Gender differences in migration flow predictions: Potential origins and insights from theory

Athina Anastasiadou[1,2] and Emilio Zagheni[1]

[1]Max Planck Institute for Demographic Research
[2]Netherlands Interdisciplinary Demographic Institute (NIDI) - KNAW/University of Groningen

Abstract:

Migration is a complex process and is difficult to predict. At the same time, migration is known to be a highly gendered process reflecting different motivations, propensities, and outcomes of migration by gender. Such differences are mirrored in the fact that on average 45% of the bilateral migration corridors worldwide are male-predominant while 32% are female-predominant. As theories and models have been shaped over decades around the narrative of the male migrant, this can potentially translate into discrepancies in migration predictions by gender. Using one of the most comprehensive macro-level data sets on bilateral migration flows disaggregated by sex, we aim to explore and understand such differences in migration predictions. We compare the predictive performance of a basic gravity model to simpler deterministic methods. Our findings show that overall worse performance measures are achieved for female migration flows compared to male when applying the baseline model. In male-predominant corridors the models tend to predict the male flows better than the female ones and vice versa. For gender balanced flows they seem to perform equally poorly. Further enhancing our understanding of the underlying mechanisms resulting in different prediction accuracy has the potential to inform the work of international organizations, researchers, and policy-makers.

Keywords:

gender, sex, migration, methods, theory, prediction

1 Introduction

Migration is a complex process sensitive to the environment surrounding it, making it the most volatile demographic process and difficult to predict. In order to better understand and conceptualize this phenomenon, theories of international migration have emerged trying to explain why some people migrate and others do not. Such theories have translated into statistical methods for migration prediction of which many are applied by researchers and international organizations in their daily work. Demographers and policy-makers need reliable predictions of migration in order to anticipate migration flows and inform policies and population projections. However, migration theories do not account for gender-specific aspects of the migration process and therefore struggle to explain the migration patterns of women (Boyd & Grieco, 2003).

Exploring migration flows more closely reveals that female and male migrations behave differently and sometimes even exhibit diverging time trends. Overall, 45% of the non-zero bilateral migration corridors[1] in the data set are male-predominant on average over time, while 32% are female-predominant, with the remaining ones being gender balanced (based on the Pseudo Bayes estimates produced by Abel and Cohen (2022)). Such pronounced differences raise important questions given that researchers are trying to predict these patterns with the same types of models.

International migration follows highly gendered patterns. Findings in the migration literature point at differences in migration propensities by gender conditional on migrants' aspirations, at differences in the usage of migrant networks, and at different migration motivations and migration outcomes by gender (Anastasiadou, Kim, Sanlitürk, de Valk, & Zagheni, 2023). However, classical migration theories fail to incorporate social dividers like gender (Riosmena, 2024). Therefore, these theories struggle to explain the predominance of women in certain migration corridors and the specific conditions under which they migrate. At the same time, developing a gendered theory of migration proved challenging because the disciplines involved in migration studies mostly focused on certain types of migration, labor migration in particular (Boyd & Grieco, 2003).

This theoretical negligence is paired with a methodological one. Many migration estimation methods and models apply the same assumptions to predict female as well as male migrations. A major obstacle to developing more gender-nuanced approaches is the lack of comprehensive sources for international migration data disaggregated by sex or gender (Hennebry, KC, & Williams, 2021)[2]. Migration flow data are very relevant

[1] We refer to *flows* to describe a number of people moving from one country to another in a specific time period. With the word *corridor* we describe a migration corridor that is established between two countries over the entire study period.

[2] Gender relations and not sex determine the relative numbers of females and males migrating (Donato & Gabaccia, 2015), therefore we refer in the remainder of the article to *gender differences* instead of *sex differences*. At the same time, the data set used for the analysis refer to the term sex and therefore the reader will encounter both terms in the remainder of this article.

for researchers as they depict in detail the direction (emigration or immigration) of the move and sometimes come with information on the characteristics of the mover. In reality, countries collect these data differently and apply different definitions of migrants (Donato & Gabaccia, 2015). Nevertheless, migration flow data are the best source to broaden our understanding of the "true" gender composition of migration (Donato & Gabaccia, 2015). Flow data capture the gender distribution at the moment of migration, whereas stocks capture sex distributions influenced by later gender-specific mortality of the migrants.

Due to the theoretical and methodological shortcomings described above, we expect to find differences in the performance of common migration models for predictions of female and male flows. The lack of theoretical concepts in combination with the scarcity of gender-disaggregated migration data might inevitably result in worse migration predictions by gender and other inconsistencies observed in migration data. In order to assess our expectations, we use bilateral migration flow estimates by sex provided by Abel and Cohen (2022). We employ common deterministic and probabilistic models to make out-of-sample predictions for female and male migrants and evaluate and compare them. Later, we engage with theory and case studies to explain and contextualize our findings. Thereby, we aim to shed light on gender biases in methods and theories, to understand their structures and origins and to suggests how they can be mitigated and how gender theory can inform migration theory and methods.

2 Results

Gravity-type models perform poorly in predicting migration flows, even more so in non gender-balanced flows. Such models view migration as a function of population sizes of the destination and origin country and the distance between them and are widely used for predicting migration as their results can be interpreted rather intuitively. Extended versions of gravity models can include socio-economic and demographic covariates for the origin and destination (Kim & Cohen, 2010). We remove the last time period from the data set and fit the models to the remaining time points. Then we make out-of-sample predictions for the last period in the sample. Employing the extended gravity model structure proposed by Kim and Cohen (2010) yields different results for male and female migration flows. Predicting the number of women in worldwide migration flows with the linear gravity model results in a Mean Absolute Percentage Error (MAPE) that is about ten times larger than for the predictions achieved through the simple data-driven deterministic methods. The gravity-like hurdle model produces even larger deviations from the true migration flows. The MAPE for predicting the number of migrant women is double the MAPE of the gravity model. Predicting the number of men in worldwide migration flows results in systematically lower MAPE values than for female flows. However, the same discrepancies crystallize when comparing the results to

the deterministic methods and the hurdle model. In general, These differences are even more striking when comparing the probabilistic methods to simple deterministic ones. This finding is consistent with evidence in the literature (Welch & Raftery, 2022) and illustrates very well how information from theory does not necessarily improve the estimation. In order to further examine the errors by the share of females in each corridor, we employ the typology of Donato and Gabaccia (2015). This typology identifies three categories of gender composition in migration flows. Namely, male predominant flows (share of females below 47%), female predominant flows (share of females above 53%), and gender-balanced flows (share of females ranges from 47-53%). As Table 1 shows, the predicted number of female migrants in male predominant flows is less accurate than for males and vice versa for both probabilistic methods.

Many factors may cause these discrepancies. First, the assumptions applied in the models are fixed across genders and so are the covariates. However, as we will show in the remainder of this section, female and male migrants may behave differently in response to the same typical gravity model components. Second, gender-disaggregated data on which data-driven approaches could be developed are scarce and oftentimes indirectly estimated themselves. For the present analysis, we use migration flow estimates derived from stocks. Migrant stocks do not always come with a gender-breakdown and only few of the flows could be validated against other data. Therefore, this adds another layer of uncertainty to our analysis of these estimates. Third, theory-informed approaches like the gravity model contain gender biases from the underlying migration theories which may propagate throughout the prediction results. In other words, any gender difference that is not captured by theory will also be ignored by the models, and can translate into different results, as we observe them in Table 1.

	female predominant		**gender balanced**		**male predominant**	
	female	*male*	*female*	*male*	*female*	*male*
Historical mean	163.73	97.1	103.87	99.76	100.49	233.44
Persistence	116.88	75.72	75.65	74.57	69.82	116.27
Gravity model	1370.67	1598.85	1041.38	1043.79	1778.51	1391.11
Hurdle model	30825.2	34012.53	18740.94	18452.69	37302.54	30602.2

Table 1: MAPE results comparing the predicted values to the observed data for the pseudo-Bayes estimates. We removed flows below a total number of 500 migrants.

Understanding patterns of gender-composition diversity in migration corridors through case studies. Every migration flow is embedded in the history of its migration corridor and is a product of socioeconomic factors and political or natural events in the point of origin, as well as promising factors in the destination steered by a global system of border control and enforcement. Considering some cases studies enables us to gain insights into the specific context from which potential gender differences arise. Figure 1

presents the share of females in four different large-scale migration corridors along with the total trends in the number of migrants. While the corridors in Panel A and D seem to be predominantly gender balanced over time, the trends in the number of migrants vary drastically. While Panel B exhibits a negative trend in the share of females over time, Panel C exhibits an overall positive one. As most classical theories of migration ignore gender norms and roles and the gendered access to resources in the origin and destination of the potential migrant, they cannot fully explain the patterns in Figure 1. Migration from Mexico to the US began in the 1940s with the recruitment of farm labourers and was therefore initially male predominant (Alba, 2024). Researchers have studied this corridor extensively and found that Mexican men are more likely than women to migrate to the US (Hamilton, 2015; Chort, 2014). But it has seen a steady rise in female migrants over time. The Mexico-US corridor still is the largest migration corridor in the world (with 150,000 immigrations in 2022 (Moslimani & Passel, 2024)) and its very long history is mainly attributed to the proximity of the two countries and the difference in their levels of economic development. All traditional components of a gravity-like model and initiation frameworks like the Neoclassical Economic Theory of migration and the New Economics of Labor Migration Theory seem to do well in explaining this observed behavior, with the first one considering individual decision-making and the latter considering households as decision-making agents. However, both theories neglect within-household gender hierarchies, gender-specific obstacles, and safety considerations on the journey. This is supported by the finding that Mexican women seem to rely on male networks to make their migration decision and to move to the US, while they rely on female networks to choose a destination settle in it (Davis & Winters, 2001).
An example of a highly fluctuating share of female migrants is the corridor from Poland to Germany. Migration from Poland to Germany is characterized by low-skilled seasonal agriculture and domestic work. Female migrants from Poland make up almost half of the 24h care providers in Germany's aging society (Lutz, Palenga-Möllenbeck, & Benazha, 2021). EU accession in 2004 has probably contributed to higher migration flows from Poland to Germany but also to more seasonal migrations due to freedom of movement. The declining share of females could be attributed to better economic opportunities in Poland's growing economy. In terms of transnational care chains, the population dependency ratio included in a gravity model setup might entail heterogeneous effects for potential female and male migrants due to traditional gender roles in care work. The segmented labor markets theory can explain a substantial portion of the migration taking place in this corridor. However, women are in general more likely than men to be employed in the secondary sector of any economy and often they also work in occupations that are regarded as low-paid (and low-status) but vital. In the foreign work force there might be a clear hierarchy between the genders, as gender wage gaps propagate throughout different classes of the society. Findings from the literature point to the double burden of discrimination experienced by female migrants on the destination labor

market (Ryazantsev, Rostovskaya, & Peremyshlin, 2019; Hayfron, 2002; Lopez, 2012). These gendered considerations are not included in the theory nor in the gravity-like models where economic opportunities are usually modelled through origin and destination country's GDP.

Panel C describes the corridor from Morocco to Spain. Migration from Morocco to Spain became notable in the 1980s when labor demand in the agricultural, construction, and service sector in Spain rose. Early migrants were usually unmarried males searching for employment and far less females usually entering as family migrants. The number of female labor migrants increased only in recent waves due to higher demand in domestic work and due to family reunification programs by the Spanish government (Liu, Esteve, & Treviño, 2019). The significant drop in immigration in the 2010-2015 period can be attributed to the financial crisis in Spain. In this corridor characterized by family migration, family networks seem to play a more important role in facilitating female migration due to gender-specific roles and considerations (Heering, van der Erf, & van Wissen, 2004). The network theory of migration considers the importance of networks for facilitating migration. However, differences in terms of network types by gender are not considered by that theory nor by the gravity-like models yet. Therefore, the plain number of fellow nationals in a country of destination may be insufficient in predicting the number of migrants of different genders.

The migration corridor between Venezuela and Colombia has been characterized by humanitarian migration in recent years. Starting in the mid 2010s, Venezuela slid into the deepest crisis a country in peace has ever experienced. While traditionally Venezuela's thriving oil economy attracted Colombian labor migrants and refugees of the Colombian conflict, the tides have turned by 2015. Millions of Venezuelans left their country en mass due to a total collapse of the national economy caused by mismanagement and crude oil dependence (Rossiasco et al., 2023). The share of females remained balanced over the years, underlining the humanitarian and forced displacement character of this exodus. Classical theories of migration generally neglect the plethora of motivations for migration other than labor. In particular, motivations that are known to be more evenly distributed across genders like education, family-reunification or forced migration, are overlooked by these explanations.

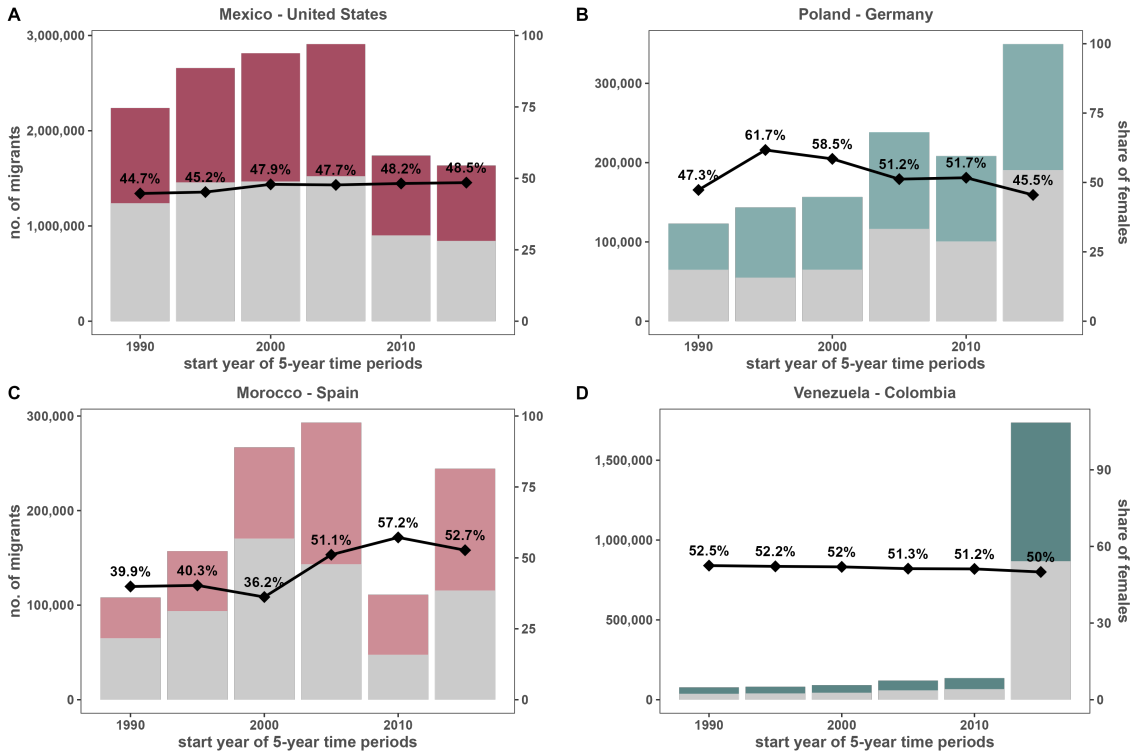

Figure 1: Number of migrants in the four corridors by year on the right y-axis. The colored bars indicate the number of female migrants. On the left y-axis, the black lines indicate the share of females on overall migrants in a given corridor and time.

Higher gender equality at origins and destinations is associated with more female-predominant flows. Gender inequalities can influence the decision and ability to migrate through gender relations and hierarchies, gender status and roles in the society of the origin country. In the destination country of the migrant, structural gender inequalities can translate into labor market discrimination and other forms of discrimination (Boyd & Grieco, 2003). For instance, previous studies found that improved gender inequality in the origin labor market, and higher economic rights in origin countries, can increase the migration of women (Baudassé & Bazillier, 2014; Neumayer & Plümper, 2021). Moreover, female migrants can experience two types of labor market discrimination in the destination country, one based on their migrant status as well as one based on their gender (Lopez, 2012; Hayfron, 2002; Alfarhan & Al-Busaidi, 2020). In a similar manner, panels A and B in Figure 2 show that higher gender equality at origins and destinations is associated with more female-predominant migration flows in our data set. More than half of the flows originating and arriving in countries with low gender equality are male-predominant. These observations suggest that levels of gender equality in the origin and destination countries can have retaining as well as enabling impacts on migrants of different genders and their migration trajectories. Classical pull- and push- frameworks reach their limits in accounting for such heterogeneous gendered impacts. The migration and gender literature also points at educational attainment as driver of women's migration intentions and moves (Kroehnert & Vollmer, 2012; Bau-

dassé & Bazillier, 2014; Heering et al., 2004). Having higher education can increase the desire to migrate (especially from a place with low returns to education) and increase access to the resources necessary to facilitate migration. Panel D in Figure 2 shows that more female-predominant flows originate in countries with higher educational attainment of women. Thus, predicting the gender-composition of flows may also be determined by educational attainment of women in the origin country. Panel C in Figure 2 suggest that more female-predominant flows take place in South to North migrations as well as in migrations within the Global North. According to classical migration theories, we would expect to see more male-predominant migration flows from South to North. The gender composition of some of these corridors might be partially owed to transnational care provisioning (as mentioned in one of the example corridors above), a phenomenon poorly captured by classical migration theories but determining a substantial amount of female labor migration (Kofman & Raghuram, 2012). Women significantly contribute to the survival of households in economies that are destabilized by transformations and withdrawal of public welfare in both the Global South as well as in the Global North (for instance, during the financial crises in 2010) (Christou & Kofman, 2022). While the data used for this analysis could only be validated for a number of countries in the Global North, more people migrate within the Global South than from South to North (Nawyn, 2016; Abel & Sander, 2014). Therefore, most likely we miss many interesting gender patterns to obtain a complete picture.

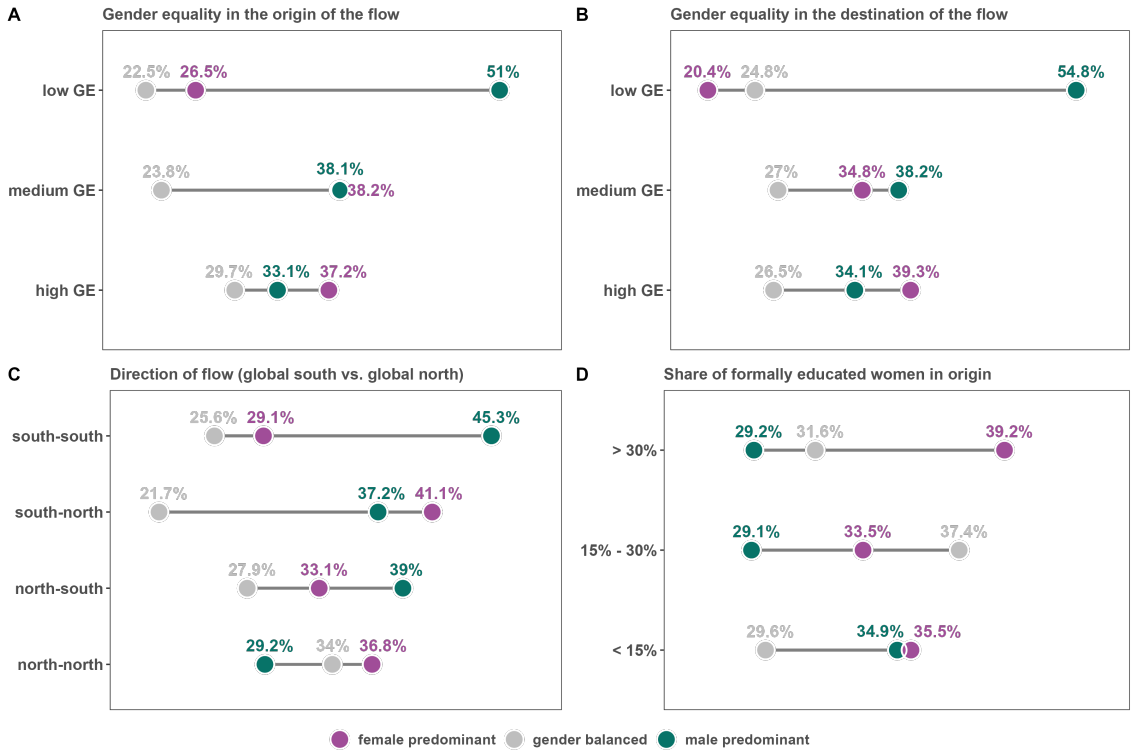

Figure 2: Gender composition of migration flows by different characteristics of origin and destination countries. In Panels A and B, gender equality (GE) is determined by the Gender Inequality Index for 2022 as provided in the United Nations Human Development Reports. In Panel C, global south and global north are derived from the World Bank definitions of high-income countries (GDP above 14,005 USD per capita). In Panel D, the share of educated women denotes the share of females in the population that achieved a Bachelor's degree or equivalent.

Globalization patterns of international migration differ by gender. Key dimensions of migration globalization were introduced by Czaika and De Haas (2014) and encompass *migration intensity* as represented by the share of the mobile population, *migration spread* as represented by the diversity in origin and destination countries of migrants, and *migration distance*, as represented by the average distance of migratory moves. All three dimensions feed into the composite migration globalization index (Czaika & De Haas, 2014). Observing those dimensions over time and by gender reveals the patterns shown in Figure 3. This process of globalization can be driven by technological and political change, for instance through increased channels of communication and interconnectedness, as well as by reduced mobility constraints and changes in labor demands and political systems (Czaika & De Haas, 2014). Migration is more intense when the relative numbers of people on the move rise. Panel B in Figure 3 shows how the rate of migration changed from 1990 to 2020 by gender. For both genders the rates exhibit a similar time trend but at a smaller rate for female than for male migrants. This goes against the common perception of an overall increase in international migration, but also against a common expectation of a feminization of migration. Migration spread can be evaluated from two perspectives, the diversification of emigration patterns

(namely, the growing number of migration destinations) and the diversification of immigration patterns (the growing number of migration origins). Panels C and D in Figure 3 illustrate these two globalization dimensions for both genders. The global immigrant spread exhibits a similarly positive trend over time for female and male migration with the female immigration spread lying slightly below the male spread indicating a diversification of migration origins regardless of gender. On the other hand, the global emigrant spread reveals slightly diverging trends by gender. For the female sample, it continuously declines until the 2010-2015 period and then accrues. This means that female emigration was, from 1990 to 2015, directed to a smaller set of destination countries and started diversifying only during the last time period. For the male sample, the global emigrant spread follows a similar pattern until it plummets to a very low level in the period 2005-10, recovers for the following period slightly to drop again to its lowest level in the period 2015-20. This indicates that male emigration was consistently directed to a smaller set of destination countries. This is partially in line with global trends observed by Czaika and De Haas (2014) with migrant stock data from 1960 to 2000. The authors conclude that immigrant populations came from an increasingly diverse array of origins while they concentrate in an increasingly small number of destinations over time (Czaika & De Haas, 2014). These developments can have several reasons. One is that some countries have transformed from emigration to immigration countries (consider,for instance, the above examined corridor from Venezuela to Colombia), the decreasing significance of post-colonial migration patterns, decreasing emigration restrictions and more (Czaika & De Haas, 2014).

A traditional predictor of gravity models for migration among other origin and destination country factors is geographical distance and it is also one of the three key dimensions of migration globalization as defined by Czaika and De Haas (2014). While geographical distance is often the only variable that proxies the process of the migrants' journey in traditional gravity-type models of migration, it might be a less relevant factor in a globalized world compared to other forms of distance, like cultural, linguistic, and legal distances (Czaika & De Haas, 2014). And its relevance might also be nuanced by gender. Migrant women were found to be less sensitive to this default determinant of a gravity model (Beine & Salomone, 2013). Entry regulations in destination countries and gender-specific migration recruitment policies are often shaped by traditional gender roles and stereotypical views of women's societal roles, influencing how migrants move across borders (Boyd & Grieco, 2003). Panel A in Figure 3 shows how the average distance covered by migrants regardless of their gender increased between 1995 and 2010, it shrank dramatically in the 2010-2015 period and did not recover in the same way for female migration.

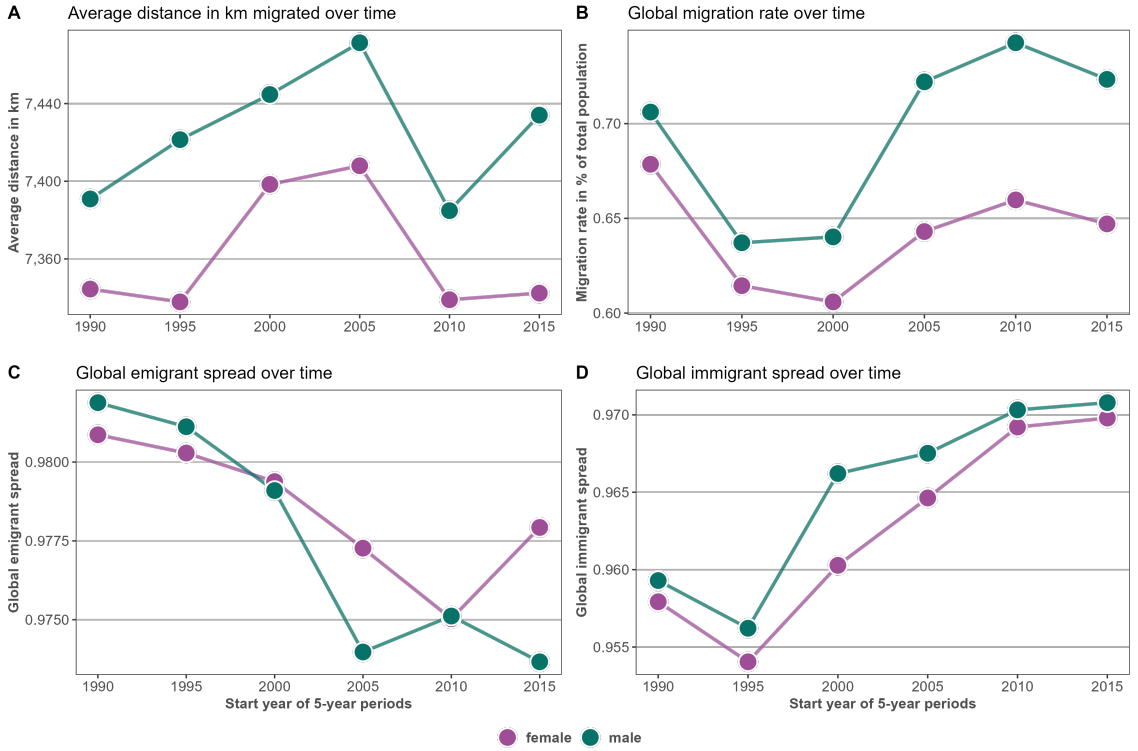

Figure 3: Panel A describes the average distances covered by the migrants in migration flows larger than zero. In Panel B, the global migration rate is calculated by dividing the number of migrants by the world population of that year. The *global emigrant spread* measures the level at which global migration is dispersed across destination countries, while the *global immigrant spread* measures the extent to which global migration comes from different origin countries.

Exploring the patterns of distance by gender composition of migration flows results in Figure 4. All types of flows appear over a wide range of distances and seem to be more clustered around shorter distances. However, heavily male predominant flows are clearly happening over shorter distances, while female predominant flows seem to be equally distributed over the range of all possible distances.

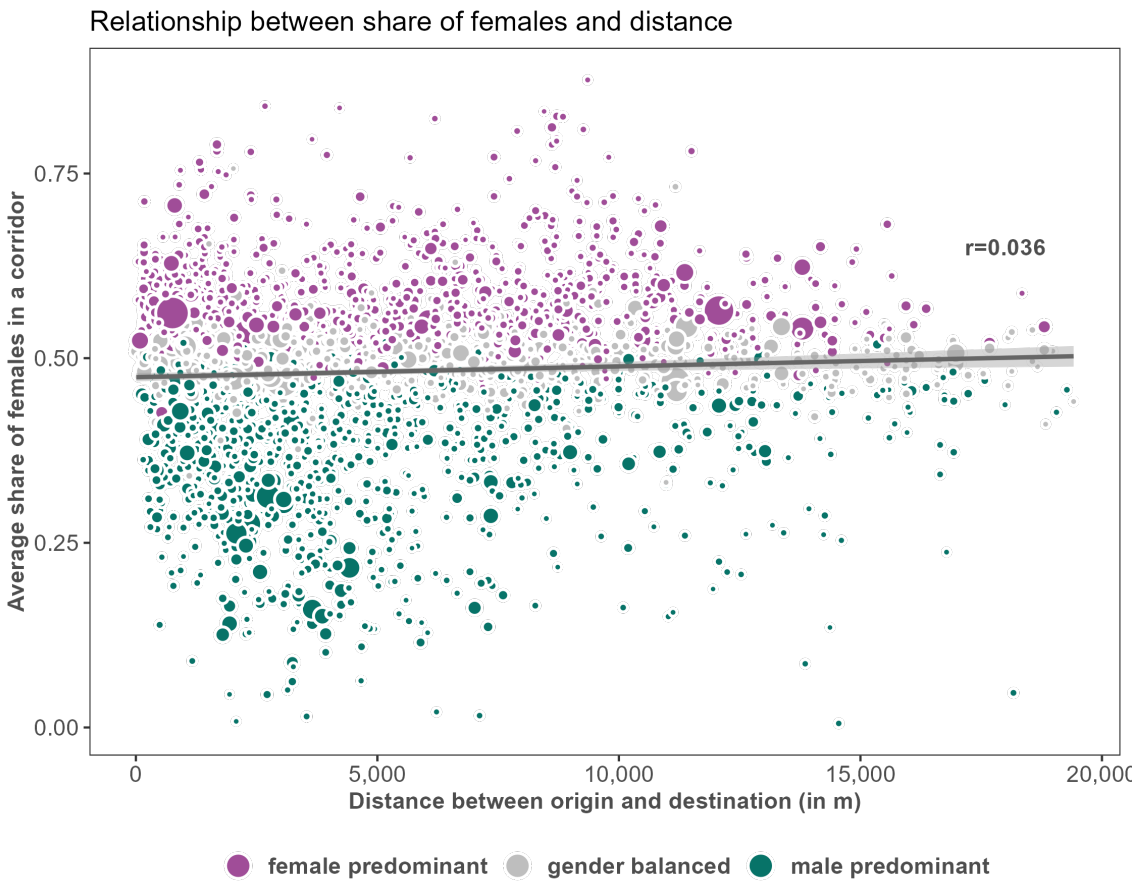

Figure 4: This figure illustrates the relationship between distance and the average share of females in all flows of a migration corridor (only total flows larger than 500 migrants). The gender-composition type was assigned based on the most common type over time in each corridor. The size of the dots indicates the average number of migrants in the respective corridor.

3 Discussion

Demographers can still learn a lot from viewing migration theories and methods through the prism of gender inequalities. As the present work aims to show, classical migration theories and common methods for migration predictions neglect the dimension of gender which inevitably leads to less accurate outcomes and expectations by gender. The main result of our analysis suggests that common methods for migration flow predictions perform differently for migration corridors with different gender compositions. And even more so did the probabilistic methods. In particular, predictions of the number of female migrants in female-predominant flows are more accurate than for males in female-predominant flows and vice versa. In order to investigate the origins of such aggregate discrepancies, we chose four selected migration corridors to illustrate gender dynamics that can explain changes in the gender composition of migration flows. By descriptively exploring the bilateral migration flow estimates derived by Abel and Cohen (2022) we concluded that many variables included in common gravity-like models

for migration prediction can influence the migration decisions of migrants differently by gender. Thereby, we explicitly criticize gravity-type models in their theoretical and statistical foundations for neglecting the dimension of gender. Data-driven methods can only be as good as the data they are based on. Methods that claim to incorporate theoretical insights perform poorly in explaining patterns of diversity in the gender composition of migration flows. We recommend researchers to question whether the assumptions of their models hold in the same way for different types of corridors with different gender compositions.

We encourage future research to consistently analyze patterns by gender where data allow for that and to thoughtfully curate the variables of predictive models based on their gender-sensitivity. Hegemonic masculinity in migration research can be challenged by incorporating insights from gender theory into existing explanatory frameworks for migration, but also by developing new theories that are able to capture current patterns of globalization of international migration. This, together with enhanced gender-disaggregated migration data collection efforts, will enable migration scholars to develop methods that perform equally well for all genders. When still employing gravity-type models for their straightforward interpretation and easy applicability, we encourage researchers to include gender-sensitive predictors alongside the traditional covariates (e.g., from the Gender Statistics Database of the World Bank).

This work does not comes without limitations. The picture drawn by the bilateral flow estimates should be interpreted with caution. Since the estimates were derived based on migrant stocks from time points that are five years apart, some gendered patterns might be obscured. Corridors like those used as case studies in this work are known for enhanced seasonal and return migration patterns. Moreover, Abel and Cohen (2022) use sex-specific inputs, namely sex-specific population counts, numbers of births and deaths where available, to derive the sex-specific bilateral flow estimates. For counties where information on the sex characteristics of migrants are missing, UN DESA imputed the data based on a regional or country model (Abel & Cohen, 2022). Not capturing gender-differences directly but inferring them might additionally lead to a general underestimation of these differences. All the assumptions have not been altered for female and male migration flows across the different estimation methods. Potential sources of gender bias in the estimation methods therefore include measurement error and imputation methods applied to the stock estimates by UN DESA. Another important limitation of the data set is the missing information about return migration flows.

With this work, we aim to inform migration theory and estimation methodology by critically examining the steps in the prediction process and by highlighting their gaps and shortcomings. We hope to inspire migration scholars and demographers to critically view the data and methods they use for their work and to encourage governments and international organizations to enhance gender-disaggregated data collection efforts so we can more thoroughly understand patterns of international migration by gender in the

future. Because we believe that migration *"[...] statistics too are a language that can be analyzed with the methods of gender studies." (Donato & Gabaccia, 2015, p.11)*.

4 Data and Methods

For the analysis, we made use of the sex-disaggregated bilateral migration flow estimates derived by Abel and Cohen (2022) based on the pseudo-Bayesian method. We chose this data set as it has maximal coverage and unlike other migration flow data sets is broken down by gender.

4.1 Deterministic Methods

Two simplistic deterministic extrapolation methods are employed in the analysis of the present work to predict the size of the migration flow in the next time period. One of them is the *historical mean flow model* which describes the mean of all previous flows (Welch & Raftery, 2022).

$$\bar{m}_{i,j,t} = \frac{1}{T}\sum_{t}^{T-1} m_{i,j,t} \tag{1}$$

where i denoted the origin, j denotes the destination, and t denotes the time period in which the flow takes place.

The second deterministic method is the *persistence model* which basically projects the most recent observation of a migration flow to the next period.

$$\bar{m}_{i,j,t} = m_{i,j,t-1} \tag{2}$$

4.2 Probabilistic Methods

The first probabilistic method employed for analysis is the *gravity model*. This model is popular in the social sciences to predict aggregate migration flows and it is based on Newton's law of gravity (Kim & Cohen, 2010). This relationship is consistent with Ravenstein's laws of migration stating that migration between two places is inversely related to the distance between them and proportional to their population sizes (Willekens, 2016).

$$\begin{aligned} log(m_{i,j,t}) = \beta_0 + \beta_1 log(GDP_{i,t}) + \beta_2 log(GDP_{j,t}) + \beta_3 log(D_{i,j}) + \beta_4 log(PSR_{i,t}) \\ +\beta_5 log(PSR_{j,t}) + \beta_6 log(URB_{i,t}) + \beta_7 log(URB_{j,t}) + \beta_8 log(IMR_{i,t}) \\ +\beta_9 log(IMR_{j,t}) + +\beta_{10} log(LA_i) + \beta_{11} log(LA_j) + \beta_{12} LB_{i,j} \\ +\beta_{13} OL_{i,j} + \beta_{14} COL_{i,j} + \beta_{15}(t-2005) + \beta_{16}(t-2005)^2 + \epsilon_{i,j,t} \end{aligned} \tag{3}$$

For analysis, we included country-level data derived from the World Development Indicators on the population of the countries of origin and destination POP, the share of the population residing in urban areas URB, the infant mortality ratio IMR, the potential support ratio (i.e. the number of persons aged 15-64 per person aged 65+ multiplied by 100) PSR, and the land area in squared kilometers LA. All these variables were obtained from the World Bank's World Development Indicators at the start of each period t (World Bank, n.d.). Indicators on shared land borders between countries LB, shared official language OL, and ever existing colonial relationship between the countries COL were obtained from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) gravity database (Conte, Cotterlaz, & Mayer, 2022). Geographic coordinates were obtained from the *countryref* dataset in the *coordinate cleaner* package in R and used for calculating the bilateral distances to the countries' capital cities (Zizka et al., 2019).

Another probabilistic method suitable for modelling zero-inflated count data is the *Hurdle model* (Welch & Raftery, 2022). We apply it in a gravity-like version with a binomial zero component and a Poisson count component.

$$m_{i,j,t}|m_{i,j,t} > 0 \sim Poisson(\lambda_{i,j,t}) \tag{4}$$

$$\mathbf{1}_{m_{i,j,t}>0} \sim Binomial(1, \omega_{i,j,t}) \tag{5}$$

with a covariance matrix X containing the same covariates as for the gravity model

$$log\lambda_{i,j,t} = X_{i,j,t}\beta \tag{6}$$

$$logit\omega_{i,j,t} = X_{i,j,t}\gamma \tag{7}$$

4.3 Evaluation

In order to evaluate the accuracy of the model predictions, we rely on the out-of-sample mean absolute error for point forecasts as applied in Welch and Raftery (2022) to evaluate the predictions across models.

$$MAPE(M, \tilde{M}) = \frac{100}{F} \sum_{i \neq j} \frac{\mid m_{i,j} - \tilde{m}_{i,j} \mid}{m_{i,j} + 1} \tag{8}$$

with F denoting the total number of flows contained in the flow matrix M. This measure has the advantage that it puts the error into context by considering the total size of the given flow.

4.4 Global migration spread

The global immigrant and emigrant spread were calculated in the following way according to Czaika and De Haas (2014).

$$ES^{global} = 1 - \Sigma_{i=1}^{n} \left(\frac{EM_i}{M} \right)^2 \tag{9}$$

$$IS^{global} = 1 - \Sigma_{i=1}^{m} \left(\frac{IM_i}{M} \right)^2 \tag{10}$$

with n denoting the number of destinations and m denoting the number of origins. M denotes the sum of all immigrants and emigrants respectively.

4.5 Validation

The same models have been applied to three other data sets. Results can be found in the Annex in Table 2.

References

Abel, G. J., & Cohen, J. E. (2022). Bilateral international migration flow estimates updated and refined by sex. *Scientific Data*, *9*(1), 1–11.

Abel, G. J., & Sander, N. (2014). Quantifying global international migration flows. *Science*, *343*(6178), 1520–1522.

Alba, F. (2024, May). Mexico at a crossroads once more: Emigration levels off as transit migration and immigration rise. *Migration Information Source*. Retrieved from `https://www.migrationpolicy.org/article/mexico-crossroads-emigration-transit`

Alfarhan, U. F., & Al-Busaidi, S. (2020, May). Women's earnings between migration status and glass ceilings: a double penalty? *Applied Economics Letters*, *27*(8), 629–632. doi: 10.1080/13504851.2020.1728222

Anastasiadou, A., Kim, J., Sanlitürk, A. E., de Valk, H., & Zagheni, E. (2023). Sex-and gender-based differences in the migration process: a systematic literature review. *MPIDR Working Papers*(WP-2023-039).

Baudassé, T., & Bazillier, R. (2014). Gender inequality and emigration: Push factor or selection process? *International Economics*, *139*, 19–47.

Beine, M., & Salomone, S. (2013). Network Effects in International Migration: Education versus Gender. *The Scandinavian Journal of Economics*, *115*(2), 354–380. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9442.2012.01733.x` doi: https://doi.org/10.1111/j.1467-9442.2012.01733.x

Boyd, M., & Grieco, E. (2003). Women and Migration: Incorporating Gender into International Migration Theory. , 7.

Chort, I. (2014). Mexican migrants to the us: What do unrealized migration intentions tell us about gender inequalities? *World Development*, *59*, 535-552. Retrieved from https://www.sciencedirect.com/science/article/pii/S0305750X14000631 doi: https://doi.org/10.1016/j.worlddev.2014.01.036

Christou, A., & Kofman, E. (2022). Gender and migration: An introduction. In *Gender and migration: Imiscoe short reader* (pp. 1–12). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-91971-9_1 doi: 10.1007/978-3-030-91971-9_1

Conte, M., Cotterlaz, P., & Mayer, T. (2022). *The CEPII Gravity database.* CEPII Working Paper No 2022-05. Retrieved from http://www.cepii.fr/DATA_DOWNLOAD/gravity/doc/Gravity_documentation.pdf

Czaika, M., & De Haas, H. (2014). The globalization of migration: Has the world become more migratory? *International migration review*, *48*(2), 283–323.

Davis, B., & Winters, P. (2001). Gender, Networks and Mexico-US Migration. *The Journal of Development Studies*, *38*(2), 1–26. doi: 10.1080/00220380412331322251

Donato, K. M., & Gabaccia, D. (2015). *Gender and International Migration*. Russell Sage Foundation. (Google-Books-ID: D8iZBgAAQBAJ)

Hamilton, E. R. (2015). Gendered disparities in mexico-us migration by class, ethnicity, and geography. *Demographic Research*, *32*, 533–542.

Hayfron, J. E. (2002). Panel estimates of the earnings gap in Norway: do female immigrants experience a double earnings penalty? *Applied Economics*, *34*(11), 1441–1452. doi: 10.1080/00036840110101429

Heering, L., van der Erf, R., & van Wissen, L. (2004, March). The role of family networks and migration culture in the continuation of Moroccan emigration: a gender perspective. *Journal of Ethnic and Migration Studies*, *30*(2), 323–337. doi: 10.1080/1369183042000200722

Hennebry, J., KC, H., & Williams, K. (2021). Gender and migration data: A guide for evidence-based, gender-responsive migration governance.

Kim, K., & Cohen, J. E. (2010, December). Determinants of International Migration Flows to and from Industrialized Countries: A Panel Data Approach beyond Gravity. *International Migration Review*, *44*(4), 899–932. Retrieved 2023-10-24, from https://doi.org/10.1111/j.1747-7379.2010.00830.x (Publisher: SAGE Publications Inc) doi: 10.1111/j.1747-7379.2010.00830.x

Kofman, E., & Raghuram, P. (2012). Women, migration, and care: Explorations of diversity and dynamism in the global south. *Social Politics*, *19*(3), 408–432.

Kroehnert, S., & Vollmer, S. (2012). Gender-Specific Migration from Eastern to Western Germany: Where Have All the Young Women Gone? *International Migration*, *50*(5), 95–112. doi: 10.1111/j.1468-2435.2012.00750.x

Liu, C., Esteve, A., & Treviño, R. (2019). The living arrangements of moroccans in

spain. *Demographic Research*, *40*, 1063–1096.

Lopez, M. J. (2012). Skilled Immigrant Women in the US and the Double Earnings Penalty. *Feminist Economics*, *18*(1), 99–134. doi: 10.1080/13545701.2012.658429

Lutz, H., Palenga-Möllenbeck, E., & Benazha, A. (2021). The recruitment of migrant carers within the german welfare state. *Research Features*. Retrieved from https://researchfeatures.com/wp-content/uploads/2021/08/Helma-Lutz.pdf

Moslimani, M., & Passel, J. S. (2024, July). What the data says about immigrants in the u.s. *Pew Research Center*. Retrieved from https://www.pewresearch.org/short-reads/2024/07/22/key-findings-about-us-immigrants/#:~:text=Where%20are%20U.S.%20immigrants%20from,and%20El%20Salvador%20(3%25).

Nawyn, S. J. (2016). *Migration in the global south: Exploring new theoretical territory* (Vol. 46) (No. 2). Taylor & Francis.

Neumayer, E., & Plümper, T. (2021). Women's economic rights in developing countries and the gender gap in migration to germany. *IZA Journal of Development and Migration*, *12*(1).

Riosmena, F. (2024). Worlds in motion redux? expanding migration theories and their interconnections. *Population and Development Review*.

Rossiasco, P., de Narváez, P., Aguilera, A., Granados, G., Guerra, P., & Samad, T. (2023). Adapting public policies in response to an unprecedented influx of refugees and migrants: Colombia case study of migration from venezuela. *Background paper to the World Development Report*.

Ryazantsev, S., Rostovskaya, T., & Peremyshlin, S. (2019, 12). Gender aspects of labor migration in Russia: trends, implications, regulation. *Woman in russian society*, 53–65. doi: 10.21064/WinRS.2019.4.5

Welch, N. G., & Raftery, A. E. (2022, August). Probabilistic forecasts of international bilateral migration flows. *Proceedings of the National Academy of Sciences*, *119*(35), e2203822119. Retrieved 2022-08-30, from https://www.pnas.org/doi/abs/10.1073/pnas.2203822119 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.2203822119

Willekens, F. (2016). Migration flows: Measurement, analysis and modeling. *International handbook of migration and population distribution*, 225–241.

World Bank. (n.d.). *World development indicators.* Retrieved from https://databank.worldbank.org/source/world-development-indicators

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., ... Antonelli, A. (2019). Coordinatecleaner: standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*(10), -7. Retrieved from https://github.com/ropensci/

`CoordinateCleaner` (R package version 3.0) doi: 10.1111/2041-210X.13152

5 Appendix

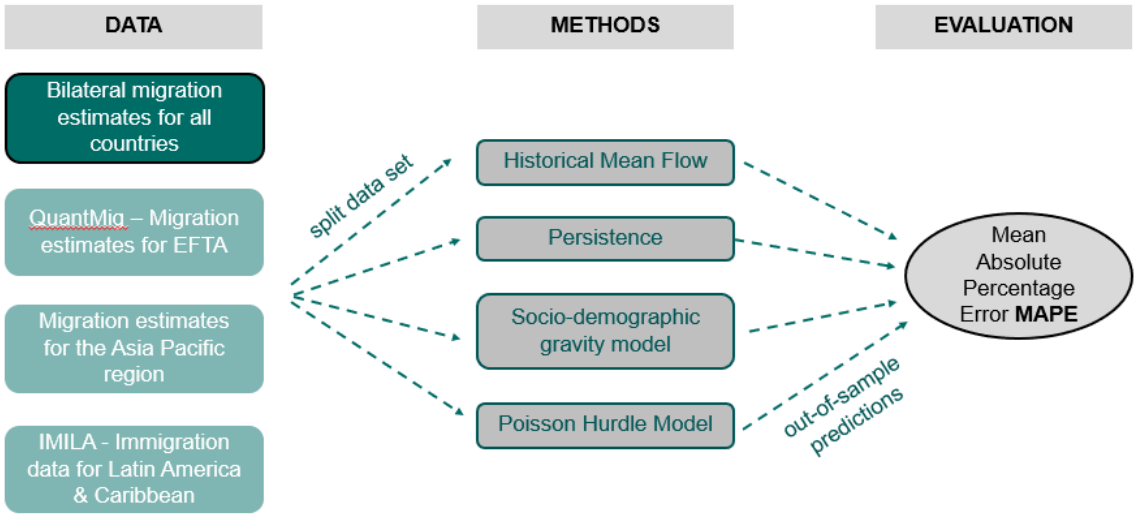

Figure 5: Methodological approach.

	female predominant		gender balanced		male predominant	
	female	*male*	*female*	*male*	*female*	*male*
Global bilateral migration flows	1370.67	1598.85	1041.38	1043.79	1778.51	1391.11
QuantMig estimates	220.67	295.45	342.41	330.66	377.5	290.24
Asia Pacific estimates	1452.6	1752.28	947.17	936.2	3168.13	2400.35
Latin America & Caribbean	2157.32	4449.78	2838.05	2839.77	2146.78	1771.71

Table 2: MAPE results of the gravity model across additional bilateral migration flow data sets disaggregated by gender.