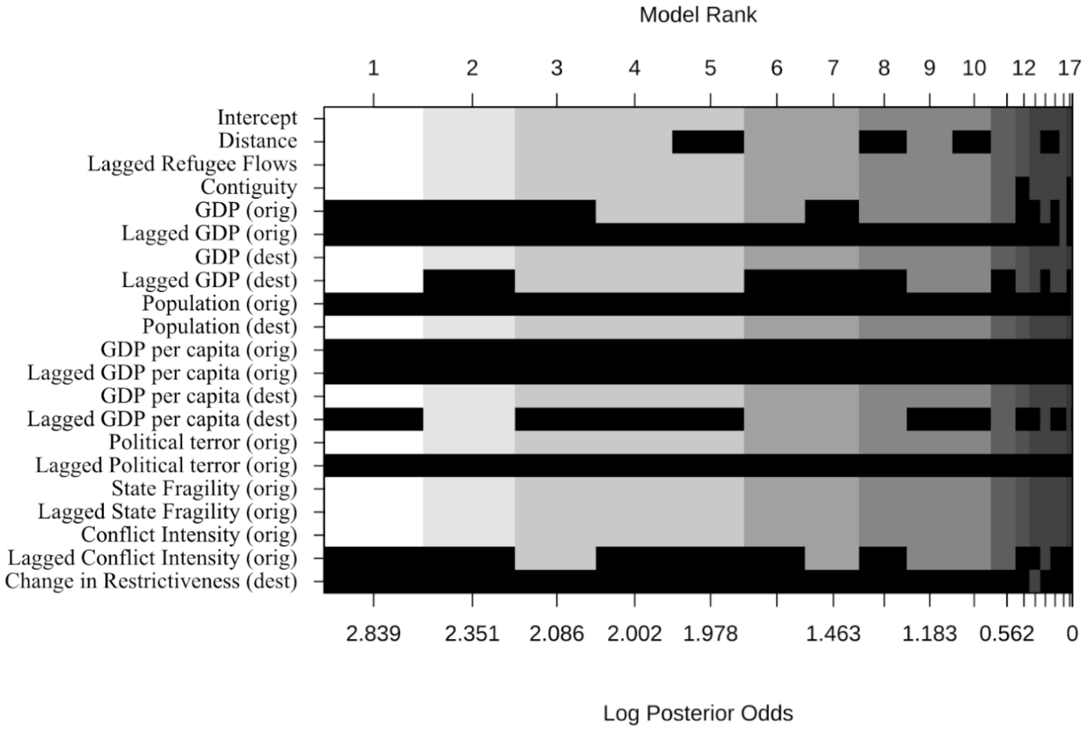
Table ii: Countries and years composing the test set of each simulation

Case	Details
Rohingya Refugee Crisis	• Origin countries: Myanmar • Destination countries: all • Years: 1989-2012
Rwandan Genocide (1994)	• Origin country: Rwanda • Destination countries: all • Years: 1994 - 2013
Yugoslav Wars (1991-1999)	• Origin countries: Bosnia and Herzegovina, Croatia • Destination countries: all • Years: 1990 - 2013
Ukraine Refugee Crisis (2022-)	• Origin countries: Ukraine • Destination countries: Austria, Belgium, Switzerland, Czech Republic, Denmark, Spain, Estonia, Finland, France, UK, Hungary, Ireland, Lithuania, Latvia, Moldova, Netherlands, Norway, Poland, Portugal, Russia, Slovakia, Slovenia, Sweden, Turkey • Years: 2023

Appendix D

Comparing BMA results across different methods for prior selection

Figure ii: BMA results using Zellner-Siow prior



Appendix E

Comparing k-Fold Cross Validation results between imputation methods

Table iii: kFCV results (using mean imputation instead of CART imputation)

k = 24	**Average RMSE - $CV_{(k), RMSE}$**	**Average R_2 - $CV_{(k), R^2}$**
Baseline gravity model	1.551051	0.8817866
Ethno-cultural conscious gravity model	1.578868	0.8824293

Appendix F

Colophon

The raw data files, my imputed dataset, and all reproducible code for my analysis can be found in the following GitHub repository

All statistical analysis was conducted in R, using R-Studio. This document was written in the Times Roman typeface using LaTeX and BibTeX, composed with Overleaf.

Bibliography

Abel, G. J. and Sander, N. (2014). Quantifying global international migration flows. *Science*, 343:1520–1522.

Adhikari, P. (2012). The plight of the forgotten ones: Civil war and forced migration. *International Studies Quarterly*, 56:590–606.

Afifi, T. and Warner, K. (2008). The impact of environmental degradation on migration flows across countries.

Aksoy, C. G., Poutvaara, P., and Schikora, F. (2020). First time around: Local conditions and multi-dimensional integration of refugees.

Alexander, D. L. J., Tropsha, A., and Winkler, D. A. (2015). Beware of r2: Simple, unambiguous assessment of the prediction accuracy of qsar and qspr models. *Journal of Chemical Information and Modeling*, 55:1316–1322.

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, 3(1):133–160.

Apodaca, C. (1998). Human rights abuses: Precursor to refugee flight? research note. *Journal of Refugee Studies*, 11(1):80–93.

Backhaus, A., Martinez-Zarzoso, I., and Muris, C. (2015). Do climate variations explain bilateral migration? a gravity model analysis. *IZA Journal of Migration*, 4(1):3.

Barthel, F. and Neumayer, E. (2014). Spatial dependence in asylum migration. *Journal of Ethnic and Migration Studies*, 41:1131–1151.

Beine, M., Docquier, F., and Özden, (2011). Diasporas. *Journal of Development Economics*, 95:30–41.

Belsey-Priebe, M., Lyons, D., and Buonocore, J. J. (2021). Covid-19s impact on american women's food insecurity foreshadows vulnerabilities to climate change. *International Journal of Environmental Research and Public Health*, 18:6867.

BenEzer, G. and Zetter, R. (2014). Searching for directions: Conceptual and methodological challenges in researching refugee journeys. *Journal of Refugee Studies*, 28:297–318.

Beyer, R. M., Schewe, J., and Lotze-Campen, H. (2022). Gravity models do not explain, and cannot predict, international migration dynamics. *Humanities and Social Sciences Communications*, 9(1):56.

Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W., and Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the united kingdom. *Journal of Forecasting*, 38.

Bleiker, R., Campbell, D., Hutchison, E., and Nicholson, X. (2013). The visual dehumanisation of refugees. *Australian Journal of Political Science*, 48:398–416.

Borghesi, S. and Vercelli, A. (2003). Sustainable globalisation. *Ecological Economics*, 44:77–89.

Borselli, M. and van Meijl, T. (2020). Linking migration aspirations to integration prospects: the experience of syrian refugees in sweden. *Journal of Refugee Studies*, 34.

Breiman, L. (2017). *Classification and regression trees*. Routledge.

British Refugee Council (2022). The uk's role in the international refugee protection system - july 2022.

Burgette, L. F. and Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology*, 172.

Böcker, A. and Havinga, T. (1997). Asylum migration to the european union: Patterns of origin and destination.

Cederman, L.-E., Wimmer, A., and Min, B. (2010). Why do ethnic groups rebel? new data and analysis. *World Politics*, 62:87–119.

Chicco, D., Warrens, M. J., and Jurman, G. (2021). The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7:e623.

Cohen, J. E., Roig, M., Reuman, D. C., and GoGwilt, C. (2008). International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences*, 105(40):15269–15274.

Concern Worldwide (2023). The 10 largest refugee crises to know in 2023.

Conte, M., Cotterlaz, P., and Mayer, T. (2022). Cepii working paper.

Crozet, M. (2004). Do migrants follow market potentials? an estimation of a new economic geography model. *Journal of Economic Geography*, 4:439–458.

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

de Haas, H., Natter, K., and Vezzoli, S. (2015). Conceptualizing and measuring migration policy change. *Comparative Migration Studies*, 3.

EASO (2016). Significant pull/push factors for determining of asylum-related migration: a literature review.

Fair, J. E. and Parks, L. (2001). Africa on camera: Television news coverage and aerial imaging of rwandan refugees. *Africa Today*, 48:35–57.

Feenstra, R. C. (2015). *Advanced International Trade*.

Feldkircher, M. and Zeugner, S. (2022). Bayesian model averaging with bms for bms version 0.3.5.

Fitzpatrick, J. (1996). Revitalizing the 1951 refugee convention. *Harvard Human Rights Journal*, 9:229–254.

Gaffar, A. (2018). Approaching the rohingya crisis: Calling for a security governance framework. *World Affairs: The Journal of International Issues*, 22(1):98–121.

Garcia, A. J., Pindolia, D. K., Lopiano, K. K., and Tatem, A. J. (2014). Modeling internal migration flows in sub-saharan africa using census microdata. *Migration Studies*, 3:89–110.

Gilardi, F., Beltran, R., Cornehls, S., Gut, R., Scheiwiller, P., and Stepanski, L. (2017). Mapping global refugee movements. *University of Zurich the Swiss Agency for Development and Cooperation*.

Guglielmi, S., Seager, J., Mitu, K., Baird, S., and Jones, N. (2020). 'people won't die due to the disease; they will die due to hunger': exploring the impacts of covid-19 on rohingya and bangladeshi adolescents in cox's bazar.

Guo, C. G., Al Ariss, A., and Brewster, C. (2020). Understanding the global refugee crisis: Managerial consequences and policy implications. *Academy of Management Perspectives*, 34.

Hakovirta, H. (1993). The global refugee problem: A model and its application. *International Political Science Review / Revue internationale de science politique*, 14(1):35–57.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hawkins, D. M., Basak, S. C., and Mills, D. (2003). Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences*, 43:579–586.

House, C. (2005). The refugees convention: why not scrap it? a summary of discussion at the international law programme discussion group at chatham house on 20th october 2005; participants included lawyers, academics and representatives from governments and ngos.

Iqbal, Z. (2007). The geo-politics of forced migration in africa, 1992-2001. *Conflict Management and Peace Science*, 24(2):105–119.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *INTRODUCTION TO STATISTICAL LEARNING : with applications in r.* Springer-Verlag New York.

Karemera, D., Oguledo, V. I., and Davis, B. (2000). A gravity model analysis of international migration to north america. *Applied Economics*, 32(13):1745–1755.

Kim, K. and Cohen, J. E. (2010). Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity1. *International Migration Review*, 44(4):899–932.

Kirişci, K. (2021). The 1951 refugee convention is falling short of its mission. could the global compact on refugees help?

Kunz, F. E. (1973). The refugee in flight: Kinetic models and forms of displacement. *The International Migration Review*, 7(2):125–146.

Mabogunje, A. L. (1970). Systems approach to a theory of rural-urban migration. *Geographical Analysis*, 2:1–18.

Marbach, M. (2018). On imputing unhcr data. *Research Politics*, 5(4):2053168018803239.

Markham, L. (2022). 'a disaster waiting to happen': who was really responsible for the fire at moria refugee camp? *The Guardian*.

Marshall, L. W. (2010). Toward a new definition of 'refugee': is the 1951 convention out of date? *European Journal of Trauma and Emergency Surgery*, 37:61–66.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Edward, T. J. (1993). Theories of international migration: A review and appraisal. *Population and Development Review*, 19(3):431–466.

Mayda, A. M. (2010). International migration: a panel data analysis of the determinants of bilateral flows. *Journal of Population Economics*, 23(4):1249–1274.

Mebelli, V., Drakaki, M., and Tzionas, P. (2023). An investigation of time series models for forecasting mixed migration flows: Focusing in germany. *Operations Research Forum*, 4.

Miller, D. S. and Chtouris, S. (2017). Borderland security and migration: Balancing humanitarian response with crisis preparedness emergency management—the social and cultural challenges to homeland security. *Journal of Applied Security Research*, 12:1–6.

Milner, H. V. (2021). Is global capitalism compatible with democracy? inequality, insecurity, and interdependence. *International Studies Quarterly*, 65.

Missbach, A. and Stange, G. (2021). Muslim solidarity and the lack of effective protection for rohingya refugees in southeast asia. *Social Sciences*, 10:166.

Moore, W. H. and Shellman, S. M. (2004). Fear of persecution: Forced migration, 1952-1995. *Journal of Conflict Resolution*, 48(5):723–745.

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

Neumayer, E. (2004). Asylum destination choice: What makes some west european countries more attractive than others? *European Union Politics*, 5(2):155–180.

Newland, K. (1993). Ethnic conflict and refugees. *Survival*, 35(1):81–101.

Offe, C. (2011). From migration in geographic space to migration in biographic time: Views from europe*. *Journal of Political Philosophy*, 19(3):333–373.

OHCHR (2021). Ohchr and protecting human rights in humanitarian crises.

Pellandra, A. and Henningsen, G. (2022). Predicting refugee flows with big data: a new opportunity or a pipe dream?

Perraton, J. (2019). *The Scope and Implications of Globalisation*, pages 50–76. The Handbook of Globalisation, Third Edition. Edward Elgar Publishing.

Piesse, M. (2020). Global water supply and demand trends point towards rising water insecurity.

Poot, J., Alimi, O., Cameron, M. P., and Maré, D. C. (2016). The gravity model of migration: The successful comeback of an ageing superstar in regional science.

Poprawe, M. (2015). On the relationship between corruption and migration: empirical evidence from a gravity model of migration. *Public Choice*, 163(3):337–354.

Radović, B. (2005). A brief retrospective on the problem of refugees in the yugoslav wars 1991-99.

Ramos, R. (2016). Gravity models: A tool for migration analysis. *IZA World of Labor*.

Redding, S. and Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, 62(1):53–82.

Rose, A. K. and Wincoop, v. (2001). National money as a barrier to international trade: The real case for currency union. *American Economic Review*, 91(2):386–390.

Roser, M., Hasell, J., Herre, B., and Macdonald, B. (2021). War and peace. *Our World In Data*.

Rüegger, S. and Bohnet, H. (2018). The ethnicity of refugees (er): A new dataset for understanding flight patterns. *Conflict Management and Peace Science*, 35(1):65–88.

Schmeidl, S. (1997). Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990. *Social Science Quarterly*, 78(2):284–308.

Shultz, C., Barrios, A., Krasnikov, A. V., Becker, I., Bennett, A. M., Emile, R., Hokkinen, M., Pennington, J. R., Santos, M., and Sierra, J. (2020). The global refugee crisis: Pathway for a more humanitarian solution. *Journal of Macromarketing*, 40:128–143.

Steel, M. F. (2011). Bayesian model averaging and forecasting. *Bulletin of EU and US Inflation and Macroeconomic Analysis*, 200:30–41.

Stein, B. N. (1981). The refugee experience: Defining the parameters of a field of study. *International Migration Review*, 15:320–30.

Stewart, J. Q. (1950). The development of social physics. *American Journal of Physics*, 18(5):239–253.

Suleimenova, D., Bell, D., and Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific Reports*, 7.

Teorell, J., Sundström, A., Holmberg, S., Rothstein, B., Alvarado Pachon, N., Dalli, C. M., and Meijers, Y. (2023). The quality of government standard dataset, version jan23.

Tucker, J. (2018). Why here? factors influencing palestinian refugees from syria in choosing germany or sweden as asylum destinations. *Comparative Migration Studies*, 6(1):29.

Turkoglu, O. and Chadefaux, T. (2019). Nowhere to go? why do some civil wars generate more refugees than others? *International Interactions*, 45(2):401–420.

Ullah, A. A. (2011). Rohingya refugees to bangladesh: Historical exclusions and contemporary marginalization. *Journal of Immigrant Refugee Studies*, 9:139–161.

UNHCR (2013). What is a refugee camp? definition and statistics — usa for unhcr.

United Nations (1951). The 1951 convention relating to the status of refugees and its 1967 protocol.

Verdirame, G. and Harrell-Bond, B. (2005). *Rights in Exile: Janus-Faced Humanitarianism*. Berghahn Books.

Vogt, M., Bormann, N.-C., Rüegger, S., Cederman, L.-E., Hunziker, P., and Girardin, L. (2023). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. 59.

Wang, R. (2018). *Understanding Multicollinearity in Bayesian Model Averaging with BIC Approximation*. PhD thesis.

Zipf, G. K. (1946). The p1 p2/d hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686.