A model of dynamic flows: Explaining Turkey's inter-provincial migration*

Ozan Aksoy[†]& Sinan Yıldırım[‡]

May 3, 2023

Abstract

The flow of resources across nodes over time (e.g. migration, financial transfers, peer-to-peer interactions) is a common phenomenon in sociology. Standard statistical methods are inadequate to model such interdependent flows. We propose a hierarchical Dirichlet-multinomial regression model and a Bayesian estimation method. We apply the model to analyse 25,632,876 migration instances that take place between Turkey's 81 provinces from 2009 to 2018. We then discuss the methodological and substantive implications of our results. Methodologically, we demonstrate the predictive advantage of our model compared to its most common alternative in migration research, namely the gravity model. We also discuss our model in the context of other approaches, mostly developed in the social networks literature. Substantively, we find that population, economic prosperity, the spatial and political distance between the origin and destination, the strength of the AKP (Justice and Development Party) in a province and the network characteristics of the provinces are important predictors of migration, while the proportion of ethnic minority Kurds in a province has no positive association with in- and out-migration.

Keywords: Hierarchical Bayesian modelling; Dirichlet-multinomial regression; Markov chain Monte Carlo; Migration; Turkey

1 Introduction

Flows of items such as people, resources, and information across a finite number of units such as places, institutions, and individuals over time constitute a common type of data in sociology and the wider social sciences. While migration is one of the most frequent types of such flows, other examples include financial transfers between institutions or individuals, passengers through

*The authors contributed equally to the manuscript. We thank Zsofia Boda and Burak Sönmez for their comments on earlier drafts. Aksoy acknowledges financial support from the British Academy (grant no. SRG20/200045). An anonymised replication package with the data and the code that produce the results and with guidelines as to how our model can be fitted to other data are available at https://github.com/SocNetMigration/MigNet-MATLAB-code.git.

†Centre for Quantitative Social Science, UCL Social Research Institute, University College London, 55-59 Gordon Square, WC1H 0NU, London UK, email: ozan.aksoy@ucl.ac.uk.

‡Faculty of Engineering and Natural Sciences, Sabancı University, Orta Mahalle, 34956 Tuzla, İstanbul, Turkey, email: sinanyildirim@sabanciuniv.edu.

transportation systems, mobility tables, and the number of times participants trust or allocate a fixed resource to others during an experiment. Due to the highly interdependent nature of such situations, certain modelling problems arise (e.g. destinations 'compete' over a fixed amount of flow) which are not straightforward to address with standard statistical models (Block et al., 2022).

Here we first aim to contribute methodologically to the analysis of such dynamic flows. We frame our model within a migration context. Indeed we developed our model with a motivation to analyse real-world migration data. Nevertheless, our model can be applied to other situations that similarly involve dynamic discrete flows across a finite number of units. One of the most widely used methods of analysing migration flows is the so-called *gravity model* (Barthélemy, 2011; Karemera et al., 2000; Expert et al., 2011). In this model, the (logarithm) of the number of people migrating from i to j is modelled as a function of the characteristics of i and j, and the distance between i and j. However, this approach does not account for the fact that when a person migrates from i to j, they mechanistically cannot migrate to another destination k. Likewise, the characteristics of k may affect migration from i to j, too. For example, k can be a competitor of j in receiving migration from i. This dependence in migration probabilities across destinations is ignored in the gravity model, while it is taken into account in our model.

In this study, we propose a Dirichlet-multinomial model, which is an extension of a class of econometric discrete choice models (Guimaraes and Lindrooth, 2007; Alamá-Sabater et al., 2017). Our model differs from those existing choice models in the treatment of non-migration. The existing models base their estimates *conditional* on migration taking place. We believe that any inference about the causes of migration would be problematic if one ignores those who choose to not move. This is because in that case, observations for which the causes of migration might have been present but did not generate migration, i.e., non-migration counts, are omitted. In other words, ignoring non-migration results in 'selection on the dependent variable'. In our model, the probabilities of migration to the available destinations as well as the probability of not migrating are directly modelled as functions of the covariates. We further extend our model with random intercepts for provinces to account for the longitudinal nature of migration flows. We fit our model within the Bayesian framework via a Markov chain Monte Carlo (MCMC) algorithm.

There is a long array of models that are developed mostly in the social networks literature which can also be applied to dynamic flows. These include Stochastic Actor Oriented Models (Snijders, 1996, 2001), Dynamic Network Actor and other similar Relational Event Models (Stadtfeld and Block, 2017; Stadtfeld et al., 2017b; Block et al., 2019), Exponential Random Graph Models (Lusher et al., 2013) and its various extensions and forms (Westveld and Hoff, 2011; Desmarais and Cranmer, 2012; Almquist and Butts, 2014; Krivitsky and Handcock, 2014; Krivitsky and Butts, 2017; Block et al., 2022), Latent Space and Latent Factor Models (Hoff et al., 2002; Minhas et al., 2019). Some of these models are theoretically flexible and very general but may be difficult to adapt to the specific case; others are computationally demanding. The application of some will require significant programming. Nevertheless, those models offer a powerful way of dealing with relational data. We believe that our approach provides a relatively straightforward and natural way to deal with dynamic flows, at the same time offering computational simplicity while taking into account key features of the data. We discuss similarities and differences between our model and those network models in Section 2.2. We believe that we contribute to the methodological literature by enriching the methodological arsenal for dealing with dynamic flows.

We also aim to contribute substantially to the migration literature. Various determinants of voluntary migration have been identified in this literature (Boyle et al., 1998). These include characteristics of the origin and the destination such as economic prospects (Borjas, 1999; Levy, 2010), immigration legislation (Palmer and Pytlikova, 2015); and characteristics of the origin-destination pair such as linguistic, cultural and physical distance (Levy, 2010; Windzio, 2018; Expert et al., 2011) between the origin and the destination. Migration can be *international* as well as *intra*-national. Yet, the migration literature is based mostly on international migration and hence the dynamics of internal migration, particularly in non-Western countries are understudied (Kuhn, 2015; Bell and Muhidin, 2009). The focus on international migration is understandable, for it seems highly consequential in shaping public opinion and politics (Chan et al., 2020; Gimpel and Schuknecht, 2001). The scarcity of research on internal migration, however, is surprising. Internal migration may not be as economically advantageous for the migrant or topical as international migration is. However, it is far less costly and risky. Consequently, compared to international migration, much larger shares of populations are affected by internal migration, particularly in non-Western countries (Kuhn, 2015). For example, according to the official statistics, in 2018 a total of 323,918 people emigrated abroad from Turkey, our study context, whereas in the same year 3,057,606, nearly ten times more, moved from one Turkish province to another. Internal migration is thus a strong determinant of local population structures, and it is important to understand its drivers.

We contribute to the migration literature by addressing this gap. Using a unique dataset compiled from the administrative, registry, and survey data, we analyse the dynamics of more than 25 million migration instances between the 81 provinces of Turkey from 2009 to 2018. Almost all empirical work on internal migration in Turkey, however scarce, was conducted with data spanning until 2000 (Gedik, 1997; Filiztekin and Gökhan, 2008; Yazgi et al., 2014). The more recent work is descriptive and focuses on larger inter-regional migration (Akın and Dökmeci, 2015). Turkey has gone through a large economic, demographic, and political transformation in the meantime (Aksoy and Billari, 2018), which has affected migration too (Çoban, 2013). It is thus unknown if these earlier insights into Turkey's internal migration apply to more recent migration patterns. Internal migration in Turkey is relevant for Europe too. Turkey's ascension to the European Union (EU) has stalled. One of the fears as to the expansion of the EU is large-scale immigration (Strasser, 2008). Understanding the determinants of inter-province migration in Turkey will help predict the extent of potential migration from Turkey to Europe and its likely destinations within the EU, should Turkey ever be part of the EU (Filiztekin and Gökhan, 2008).

Overall, by providing novel insights into the internal migration of Turkey and a convenient model of migration flows, our study contributes to the literature both substantively and methodologically.

2 A model of migration flows

We now describe our model in a migration context. Note, however, that our model can be applied to many other settings with minimal adjustment, e.g. provinces below can be replaced by individuals or institutions, and migration can be replaced by the number of financial transfers or interactions. In fact, in many of those alternative settings, one could directly proceed with modelling the $H_t(i,j)$'s below. The migration case needs a slight adjustment, as we will describe now, for the stock of people who could migrate but did not is dynamic due to changes in births and deaths.

Assume that we have $n > 1$ provinces and $T > 1$ consecutive years and that for each year we have data on the population of and migration between the provinces. Specifically, we observe

$$P_t(i) = \text{the population of province } i \text{ at the beginning of year } t$$

for all $i = 1, \dots, n$ and $t = 1, \dots, T$; and

$$H_t(i,j) = \text{the number of people who migrated from province } i \text{ to province } j \text{ in year } t$$

for all $1 \leq i, j \leq n$ pairs with $i \neq j$, respectively. Let us define

$$E_{t,t+1}(i) := P_{t+1}(i) - \left[P_t(i) - \sum_{j \neq i} H_t(i,j) + \sum_{j \neq i} H_t(j,i) \right], \quad i = 1, \dots, n,$$

which is the net change in population for province i after taking into account the inter-province migration during year t. The net change captures births, deaths, and people who are registered in the system for the first time for any other reason. Taking the net change into account, we define the *adjusted population* of province i for year t as

$$X_t(i) := P_t(i) + E_{t,t+1}(i) = P_{t+1}(i) + \sum_{j \neq i} H_t(i,j) - \sum_{j \neq i} H_t(j,i), \quad i = 1, \dots, n.$$

We consider this adjusted population to find the number of people who did *not* migrate from province i during year t as

$$H_t(i,i) = X_t(i) - \sum_{j \neq i} H_t(i,j), \quad i = 1, \dots, n.$$

Therefore, given the assumptions, we can summarise the relation between the quantities $H_t(i,j)$s and $X_t(i)$s as

$$X_t(i) = \sum_{i=1}^{n} H_t(i,j), \quad i = 1, \dots, n. \tag{1}$$

As a prelude to our proposed model, let us assume for now the existence of a probability of migrating from province i to province j in year t, denoted by $\rho_t(i,j)$, and that $\rho_t(i,j)$ is same for all citizens in province i. This yields a multinomial distribution for $(H_t(i,1), \dots, H_t(i,n))$ for each t and i, specified as

$$(H_t(i,1), \dots, H_t(i,n)) \sim \text{Multinomial}(X_t(i); \rho_t(i,1), \dots, \rho_t(i,n)).$$

A multinomial regression can be constructed by modelling $\rho_t(i,j)$ as a function of the factors about provinces i and j individually, as well as the factors about the relation between i and j. Generically, we assume that there are $d_u \geq 1$ and $d_v \geq 1$ factors that affect the likelihood of migrating from and to a province, respectively. Those factors can vary across the years and provinces, hence are denoted by $u_t(i) \in \mathbb{R}^{d_u}$ and $v_t(i) \in \mathbb{R}^{d_v}$, respectively. Note that some factors can both push and pull migration (e.g. (un)employment) and hence appear both as $u_t(i)$ and $v_t(i)$. Furthermore, for each (i,j) pair we also have $d_z \geq 1$ factors of sort $z_t(i,j) \in \mathbb{R}^{d_z}$ defined as pair-level factors that may affect migration from i to j. Those factors altogether determine the probability $\rho_t(i,j)$ of migrating from i to j in year t as

$$\rho_t(i,j) = \frac{\exp\{\theta_0 + \theta_1 \cdot u_t(i) + \theta_2 \cdot v_t(j) + \theta_3 \cdot z_t(i,j)\}}{1 + \sum_{j' \neq i} \exp\{\theta_0 + \theta_1 \cdot u_t(i) + \theta_2 \cdot v_t(j') + \theta_3 \cdot z_t(i,j')\}}.$$

Here, $(\theta_0, \theta_1, \theta_1, \theta_3)$ is the vector of model parameters. The components $\theta_1 \in \mathbb{R}^{d_u}$, $\theta_2 \in \mathbb{R}^{d_v}$, and $\theta_3 \in \mathbb{R}^{d_z}$ correspond to the sending ($\rightarrow$), the receiving ($\leftarrow$), and the joint (pair-level) factors ($\leftrightarrow$), respectively. The scalar component $\theta_0 \in \mathbb{R}$ is the 'base' parameter, which solely determines the probability of migration at all when all factors are zero. Finally, exponentiation is used to ensure that we have positive terms. The resulting model is an instance of the multinomial logistic regression model (Theil, 1969).

A limitation of the above expression may be the deterministic relation between the factors and the migration probability $\rho_t(i,j)$. We would like to allow for variability in those probabilities, such that even under the same factor values the probabilities $\rho_t(i,j)$s may differ across province-year pairs, for example, due to unsystematic residual factors that our model may fail to cover (Nelson, 1984). This can be done by modelling the probabilities using a Dirichlet distribution for each of those province-year pairs. Specifically, we let $\rho_t(i,j)$s be random variables themselves, with $(\rho_t(i,1), \dots, \rho_t(i,n))$ having a Dirichlet distribution independently for each t and i, that is,

$$(\rho_t(i,1), \dots, \rho_t(i,n)) \sim \text{Dirichlet}(\alpha_t(i,1), \dots, \alpha_t(i,n)).$$

This time, it is the parameters of the Dirichlet distribution that we model via regression.

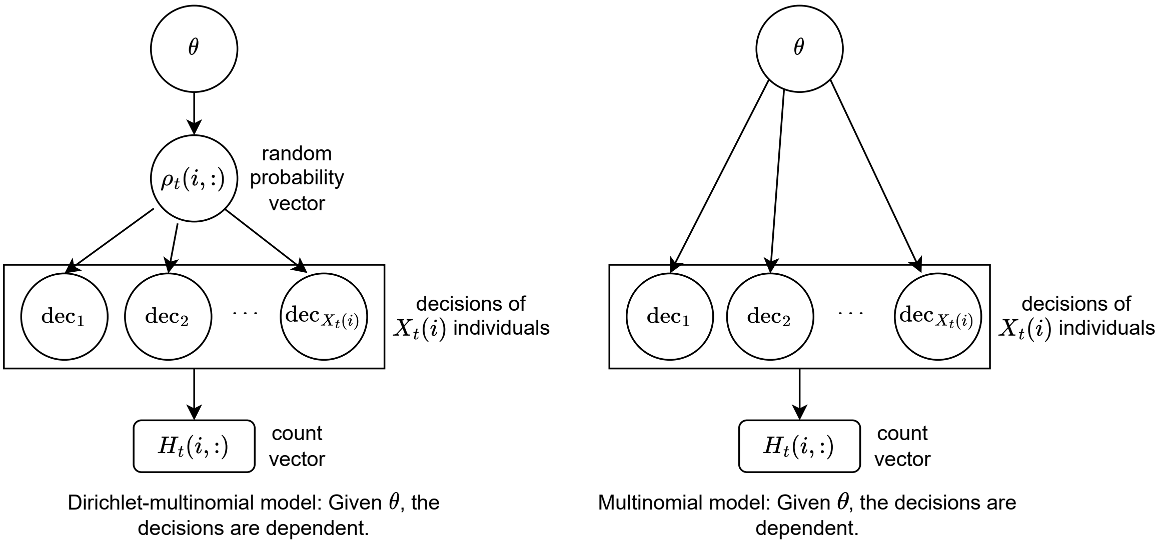

Figure 1: Left: Dirichlet-multinomial model, Right: Multinomial model

Specifically,

$$\alpha_t(i,j) = \begin{cases} \exp\{\theta_0 + \theta_1 \cdot u_t(i) + \theta_2 \cdot v_t(j) + \theta_3 \cdot z_t(i,j) + \theta_4\} & i \neq j \\ \exp\{\theta_4\} & i = j. \end{cases} \tag{2}$$

The resulting model is a Dirichlet-multinomial regression. Observing the expected probabilities

$$E(\rho_t(i,j)) = \frac{\alpha_t(i,j)}{\sum_{j'=1}^{n} \alpha_t(i,j')},$$

we note that, in this model, the vector $(\theta_0, \theta_1, \theta_2, \theta_3)$ lends itself to a similar interpretation as in the multinomial logistic regression model introduced earlier. The extra scalar parameter $\theta_4 \in \mathbb{R}$ is present in each $\alpha_t(i,j)$, hence does not affect the expected probabilities. The parameter θ_4 instead captures the variances of the probabilities – the larger θ_4, the smaller the variances.

Due to its flexibility, we proceed with the Dirichlet-multinomial regression model described above. Note that, from an inference perspective, the Dirichlet-multinomial specification does not introduce additional complications compared with the multinomial logistic specification. As it is well known, after integrating out $\rho_t(i,j)$s, $H_t(i,j)$s follow a Dirichlet-multinomial distribution,

$$(H_t(i,1), \dots, H_t(i,n)) \sim \text{Dirichlet-multinomial}(X_t(i); \alpha_t(i,1), \dots, \alpha_t(i,n)), \tag{3}$$

which can easily be computed.

Figure 1 (left) depicts the Dirichlet-multinomial model for a single province and single time step. According to the Dirichlet-multinomial model, people's decisions *within a province* are *dependent* even when conditioned on θ and the covariates. This is because, given θ and the covariates, the same *random* migration probability vector $\rho_t(i,:)$, drawn from Dirichlet distribution, applies to all individuals in the same province, introducing a positive dependency among their choices. This is in contrast to the simpler multinomial model (Figure 1, the model on the right), where the decisions *are* independent given θ.

It is possible to quantify the mentioned dependency in the Dirichlet-multinomial model in terms of closed-form conditional distributions. For example, given that an individual I_1 has migrated to province j from province i during time period t, the probability distribution of the decision of another individual I_2 follows a categorical distribution with

$$\Pr(I_2 \text{ migrates } i \to k | I_1 \text{ migrates } i \to j) = \begin{cases} \frac{\alpha_t(i,j)+1}{1+\sum_{j'=1}^{n} \alpha_t(i,j')} & \text{if } k = j \\ \frac{\alpha_t(i,k)}{1+\sum_{j'=1}^{n} \alpha_t(i,j')} & \text{if } k \neq j \end{cases}, \quad k = 1, \ldots, n.$$

As another example, given t among $P > 0$ individuals the migration counts to the n provinces are $A_{1:n} = a_{1:n}$ (so that $a_1 + \ldots + a_n = P$), the probability distribution of the migration counts $B_{1:n}$ of $R > 0$ other individuals in the same province are jointly distributed as

$$B_{1:n} | (R, A_{1:n} = a_{1:n}) \sim \text{Dirichlet-multinomial}(R; \alpha_t(i,1) + a_1, \ldots, \alpha_t(i,n) + a_n)$$

Both examples indicate positive dependency among the individuals' decisions in the same province at the same time period.

The presence of positive dependency between the decisions may be considered another positive feature of the Dirichlet-multinomial model, as it captures the possibility that individuals living in the same province may be influenced by each other in deciding to migrate to a particular destination or not migrate, for example through peer effects or familial migration decisions. Furthermore, the strength of this dependency is determined by θ_4: it decreases as θ_4 increases. Hence, we can gauge the amount of within-province interpersonal influence by estimating θ_4 using the data.

Hierarchical specification: Our data are multilevel, that is, migration from i to j for all i and j are observed over the $T > 1$ years. Accordingly, we expand the specification in (2) to address this multilevel structure by assuming that the base parameter θ_0 is a random variable defined at the province level (i.e. a 'random intercept'). More specifically, we modify equation (2) as

$$\alpha_t(i,j) = \begin{cases} \exp\{\theta_0(i) + \theta_1 \cdot u_t(i) + \theta_2 \cdot v_t(j) + \theta_3 \cdot z_t(i,j) + \theta_4\} & i \neq j \\ \exp\{\theta_4\} & i = j. \end{cases} \tag{4}$$

Next, by our aim of Bayesian inference, we build up the prior distributions for the model parameters introduced so far. The parameters θ_1, θ_2, θ_3, and θ_4 are assumed *a priori* independent and have priors $\theta_i \sim \eta_i$, $i = 1, \ldots, 4$. Further, independently from θ_1, θ_2, θ_3, and θ_4, the base parameter $\theta_0(i)$ for each $i = 1, \ldots, n$ are also assumed independent with a common normal distribution

$$\theta_0(i) \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu_0, \sigma_0^2\right), \quad i = 1, \ldots, n. \tag{5}$$

The parameters μ_0 and σ_0^2 are themselves treated as random variables, that are independent with $\mu_0 \sim \mathcal{N}(\mu_h, \sigma_h^2)$ and $\sigma_0^2 \sim \mathcal{IG}(a_h, b_h)$, where the latter is the inverse-Gamma distribution with shape and scale parameters a_h and b_h, respectively. We assume that the hyperparameters μ_h, σ_h^2, a_h, and b_h are known.

The full parameter vector of the resulting hierarchical model is the $(1 + 1 + n + d_u + d_v + d_z + 1) \times 1$ vector

$$\boldsymbol{\theta} = (\mu_0, \sigma_0^2, \theta_0(1), \dots, \theta_0(n), \theta_1, \theta_2, \theta_3, \theta_4).$$

This completes the description of our model of migration in a closed system. We refer to our model whose Dirichlet-multinomial parameters are given by (4) as a *hierarchical specification*, for it models the base parameter $\theta_0(i)$ as random across provinces.

Note that one could expand the model above by making $\theta_0(i)$, the base parameter, a polynomial function of time, of order $d_b \geq 0$. In that case, the base parameter could be expressed as

$$\theta_0(i) = \sum_{k=0}^{d_b} \theta_{0,k}(i) t^k,$$

where the polynomial coefficients for province i would be modelled as random with

$$\vartheta_i := (\theta_{0,0}(i), \dots, \theta_{0,d_b}(i)) \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu_0, \Sigma_0\right), \quad i = 1, \dots, n,$$

and μ_0, Σ_0 would be multivariate random variables with suitable priors, such as $\mu_0 \sim \mathcal{N}(\mu_h, \Sigma_h)$ and $\Sigma_0 \sim \mathcal{W}^{-1}(\nu_h, \Psi_h)$, the inverse-Wishart distribution with ν_h degrees of freedom and the scale matrix Ψ_h. With $d_b = 1$, for example, we would have a model with 'random slopes' of time.

In our implementations, we will not present results with this polynomial specification. Our results below will be restricted to the 'random intercept' specification where the base parameters are random across provinces as in (5). This is mainly because province log populations, which are included in the model as predictors, are increasing almost as a linear function of time, rendering the estimation of a separate time slope unnecessary and numerically difficult. We did fit specifications with linear as well as quadratic random time slopes which did not make noticeable differences in our models, apart from creating convergence problems. That being said, the polynomial specification for the base parameter may be useful in other applications; hence we will describe our inference method below by keeping general and considering the polynomial specification.

2.1 Inference

Before presenting the study context and data, we describe our inference method. To quantify the uncertainties in the estimation of $\boldsymbol{\theta}$ conveniently, we aim for Bayesian inference of $\boldsymbol{\theta}$ given the data

$$\mathcal{D} = \{H_t(i, j); \quad i, j \in \{1, \dots, n\}, t \in \{1, \dots, T\}\}.$$

Note that the (adjusted) province populations $\{X_t(i) : t = 1, \dots, T; i = 1, \dots, n\}$ are implicitly contained in $\mathcal{D}$ also, thanks to the relation given in (1). More specifically, we aim to find the posterior distribution for the hierarchical model

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\boldsymbol{\theta})p(\mathcal{D}|\boldsymbol{\theta}).$$

Here, $\boldsymbol{\theta} = (\mu_0, \Sigma_0, \vartheta_1, \dots, \vartheta_n, \theta_1, \theta_2, \theta_3, \theta_4)$ is the parameter vector with the hierarchical specification for the polynomial coefficients of the province-dependent base parameters; $p(\boldsymbol{\theta})$ is the probability density of the prior distribution given by

$$p(\boldsymbol{\theta}) = \left[\prod_{j=1}^{4} \eta_j(\theta_j)\right] \left[\prod_{i=1}^{n} \mathcal{N}(\vartheta_i; \mu_0, \Sigma_0)\right] \mathcal{N}(\mu_0; \mu_h, \Sigma_h)\mathcal{W}^{-1}(\Sigma_0; \nu_h, \Psi_h);$$

and $p(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood given by a product of Tn Dirichlet-multinomial probabilities as

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{t=1}^{T}\prod_{i=1}^{n} p(H_t(i,:)|\alpha_t^{\boldsymbol{\theta}}(i,:)), \tag{6}$$

where the superscript $\boldsymbol{\theta}$ over $\alpha_t^{\boldsymbol{\theta}}(i,:)$ is used from now on to indicate the dependency on $\boldsymbol{\theta}$ explicitly and $p(H_t(i,:)|\alpha_t^{\boldsymbol{\theta}}(i,:))$ is the Dirichlet-multinomial probability

$$p(H_t(i,:)|\alpha_t^{\boldsymbol{\theta}}(i,:)) = \frac{X_t(i)!\Gamma\left(\sum_{j=1}^{n}\alpha_t^{\boldsymbol{\theta}}(i,j)\right)}{\Gamma\left(X_t(i) + \sum_{j=1}^{n}\alpha_t^{\boldsymbol{\theta}}(i,j)\right)} \prod_{j=1}^{n} \frac{\Gamma\left(H_t(i,j) + \alpha_t^{\boldsymbol{\theta}}(i,j)\right)}{H_t(i,j)!\Gamma\left(\alpha_t^{\boldsymbol{\theta}}(i,j)\right)}$$

where $\Gamma(\cdot)$ is the gamma function.

As the posterior distribution is analytically intractable, we develop an MCMC algorithm (see e.g. Tierney (1994)) to sample from the posterior distribution. The developed MCMC algorithm is an instance of Metropolis-Hastings-within-Gibbs (MHwG). The parameter vector $\boldsymbol{\theta}$ is divided suitably into disjoint blocks of components, and those blocks are updated in turn by either a Gibbs move (if sampling from the full conditional distribution of the block is possible) or a Metropolis-Hastings move (otherwise). We describe this process in detail in Algorithm 1, where the aforementioned blocks are taken as

$$\{\theta_1\}; \{\theta_2\}; \{\theta_3\}; \{\theta_4\}; \{\vartheta_1, \dots, \vartheta_n\}; \{\mu_0\}; \{\Sigma_0\}.$$

We note some important remarks on the computational complexity of Algorithm 1. While updating each of $\{\theta_1\}; \{\theta_2\}; \{\theta_3\}; \{\theta_4\}$, a product of nT Dirichlet-multinomial distributions, each with n categories, is computed in the numerator and the denominator of the acceptance ratio. Given the rest of the components and the observations, the variables $\vartheta_1, \dots, \vartheta_n$ are conditionally independent, therefore they can be updated in parallel, each requiring the computation of T Dirichlet-multinomial distributions, each with n categories, in the numerator and the denominator of the acceptance ratio. Finally, the blocks $\{\mu_0\}$ and $\{\Sigma_0\}$ can be updated in turn with

Gibbs moves since their full conditional distributions have closed forms, thanks to conjugacy.

Algorithm 1: MHwG for the Dirichlet-multinomial model with the hierarchical specification for its province-dependent base parameters

Input: Migration counts H_t and external factors u_t, v_t, z_t for $t = 1, \dots, T$; hyperparameters μ_h, Σ_h, ν_h, Ψ_h, prior distributions η_j for $j = 1, \dots, 4$, proposal distributions q_j for $j = 1, \dots, 4$, initial value $\boldsymbol{\theta}^{(0)}$

Output: Samples $\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(K)}$

Start with $\boldsymbol{\theta}^{(0)}$.

for $k = 1, 2, \dots$ **do**

Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k-1)}$.

for $j = 1, 2, 3, 4$ **do**

Propose $\theta'_j \sim q_j(\theta'_j | \theta_j)$

Construct the proposed parameter vector $\boldsymbol{\theta}'$ by replacing θ_j by θ'_j in $\boldsymbol{\theta}$.

Update $\boldsymbol{\theta}$ by replacing θ_j by θ'_j with acceptance probability

$$\min \left\{ 1, \frac{q_j(\theta_j | \theta'_j) \eta_j(\theta'_j) \prod_{t=1}^{T} \prod_{i=1}^{n} p(H_t(i,:) | \alpha_t^{\boldsymbol{\theta}'}(i,:))}{q(\theta'_j | \theta_j) \eta_j(\theta_j) \prod_{t=1}^{T} \prod_{i=1}^{n} p(H_t(i,:) | \alpha_t^{\boldsymbol{\theta}}(i,:))} \right\};$$

otherwise, keep θ_j as before.

for $i = 1, \dots, n$ **do**

Propose $\vartheta'_i \sim q_0(\vartheta'_i | \vartheta_i)$.

Construct $\boldsymbol{\theta}'$ by replacing ϑ_i by ϑ'_i in $\boldsymbol{\theta}$.

Update $\boldsymbol{\theta}$ by replacing ϑ_i by ϑ'_i with acceptance probability

$$\min \left\{ 1, \frac{q_0(\vartheta_i | \vartheta'_i) \mathcal{N}(\vartheta'_i; \mu_0, \Sigma_0) \prod_{t=1}^{T} p(H_t(i,:) | \alpha_t^{\boldsymbol{\theta}'}(i,:))}{q_0(\vartheta'_i | \vartheta_i) \mathcal{N}(\vartheta_i; \mu_0, \Sigma_0) \prod_{t=1}^{T} p(H_t(i,:) | \alpha_t^{\boldsymbol{\theta}}(i,:))} \right\};$$

otherwise, keep ϑ_i as before.

Sample $\mu_0 \sim \mathcal{N}(\mu_{\text{post}}, \Sigma_{\text{post}})$ where $\Sigma_{\text{post}} = \left(\Sigma_h^{-1} + n\Sigma_0^{-1}\right)^{-1}$ and $\mu_{\text{post}} = \Sigma_{\text{post}}(\Sigma_h^{-1}\mu_h + \Sigma_0^{-1} \sum_{i=1}^{n} \vartheta_i)$.

Sample $\Sigma_0 \sim \mathcal{W}^{-1}\left(\nu_h + n, \Psi_h + \sum_{i=1}^{n} (\vartheta_i - \mu_0)(\vartheta_i - \mu_0)^T\right)$

Store the new sample

$$\boldsymbol{\theta}^{(k)} = (\mu_0, \Sigma_0, \vartheta_1, \dots, \vartheta_n, \theta_1, \theta_2, \theta_3, \theta_4),$$

2.2 Related methodology

Our model's closest relative is the one proposed by Guimaraes and Lindrooth (2007). Guimaraes and Lindrooth (2007) develop a Dirichlet-multinomial regression as an extension of the multinomial logistic regression within a discrete choice random utility framework. They use this to model counts from different groups to multiple destinations (e.g. hospital choice of different groups of patients) whereby for each group multinomial probabilities for the destinations are modelled independently with a Dirichlet distribution. Note that our model can also be interpreted from a random utility framework. Consider, for example, the migration decisions of the actors at province i at time t. The quantity $\alpha_t(i, j)$ given in (2) can be interpreted as the deterministic part of a random utility an actor gets, had they moved from i to j at time t. The Dirichlet-multinomial model corresponds to the assumption that each actor makes the choice that maximises their utility, where the log-utility for the k'th actor at province i at time t had they moved to province j is in the form

$$\alpha_t(i,j) + \nu_t(i,j) + u_t(i,j,k).$$

Here, $\nu_t(i,j)$, $j = 1,\dots,n$ are the i.i.d. random utility components that affect identically all individuals at province i at time t, and, independently from those, $u_t(i,j,k)$, $j = 1,\dots,n$, $k = 1,\dots,X_t(i)$ are the i.i.d. random utility components at the individual level. (See Guimaraes and Lindrooth (2007) for the specific forms of the distributions of those random utility components.) In fact, the province-level random utility component $\nu_t(i,j)$, $j = 1,\dots,n$, can be interpreted as the source of the positive dependency between the decisions of the actors within the same province, which was mentioned earlier.

This Dirichlet-multinomial regression model is used by Alamá-Sabater et al. (2017) to explain immigration to different parts of Spain from various regions outside of Spain. Our model and the model of Alamá-Sabater et al. (2017) have critical differences. Firstly, Alamá-Sabater et al. (2017) aim to explain how much migration a certain factor 'pulls', hence only focus on factors of the receiving regions, (corresponding to our $v_t(j)$s, or the $\leftarrow$ factors). Whereas, in this study, we consider migration on a matrix of provinces. Consequently, our model includes factors in sending ($\rightarrow$) and receiving ($\leftarrow$) nodes as well as pair-level covariates ($\leftrightarrow$). A second crucial difference between our model and the model in both Guimaraes and Lindrooth (2007) and Alamá-Sabater et al. (2017) is that our model accounts for people who have *not* migrated, while the latter two omit 'self-edges'. Alamá-Sabater et al. (2017) model the flows of immigration to different provinces in Spain, conditional on immigration taking place. Likewise, Guimaraes and Lindrooth (2007) omit those who do not decide to migrate at all. We believe that an omission of self-edges and no-choice cases would create a serious issue in our setting, for in that case the analysis would be restricted to the cases for which a positive outcome is observed. A further difference between our model and those in Guimaraes and Lindrooth (2007) and Alamá-Sabater et al. (2017) is that our model is hierarchical, allowing a random intercept. Finally, while Guimaraes and Lindrooth (2007) and Alamá-Sabater et al. (2017) implement maximum likelihood estimation, we use Bayesian inference to quantify the uncertainties in our estimates effectively.

A common and notable feature of our model and those above is that they rely on the assumption of independence of irrelevant alternatives (IIA). This assumption implies that the *ratio* of the probabilities of a single actor moving from i to j and from i to k ($\frac{\rho_t(i,j)}{\rho_t(i,k)}$) depends only on the features of i, j, and k and is independent of the features of all other destinations. Note that the IIA is about the ratio of two probabilities. The plain probability of moving to a particular destination does depend on the features of all possible destinations in our model. While the IIA assumption can be violated in practice, there are existing, however imperfect, diagnostic tests (Cheng and Long, 2007). There are also extensions of the discrete choice models that relax this assumption (see, e.g. Benson et al. (2016)). An accepted practice is to use models that rely on IIA when the alternatives are plausibly distinct and independently weighed by the decision maker (Benson et al., 2016) and we believe the decision to migrate to a particular place plausibly fits these criteria.

The most common alternative model that is used to analyse migration is the co-called gravity model (Karemera et al., 2000; Barthélemy, 2011; Expert et al., 2011; Poot et al., 2016). In this model, the logarithm of the number of people moved from i to j at time T is simply regressed on the features of i and j such as population and some measure of the distance between i and j. The estimation is straightforward, for it can be performed within the standard maximum likelihood framework that is used for regression modelling. The disadvantages are as follows. Firstly, the model does not incorporate systemic effects, for example, origin populations constrain the total size of out-migration which is not directly incorporated in the model (Poot et al., 2016). In addition, destinations compete for migration, for if an individual migrates to j, they simultaneously cannot migrate to k. Naturally, then the properties of k should also affect the decision to migrate to j which are typically ignored in the gravity model. Note that this is a more binding assumption than the IIA discussed above which is about the ratio of two probabilities. Furthermore, non-migration (zero cell counts or the diagonals–those who stay) is mostly omitted in this model, conditioning the estimates on migration having taken place and creating a selection on the dependent variable, just as in the models of Guimaraes and Lindrooth (2007) and Alamá-Sabater et al. (2017) discussed above. Nevertheless, this is a very popular model of migration, hence we will compare the predictive power of our model with that of the gravity model below.

A class of models which share the Bayesian nature of our model and also deal with migration can be found in Azose and Raftery (2015) and Azose and Raftery (2018). The main aim of those models is to forecast country-level future net migration based on past migration via an auto-regressive hierarchical model. Azose and Raftery (2018) improve those forecasts by developing a procedure that estimates cross-country correlations in net migration rates. Those models differ from ours, in that explaining the determinants of migration is not part of their purpose.

The literature on social networks, too, offers a long list of statistical models that can be applied to flows. A canonical model is the Exponential Random Graph Model (ERGM) (Lusher et al., 2013) which takes the entire data as a network of nodes (e.g. provinces) and edges (e.g. migration flows). In ERGMs, the whole network is modelled as a function of node, edge, and network-level covariates. ERGM enjoys greater generality than our model; because, using it, one can naturally include various local or global network statistics as explanatory variables and incorporate complex contemporaneous dependencies in the data (i.e. network dependencies in a cross-sectional measurement). The standard ERGM deals with binary edges and cross-sectional settings, so it cannot be directly applied to our problem. ERGM, however, has been extended to deal with longitudinal data (Krivitsky and Handcock, 2014) and with count or multi-layered networks (Block et al., 2022; Krivitsky et al., 2020; Krivitsky and Handcock, 2014; Krivitsky, 2012) and the more general weighted edges as in Generalised Exponential Random Graph

Models (GERGM) (Desmarais and Cranmer, 2012).

While an ERGM type of model would be very flexible and general, it can be challenging to perform inference in ERGMs (Almquist and Butts, 2013, 2014), particularly for various generalisations of it. The major challenge may be having to sample the whole network (typically several times) at every iteration, which can only be done approximately, for example by using Gibbs sampling, unless the network is very small–note that we also use a Gibbs sampler, but in our case, the inference is much simpler as we will elaborate shortly. Consequently, while ERGM is generalised to deal with longitudinal data and weighted edges, most applications are constrained to either longitudinal binary networks, cross-sectional networks with weighted edges, or relatively small networks. Indeed, ERGMs have been applied to the study of migration flow networks (Windzio, 2018; Windzio et al., 2021; Leal, 2021). However, likely due to those computational complexities, the authors had to simplify the flows so that an ERGM can be fitted. For example, Windzio (2018); Leal (2021) dichotomise flows, Windzio et al. (2021) use valued edges but ordinal categories are imposed to migration flows and the data are analysed only cross-sectionally even though they are longitudinal. Block et al. (2022) develop a weighted ERGM for mobility networks, but the application is restricted to cross-sectional observations. Abramski et al. (2020) successfully applies the GERGM developed by Desmarais and Cranmer (2012) to study refugee migration patterns, but the setup is again simplified with only 12 countries and a cross-sectional analysis. We have also tried to implement an ERGM for count edges (Krivitsky, 2012; Krivitsky et al., 2021) to our data. Unfortunately, it failed to converge even for a single year of our migration data, let alone the 10 years our data spans.

A further issue that needs to be tackled in an ERGM-type model is that the total flow of out-migration is bounded by the population of the origin, and this constraint may be difficult to implement in an ERGM. In our model, these 'row-sums' are naturally handled through the Dirichlet-multinomial distribution.

We must add that there are exciting recent developments in the ERGM literature, especially the pseudo-likelihood parameter estimation procedures for weighted ERGMs (Huang and Butts, 2021) and its applications to migration networks (Huang and Butts, 2022) are fast developing. These developments can alleviate the computational and other practical constraints for count data and make ERGMs more feasible for the study of migration flows.

Another class of models that are developed in the social networks literature that have some similarities with our model comprise the Stochastic Actor-Oriented Model (SAOM) (Snijders, 2001), the Dynamic Network Actor Model (DyNAM) (Stadtfeld and Block, 2017), the Relational Event Model (REM) (Butts, 2008) and other similar models (Almquist and Butts, 2014). These models are designed to analyse longitudinal dynamic binary network data, though extensions for count data exist (Stadtfeld et al., 2017a). In the actor-based SAOM and DyNAM, two processes are modelled separately: the timing of an event (e.g. a tie to be formed), and the target of the event (e.g. whom a person sends a tie) which is modelled with a multinomial specification. In the tie-based REM, the timing and the position of an event (e.g. which two nodes are connected) are modelled simultaneously. In our model, the exact timing of migration is unknown, apart from it taking place during a given year. Moreover, many migration flows take place during a given year. To apply the dynamic network models discussed here, however, one needs to be able to calculate the conditional distribution of the migration data of a year given the state of variables in the previous year. This is not available in our case, since we do not know the times of individual migration events and integrating them out is analytically intractable. One can only approximate those conditionals, making this class of models inconvenient for us to apply in our case.

There is a final class of models, developed again mainly in the social networks literature. These include latent space (Hoff et al., 2002; Hoff and Ward, 2004) and latent factor models (Minhas et al., 2019). Most recently, and perhaps most relevantly, Minhas et al. (2019) describe a latent factor model in which cell outcomes in a matrix (e.g. migration flows in our case) are modelled as a function of the sender and proposer characteristics plus a latent factor matrix. The elements of this matrix are the sender and receiver latent factors multiplied by a sender-receiver pair parameter. This latent multiplicative factor captures higher-order dependence structures which are not explained by the observed covariates included in the model. Minhas et al. (2019) then propose a Bayesian estimation procedure which assumes certain prior distributions for the parameters, including the elements of the latent factor matrix. This class of models is very flexible too. Applications, however, are often restricted to binary data, though an extension to the multinomial and count data is possible. In addition, diagonals are often not modelled explicitly (see, e.g. Minhas et al. (2019) and it is not straightforward to constrain the row-sums which are bounded in our data by the province populations.

Overall, we argue that the methodology and the modelling approach of these relational and network models are too general for our problem. In contrast, our specific model is directly related to the migration (and other similar) flows and can be derived from a small number of assumptions (such as Multinomial probabilities and independence of the actors in different provinces). For example, staying within the notation in our manuscript, the SAOM in Snijders (1996) takes (H_t, Z_t) as the state and assumes that this is a Markov process. H_t corresponds to the relation matrix in SAOM, and Z_t the exogenous variables. However, in our model, (H_t, Z_t) is not a Markov process, in that H_t depends on the cumulative effect of the previous $H_{1:t-1}$'s, since those sum up to the populations. In this sense, our model is similar to the model discussed by Hanneke et al. (2010) which conditions the current network on the earlier network realisations. Consequently, the parameters of our model can be inferred with a fairly standard MCMC algorithm, see Algorithm 1. This convenience is due to the assumption in our model that out-migration in two *different* provinces are independent, *conditional* on covariates which include factors related to migration in the previous year (however, recall that our model addresses the dependency of migration decisions among people who live in the same province). This assumption is plausible in the migration context: when a person in province i is deciding to migrate to a certain province j (or not to migrate at all) at time point t, this decision is likely independent of another person's decision in province $k \neq i$ at *the same time-point*, provided that we take key relevant factors into account. Note that in our model a person's decision in province i *can* depend on another person's decision in province j that was made in the previous year, for we include factors that are functions of past migration as predictors of future migration. For example, if we think that a person in province i is more likely to migrate to province j if others from other provinces also migrate to j, we can include this dynamic in the model explicitly as we will do below by adding a *popularity* measure of a province. Hence, we can ameliorate possible misfits due to the conditional independence assumption by including relevant covariates in the model. Also note that in our model a person's decision in province i *can* depend on another person's decision in the same province and time period, due to the Dirichlet-multinomial specification as discussed above.

We should also add that the conditional independence assumption of our model is less problematic, the more frequent the temporal measurements are. We have annual migration data here, which we believe is frequent enough given that most migration decisions are not made on a whim. If, however, the researcher has only cross-sectional data or data that are collected very infrequently, such as decades apart, then complex dependencies in migration flows between actors from different provinces may not be captured in our model. In such cases with cross-sectional data or data collected decades apart, however, the researcher would have a simpler data structure for which the more complex count ERGM-type models may work better.

We believe that those models developed in the social networks literature reviewed above are rich, very general, flexible; and in principle can be applied to the type of data we have here, especially given the rapidly developing work on estimation in valued ERGMs. A systematic comparison of all possible alternative modelling approaches, however, is beyond the scope of the current study.

3 Determinants of migration

We now move on to the substantive discussion of the expected drivers of migration and describe the Turkish context to which our data belongs.

Migration is a complex phenomenon with multiple causes, hence there is no single theory for it. A non-exhaustive list of the determinants of voluntary migration can be grouped into economic, demographic, geographical, cultural, and social network factors (Boyle et al., 1998). In classical economic theory, income maximisation relative to the costs of moving is a key driver of migration (Borjas et al., 1992). Economic prospects, such as a high per capita income, employment opportunities, or wages thus should pull migrants. Conversely, poor economic prospects should push migration.

Migration flows are also shaped by population. The larger the population of the origin and the destination, the larger will be the flow between the two (Barthélemy, 2011; Levy, 2010). This prediction is due to the empirical regularity that flows correlate positively with stocks in either direction (Poot et al., 2016). A large population means that there is more capacity to send and receive migrants.

While the increasing population may increase migration flows, the spatial distance between the origin and the destination constrains it. Next to the gravity model, spatial network models, too, show that distance affects ties between individuals, see, e.g. Butts et al. (2012). The effect of spatial distance operates via various channels (Schwartz, 1973). First, it directly increases the cost of moving, a key factor in the economic model of migration too. Second, it hinders maintaining contact with friends and family in the origin, hence imposing also a 'psychological' cost on migration. Third, the larger the spatial distance, the lower the information flow between the origin and the destination. A lower information flow means that it is more difficult to hear about available opportunities at the destination. The combined effect of population and spatial distance on migration is sometimes referred to as the 'gravity law' whereby the 'mass' (i.e. population) of the origin and the destination relative to the spatial distance between the two determines flows (Barthélemy, 2011; Levy, 2010).

The 'gravity law' focuses exclusively on spatial distance but it is known that cultural or linguistic distances constrain migration too. The similarity between the origin and the destination concerning language and religion, for example, improves the labour market integration of migrants (Van Tubergen et al., 2004; Windzio, 2018). Furthermore, cultural similarity facilitates migration due to homophily (McPherson et al., 2001), and by reducing discrimination and acculturation costs for migrants (Van Tubergen et al., 2004).

Finally, social networks are highly consequential for migration (Massey and España, 1987). This is true for the social network of the individual migrant (Massey and España, 1987) as well as macro-level network features of the origin and the destination (Windzio, 2018; Levy, 2010). Knowing people who migrated previously reduces the cost of migration for the prospective migrant. The social network in the destination can introduce the prospective migrant to possible employers, suggest places to live, offer information on vacancies, and provide social support which reduces the psychological costs of migration (Massey and España, 1987). These effects make migration self-perpetuating: the more migration there is from i to j, the more migration is expected to take place from i to j in the future (Massey, 1990).

One would expect the reverse to be true as well: the more migration there is from i to j, the more migration is expected to take place from j to i too. In other words, one expects a *reciprocity* effect. This is firstly due to return migration (Borjas and Bratsberg, 1996; Danchev and Porter, 2018). But once a link is established between i and j through large-scale migration, natives in j will be introduced to the opportunities and the networks in i as well. This may facilitate migration from j to i even among the natives of j.

Betweenness centrality (Freeman, 1977) and in-degree assortativity (Newman, 2002) are other network characteristics that are expected to affect migration. The betweenness centrality of a node in a network indicates how many (weighted) shortest paths pass through the node. A high level of betweenness centrality of an area implies that the area acts as a migration hub. This also means that there is high diversity, for people from different origins come to the area and likewise go to different destinations. This may offer new economic opportunities that attract further migration. Indeed, Damelang and Haas (2012) show using an instrumental variable design that in Germany cultural diversity enhances the labour market success of immigrants.

Finally, if two provinces i and j have attracted migration from similar other provinces, as measured by in-degree assortativity, one expects more migration to take place between j and k too. This is because in that case both j and k will include networks of people from similar origins. This will, in turn, facilitate information flow and reduce the costs of migration between j and k.

The aforementioned factors are expected to affect migration generically, international as well as intra-national. Below we discuss what we know about internal migration in Turkey.

3.1 Internal migration in Turkey

Historically, Turkey has sent a large number of emigrants to Europe. Most research has thus focused on Turkish immigrants' integration into Europe. Since the 1980s, however, emigration from Turkey to Europe has decreased (İçduygu and Sert, 2009). Far larger shares of the population migrate internally. Our data show that annually around 2.5 million people migrate between the 81 provinces of Turkey. Annual emigration abroad, on the other hand, is around 250 thousand.

Work on the determinants of internal migration in Turkey is scarce (Koramaz and Dökmeci, 2017b). Most of the existing research is carried out by urban planners whose main focus is on spatial issues (Filiztekin and Gökhan, 2008). Also, almost all existing studies analyse data up to the year 2000, i.e. the last year a population census was conducted. Studies that make use of more recent migration data are mainly descriptive; for example, Akın and Dökmeci (2015) classify the 12 regions of Turkey based on inter-regional migration patterns.

Gedik (1997) shows that starting from the 1980s the vast majority of migration in Turkey takes place from city to city as opposed to rural to urban. This pattern is also confirmed by Evcil et al. (2006). As expected, economic factors such as per capita Gross Domestic Product (GDP), wages, industrial workforce, and unemployment rates are found to be important determinants of migration (Filiztekin and Gökhan, 2008; Gezici and Keskin, 2005). Evcil et al. (2006) argue that GDP differentials are one of the most important drivers of internal migration.

In line with the 'gravity law', populations in the origin and destination are found to be positively associated with migration flows (Filiztekin and Gökhan, 2008; Gezici and Keskin, 2005). The findings on the second component of the gravity law, namely spatial distance, are equivocal. Gedik (1997) argues that beyond the immediate neighbouring cities, the effect of spatial distance on migration is minimal. Koramaz and Dökmeci (2017a), however, report that after peaking at a distance of 200-400 kilometres, migration decreases rapidly as spatial distance increases. Koramaz and Dökmeci (2017a) also report that in provinces in the East, most of which have large shares of Kurds, the effect of spatial distance is less pronounced.

Earlier studies also indicate the importance of social networks. Gedik (1997) shows that previous migrants who are friends or relatives from the same area are as effective forces as economic factors in facilitating migration. Filiztekin and Gökhan (2008) use the stock of earlier migration between i and j as an indicator of a network effect and register a strong effect of this earlier stock on future migration. To our knowledge, no study looks at further social network characteristics of Turkey's provinces, such as betweenness centrality, reciprocity, and assortativity.

Other important social forces may affect internal migration in Turkey. The first is politics. Aksoy and Billari (2018) and Aksoy and Gambetta (2021) show that provinces and districts that are ruled by Erdoğan's AKP (Justice and Development Party) are more efficient in providing local services and social assistance than those controlled by the opposition. This can potentially lead to 'welfare migration' from opposition to AKP municipalities. Aksoy and Billari (2018) do not find evidence for such welfare migration, but migration is not their focus and hence their analysis is rather descriptive. Political polarization is also rife in Turkey. A recent poll shows that a vast majority of participants (78%) do not "approve of their daughter marrying a supporter" of a party other than their own (Erdoğan and Semerci, 2017). We are not aware of any previous work that systematically focuses on the effect of politics and political distance on migration in Turkey. Yet, we expect Turkey's political divides to play an important role in migration decisions.

A further potentially important factor is ethnicity. There are sizeable Kurdish populations living in the large cities of Turkey. This is due to both economic as well as forced migration. Violent conflicts in the 1990s resulted in the forced displacement of Kurds (Ergin, 2014). Kurds tend to have higher levels of unemployment, poverty, and fertility (Koc et al., 2008) which are factors that traditionally push people to migrate. Due to historical and political reasons, data on ethnicity in Turkey are scarce, so much so that no population census since 1965 includes a question on ethnicity (Koc et al., 2008). Hence, it is unknown if Kurdish migration is particularly high, once economic factors and population are accounted for. In this study, we will provide the first comprehensive test.

4 Data and descriptive results

4.1 Data and variables

We compile a new dataset for this study as described below. All variables are at the province level and there are 81 provinces. All variables are obtained from the Turkish Statistical Institute (TurkStat), except the proportion of Kurds in a province, which is estimated from two representative surveys.

4.1.1 Dependent variable: migration counts

The number of individuals who moved from one province to another in a given year is available from TurkStat from 2009. These data come from Turkey's *Address Based Population Registration System* (ABPRS) which replaced the census in 2007. By law, each Turkish citizen is required to be registered at a single primary address. Any change in one's primary address is required to be updated in ABPRS within 20 working days, online or in-person. Not complying results in penalties. There are other incentives to keep the primary address up-to-date, for example, school access is determined by residence and any official communication takes place via the registered address. Migration that is not yet registered in the system is not captured by these data. Hence, our dependent variable should be interpreted as 'official' internal migration.

We also calculate the number of people who did not migrate. We do so by subtracting net migration (out- minus in-migration) from the population of a province in a year. We also apply a correction by adjusting for the number of births, deaths, or those who register for the first time in the province in a given year (see section 2 above and the part leading up to equation (1) there).

This gives us an 81×81 matrix for each year from 2009 to 2018, with off-diagonals indicating the number of people migrating between provinces, and the diagonals indicating those who stay.

4.1.2 Explanatory variables

We predict migration with the following explanatory variables. In our model, all dynamic (i.e. time-varying) explanatory variables are lagged by one year.

As indicators of economic prospects, we obtain annual provincial *GDP* per capita and the *unemployment rate*. Unemployment rates are only available at the Nuts-2 level which comprises 26 large regions.

As the elements of the 'gravity law', we use annual *population* in and the *spatial distance* between provinces. Spatial distance is measured as the driving distance in kilometres between province centres.

To measure the '*political distance*' between provinces, we calculate the absolute difference in vote shares of the political parties in the 2004, 2009, and 2014 mayoral elections. Political distance is a dynamic variable.

We calculate the following network characteristics using the (one-year lagged) migration matrix. The self-perpetuating nature of migration is captured by the number of people migrated from i to j in the *previous year* and by total in-migration in j, i.e. *popularity* (or equivalently weighted in-degree). *Reciprocity* for migration from i to j is calculated as the number of migrants from j to i in the previous year. *Betweenness centrality* of a province is calculated as the normalised number of geodesics (shortest paths) going through a province in the migration network. In-degree assortativity is operationalised as the correlation between the migration in-flows of i and j. The higher this correlation between two provinces, the more similar the provinces are in terms of the origin and flows of incoming migrants, so this variable can be interpreted as in-flow *similarity* too.

We also obtain the proportion of municipalities in a province that is controlled by the *AKP* after the 2004, 2009, and 2014 elections as a dynamic variable. This variable has a [0 1] range corresponding to the provinces in which AKP controls none and all of the municipalities in a province in a given year.

Finally, we estimate the percentage of *Kurds* in a province as a dynamic variable. There is no official data on ethnicity, so we use Turkey's Demographic and Health Survey 2008 and 2013 waves (Koc et al., 2008). Both surveys are representative and sample women of reproductive age (15 to 49) (Hacettepe University Institute of Population Studies, 2013). The survey includes a question about the respondent's mother tongue using which we identify if someone is Kurdish. We use the 2008 and 2013 waves to estimate the proportion of Kurds in a province in and before those years.

To facilitate estimation and interpretation, we normalise all explanatory variables, except time, to the [-1 1] range, by centring around the mean and dividing by the maximum of the absolute of the centred values, hence one of the $-1, 1$ boundaries is attained. In this way, the coefficients of those factors will be about a change associated with a change from the average value to the absolute maximum observed value of a factor. Finally, we worked with *popularity*, *population*, *previous year*, and *reciprocity* in the log domain.

4.2 Descriptive results for migration

Figure 2 shows out- and in-migration in the 81 provinces of Turkey. The figure shows the absolute level of migration, migration as the percentage of the number of inhabitants at the start of the year, and the 2018-2009 difference in the latter. The figure shows that large provinces, such as İstanbul, Ankara, İzmir, Adana, and Antalya have high levels of out- and in-migration (see Figure 3 for province names). This is in line with gravity law. There is almost a perfect correlation between out- and in-migration levels across provinces.

Migration as a percentage of the number of inhabitants, however, shows a different pattern. Proportional to their populations, large provinces (e.g. İstanbul, Ankara, İzmir, Adana, and Antalya) have relatively low migration. Smaller provinces in the centre-east, such as Gümüşhane and Tunceli have the highest migration per capita. The difference between absolute and relative migration will become important when we present our model estimates below.

The changes displayed in Figure 2 (bottom panels) show that, on average, the change in out-migration per capita is rather stable. However, it is slightly negative in some places and slightly positive in others, the largest change is in Gümüşhane.

Figure 2 shows only the origins and destinations of migration. Figure 3 shows absolute migration flows. In Figure 3 we omit small flows for clarity. These small flows will be part of the analyses in Section 5. Figure 3 shows that İzmir, Anlatya, Tekirdağ, Ordu, Tokat, and Ankara are popular destinations, especially for those from İstanbul. There are also local clusters, such as Şanlıurfa, Konya, Gaziantep, Adana and Mersin that attract and send migrants.

Finally, Figure 4 shows the bi-variate correlations for the one-way (i.e. variables defined for a province–'node' characteristics) and two-way factors (i.e. variables defined for a pair of provinces–'edge' characteristics). As expected, many predictors are correlated. Below we will analyse how these factors are associated with migration flows, our dependent variable, after describing our model of migration in detail.

5 Results

In this section, we provide the results of the implementation of our methodology using the migration data described in Section 4. The experiments were performed in MATLAB, version

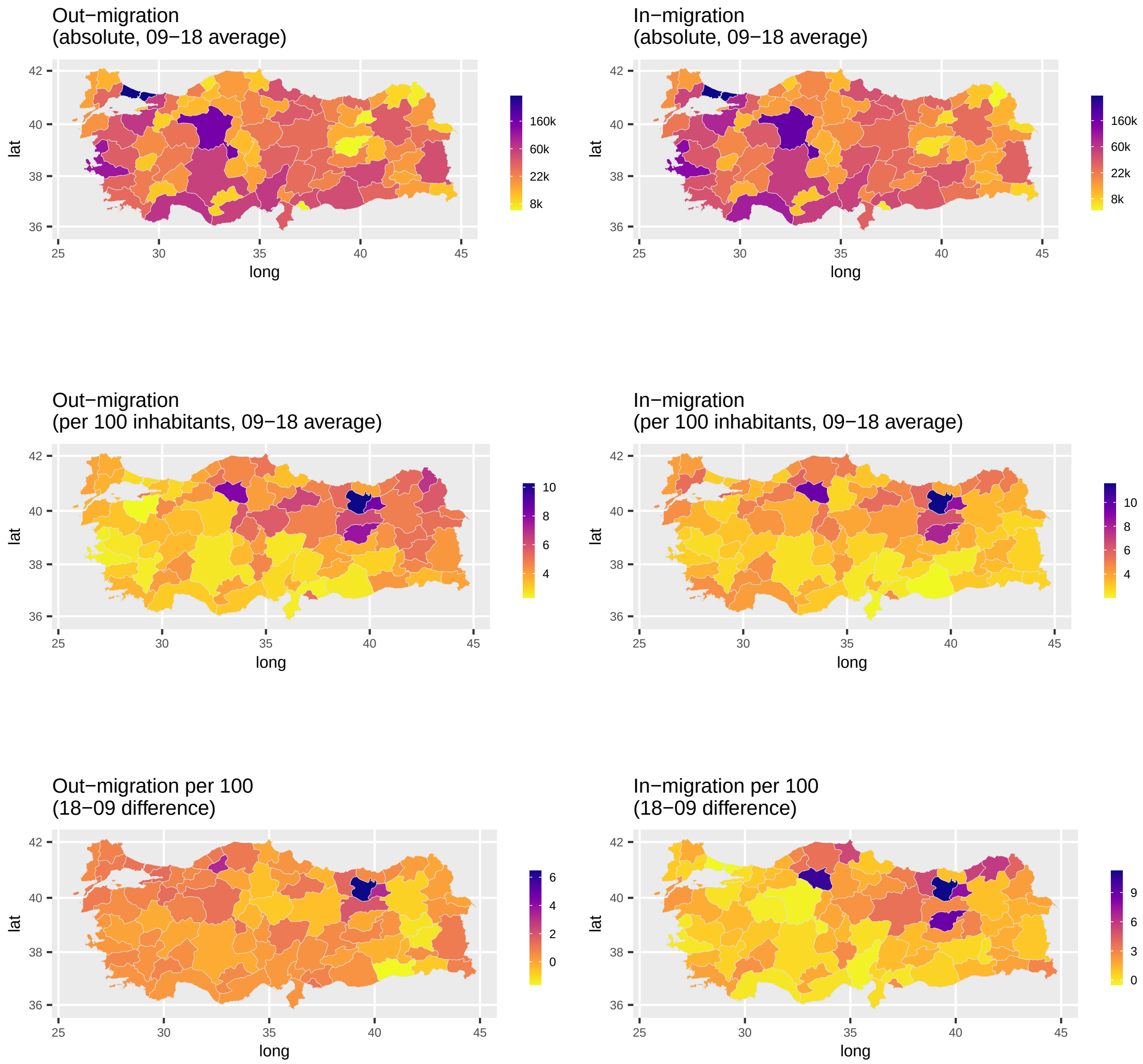

Figure 2: Absolute, per 100, and change in per 100 between 2018 and 2009 in out- and in-migration per province.

R2021b. All data and the code that produce the results below are available (anonymised for peer-review) at https://github.com/SocNetMigration/MigNet-MATLAB-code.git. This replication package also includes guidelines for preparing other data for our model.

5.1 Model estimates

We implemented the sampling method in Algorithm 1 to estimate θ. We use non-informative priors for $\boldsymbol{\theta}$ described in Appendix A. Algorithm 1 is run for 5×10^5 iterations, and the posterior distributions are estimated based on the draws obtained during the last 80% of the iterations (following a burn-in). Convergence is confirmed by visual inspection of the history of the draws

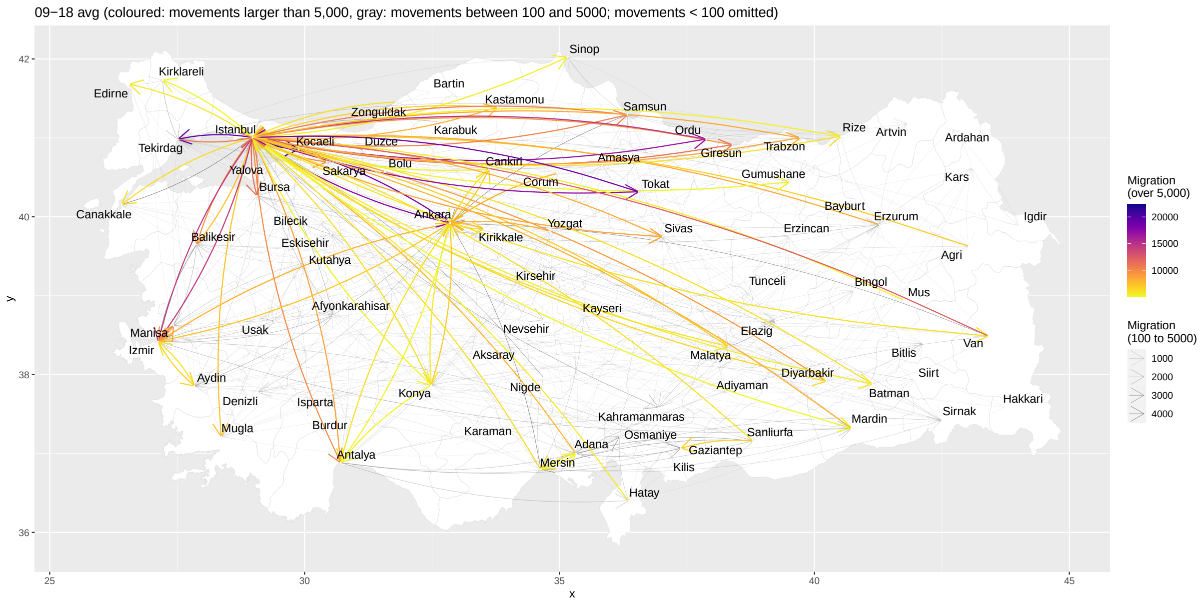

Figure 3: Migration flows between provinces.

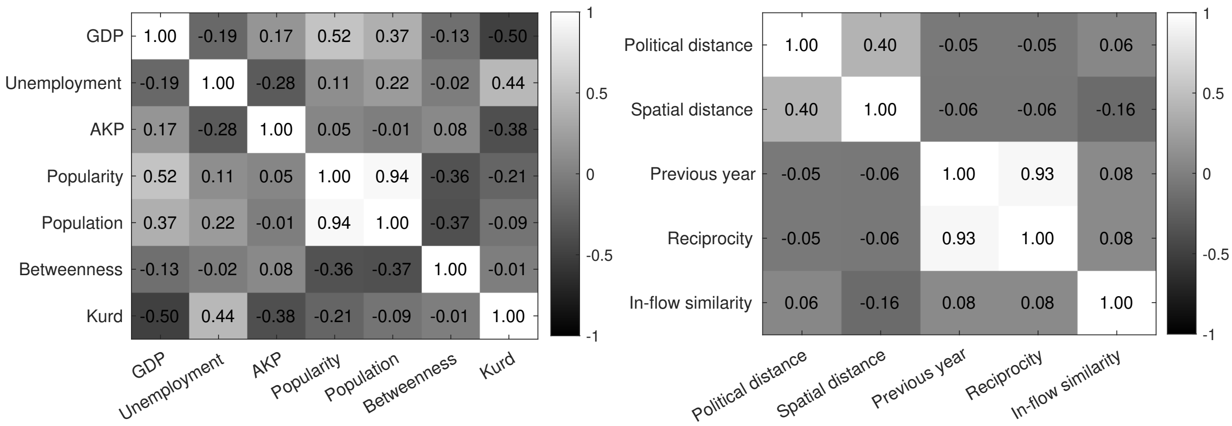

Figure 4: Sample correlation matrix for the one-way and two-way factors

from the chain.

Figure 5 shows the box plots for the estimated marginal posterior distributions of the coefficients for our predictors of migration. Table 1 displays the means, standard deviations, and 90% credible intervals of those parameters, and additionally of the random intercept and the 'scale' parameter of the Dirichlet-multinomial model (see (2) and (4)). Furthermore, the correlation structure in the posterior distribution for $\theta_1, \theta_2, \theta_3$ is shown in Figure 6. Appendix B.1 includes the histograms of the posterior distributions of the parameters.

The correlation structure in the posterior in Figure 6 is in line with the correlation matrix of the predictors given in Figure 4. A comparison of these two figures shows the typical trend that a positive correlation between two predictors yields a negative correlation between their coefficients in the posterior, and vice versa, as expected. This justifies the general cautionary rule that applies to almost all multivariate models that the effects of correlated predictors should not be considered in isolation from each other.

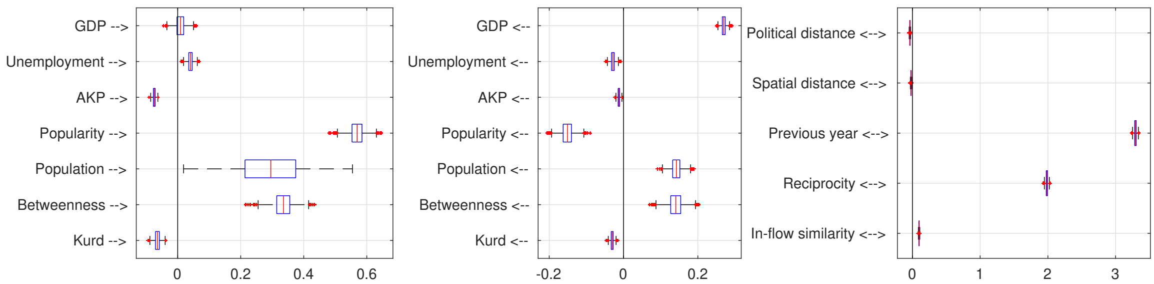

Figure 5: Box plots of marginal posterior distributions of factor coefficients in the hierarchical (random intercepts across provinces) Dirichlet-multinomial model described. See Table 1 for the posterior distributions of the intercepts.

Table 1: Means, standard deviations, and 90% credible intervals for the marginal posterior distributions of the parameters of the Dirichlet-multinomial model with the hierarchical specification.

Parameter	**mean**	**std**	90% **Cred. Int.**	**Parameter**	**mean**	**std**	90% **Cred. Int.**
GDP $\rightarrow$	0.0088	0.0155	(-0.0175 0.0335)	GDP $\leftarrow$	0.2698	0.0060	(0.2601 0.2796)
Unemp. $\rightarrow$	0.0406	0.0079	(0.0279 0.0533)	Unemp. $\leftarrow$	-0.0286	0.0055	(-0.0372 -0.0192)
AKP $\rightarrow$	-0.0742	0.0044	(-0.0816 -0.0671)	AKP $\leftarrow$	-0.0124	0.0028	(-0.0167 -0.0078)
Popularity $\rightarrow$	0.5691	0.0237	(0.5310 0.6078)	Popularity $\leftarrow$	-0.1511	0.0164	(-0.2064 -0.1538)
Population $\rightarrow$	0.2935	0.1187	(0.0899 0.4920)	Population $\leftarrow$	0.1419	0.0141	(0.1177 0.1642)
Between. $\rightarrow$	0.3358	0.0334	(0.2876 0.3892)	Between. $\leftarrow$	0.1408	0.0199	(0.1083 0.1727)
Kurd $\rightarrow$	-0.0642	0.0090	(-0.0785 -0.0487)	Kurd $\leftarrow$	-0.0305	0.0039	(-0.0370 -0.0240)
Polit. dist. $\leftrightarrow$	-0.0376	0.0040	(-0.0441 -0.0310)	base mean	-8.4527	0.1286	(-8.6645 -8.2419)
Spat. dist. $\leftrightarrow$	-0.0213	0.0037	(-0.0275 -0.0152)	base var.	1.3366	0.2274	(1.0079 1.7467)
Prev. year $\leftrightarrow$	3.2927	0.0153	(3.2675 3.3177)	Scale	10.5472	0.0064	(10.5368 10.5576)
Reciprocity $\leftrightarrow$	1.9850	0.0140	(1.9616 2.0079)				
In-flow sim. $\leftrightarrow$	0.0982	0.0040	(0.0916 0.1048)				

The results are also mostly in line with theory. Unemployment ($\rightarrow$) in the sender province is positively associated with out-migration and unemployment ($\leftarrow$) in the target has a negative association with in-migration. While GDP ($\rightarrow$) in the sending province does not seem to have a strong association with out-migration, GDP ($\leftarrow$) in the target is strongly and positively associated with in-migration.

Regarding the elements of the 'gravity law', a large population ($\rightarrow$) is associated with a higher probability of out-migration–the posterior mean of its coefficient is 0.29. Population ($\leftarrow$) in the target also has a very strong positive association with in-migration. We also find a negative association between spatial distance ($\leftrightarrow$) and migration. These results are in line with the 'gravity law'.

We also find a negative association between provincial political distance ($\leftrightarrow$) and migration. Interestingly, the size of the coefficient for political distance is slightly higher than that for spatial distance.

As expected, network characteristics are strongly associated with migration. Migration from i to j in the previous year ($\leftrightarrow$) is a very strong predictor of current migration. Its coefficient is the largest of all. Interestingly, the popularity ($\leftarrow$) of a province (total in-migration) in a given year is associated negatively with in-migration the next year, net of all other factors controlled

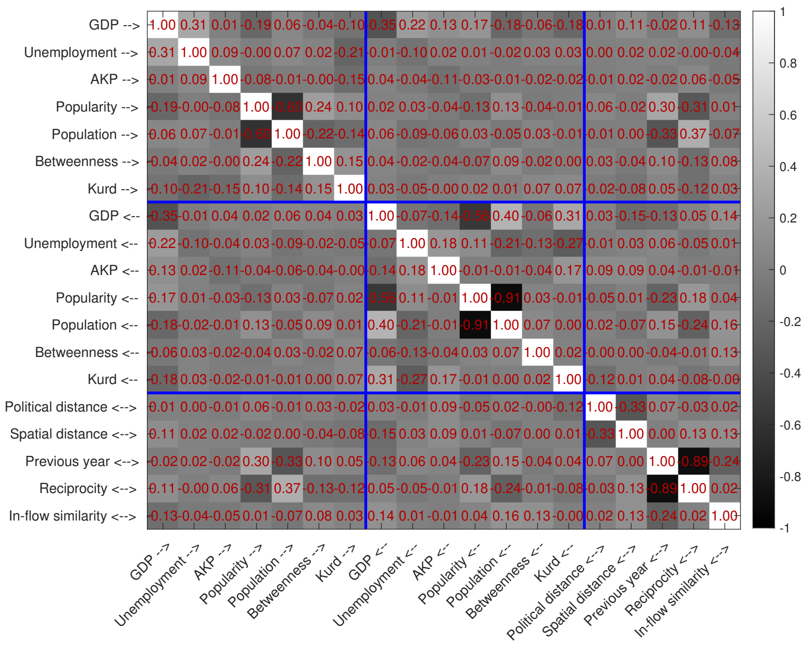

Figure 6: Posterior correlation matrix

in the model.

We also find strong reciprocity ($\leftrightarrow$): the larger the migration from j to i in a given year, the larger the migration from i to j the next year. Betweenness centrality of a province in a given year also makes a province attractive the next year ($\leftarrow$). Likewise, betweenness centrality ($\rightarrow$) of the origin is associated strongly and positively with out-migration. Hence, central provinces seem to act as hubs, attracting and initiating further migration. In-flow similarity ($\leftrightarrow$), which captures in-degree assortativity, is associated positively with migration too.

Interestingly, we find that the strength of the AKP in a province is negatively associated with both out-migration ($\rightarrow$) and in-migration ($\leftarrow$), although its coefficient is small in both cases. This shows that migration both out of and into AKP-dominated regions is low. Together with the strong negative coefficient for political distance, this finding further indicates the negative association of political divides with migration.

Finally, the proportion of Kurds in a province is associated negatively with both out-($\rightarrow$) and in-migration ($\leftarrow$). Although its coefficients are low in the absolute sense, negative values are inconsistent with the common belief that migration is a Kurdish phenomenon. Note, however, that the outcome is all migration from and to a province, Kurdish or otherwise. Hence, the evidence here is indirect. Nevertheless, if Kurds were much more mobile than Turks and other ethnicities as the common belief suggests, one would expect a positive effect of the proportion of Kurds in a province on out-migration.

The coefficients for $\leftarrow$ and $\leftrightarrow$ factors can be interpreted in terms of relative probabilities of attracting migration from a given province. For example, suppose that the maximum observed GDP value is normalised to 1 (recall that all factors are centred around the mean and normalised to the $[-1, 1]$ range). Then the mean of the posterior of the coefficient for "GDP $\leftarrow$" suggests that, from a given province, the probability of migration to a province with the highest GDP per capita is 31% higher ($\exp(0.27) = 1.31$) than the probability of migration to an average province *ceteris paribus*. Likewise, the probability of migration to the spatially farthest province is 2% lower ($\exp(-0.02) = 0.98$) than that to an average province (provided that the maximum spatial distance has been normalised to 1). The coefficients for $\rightarrow$ factors can be interpreted in terms of the relative probability of migration to any province versus not migration. For example, one unit increase in unemployment (e.g. the change from average to maximum observed value) in province i is associated with a 4% ($\exp(0.04) = 1.04$) increase in relative migration probability from i, *ceteris paribus*.

Table 1 also shows the mean and the variance of the intercept which is random across provinces. They show that there is considerable variability in the levels of out-migration probabilities across provinces, the largest in Bayburt and the smallest in Istanbul. Figure 7 shows the posterior distribution of the intercepts per province, estimated with our hierarchical specification. These provincial variations somewhat match the descriptive statistics given in Figure 2, although they are not equal because these intercepts are obtained after controlling for the predictors of migration.

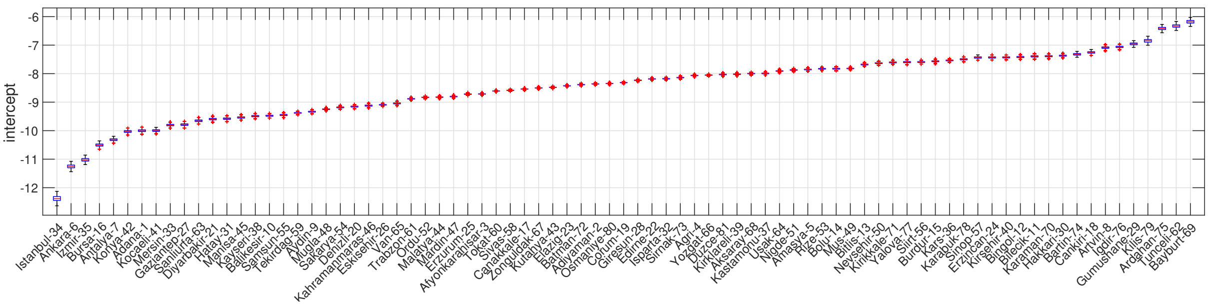

Figure 7: Box plots of the marginal posterior distributions of random intercepts for provinces in the hierarchical specification of the Dirichlet-multinomial model.

Additional estimation results: Some scholars argue that controlling for the lagged dependent variable accounts for path dependency or autocorrelation of residuals in panel regressions, though it may also bias downwards the coefficients for other explanatory variables (Keele and Kelly, 2006). This potential issue has been shown to apply to temporal ERGMs, too (Block et al., 2018). To address this potential issue, we fitted our models after excluding variables that are calculated from past migrations. These variables are popularity, betweenness, previous year, reciprocity, and in-flow similarity. The results after excluding these variables are given in Appendix B.2.

These additional results show that including the measures based on migration in the previous year in the model generally does not suppress the coefficients of the other variables. In fact, for some variables, the estimated coefficients are somewhat smaller after excluding the lagged measures (e.g. AKP $\leftarrow$). Coefficients do increase, in absolute value, for some variables (e.g. GDP $\rightarrow$, Kurd $\leftarrow$) after excluding lagged variables. But because of a lack of general suppression of coefficients due to including in the model measures based on lagged values of migration, we

Table 2: Mean squared prediction errors for the Dirichlet-Multinomial model and the gravity model with random intercepts for provinces (DM and Grav-re, respectively) and with a global single intercept (DM0 and Grav0) for different outcomes (log counts or counts, diagonals excluded or included).

Outcome	Diagonals	DM	Grav-re	DM0	Grav0
log-counts	excluded	0.1273	0.1227	0.1327	0.1242
	included	0.1257	0.1212	0.1311	0.1227
counts ($\times 10^6$)	excluded	0.0927	0.1054	0.1096	0.1165
	included	1.0847	1.5478	1.8661	2.1104

conclude that the above issue is not particularly problematic in our case. Note that one expects differences in the coefficients of variables that are estimated before and after the inclusion of lagged migration variables due to the conditional nature of the Dirichlet-multinomial regression, which we observe too. Additionally, as shown in Table 3 in Appendix B.3, the exclusion of lagged measures reduces the predictive performance substantially, further suggesting the need for including them in the model.

5.2 Comparison with the gravity model in terms of predictive power

The gravity model assumes a linear regression for the non-diagonal log-migration counts. Although the essential factors of the gravity model are population and distance, one could easily include other one-way and two-way factors in the model. Therefore, for fairness in comparison, we constructed a gravity model that includes all the factors considered in our model. This corresponds to the following relation for $i, j \in \{1, \dots, n\}$; $i \neq j$, and $t \in \{1, \dots, T\}$,

$$\log H_t(i,j) = \theta_0(i) + \theta_1 \cdot u_t(i) + \theta_2 \cdot v_t(j) + \theta_3 \cdot z_t(i,j) + e_t(i,j),$$

where the sum on the right-hand side serves as the base parameter as before, $e_t(i,j)$ are uncorrelated deviation terms, assumed to have zero mean and common variance, and $u_t(i)$, $v_t(j)$, and $z_t(i,j)$ are as defined before. (Wherever $H_t(i,j) = 0$, which happens only in 31 out of $81 \times 81 \times 81 = 65610$ cells, we substituted it with a 1 to avoid $\log 0$.) We fitted two versions of this model: first, with a global $\theta_0(i) = \theta_0$ which is constant across provinces, and the second in which $\theta_0(i)$'s are "random effects" which are assumed to be normally distributed across provinces. The latter version matches with the Dirichlet-Multinomial model we discuss above which also has "random effects" for provinces. To facilitate comparison we also fit a Dirichlet-Multinomial specification with a global θ_0.

We compared the gravity model and the Dirichlet-Multinomial model in terms of their out-of-sample predictive performances with the following process. Recall that the total data consists of 10 consecutive years between 2009 and 2018. For every single year, we excluded its migration data and used the rest of the years for training the model. The training results are then used to predict the migrations in the excluded year. For the Dirichlet-Multinomial model, we ran Algorithm 1 (or variants of this for the global and random effects intercepts) to obtain samples from the posterior distribution of θ given the training data, and we predicted the migration counts in the test year by the Monte Carlo estimation of their expectations with respect to their posterior predictive probability distribution. For the gravity model, the maximum likelihood estimates for the parameter vector, obtained from the training data, were substituted for calculating the expectation of $\log H_t(i,j)$ for the test year t. Moreover, for the gravity model, for a test year t, the non-migration counts, i.e., $H_t(i,i)$, $i = 1, \ldots, n$ are predicted by $P_t(i) - \sum_{j=1, j \neq i}^{n} \hat{H}_t(i,j)$, where $\hat{H}_t(i,j)$ is the predicted value for $H_t(i,j)$. The prediction results for the ten test years are then used to calculate squared errors [(predicted - observed)^2] and averaged over the 10 years. We calculate separate mean squared errors for migration counts with and without the diagonals. This is because in a typical gravity model diagonals are excluded.

The results show that the Dirichlet-Multinomial model has better performance than the gravity model in predicting migration counts. Prediction errors for the Dirichlet-Multinomial for count outcomes are consistently lower than that for the gravity model both when we include and exclude the diagonal (non-migration) values in the predicted outcome. When the log-counts are considered the gravity model performs slightly better than the Dirichlet-Multimonial. This suggests that the gravity model is generally mispredicting large flows. As an aside, the mean absolute error of our hierarchical model for all counts including non-migration is only 108.6 which corresponds to less than 0.05% of the standard deviation of migration counts. This shows that our model predicts migration flows reasonably well. Also note that random province-specific intercept parameters improve the performance, and hence those random effects do not lead to over-fitting.

We also compared the predictive performance of our method with that of a simpler variant which excludes variables based on the previous year's migration flows (see Table 3 in Appendix B.3). The results show that the exclusion of lagged measures hampers predictive performance substantially, pointing to the need of including them in the model as we do.

5.2.1 Baseline probability of migration and the predictive accuracy of the gravity model

The results above show that our model outperforms the gravity model in predicting migration counts. We conjecture that one of the main reasons for the relatively poor performance of the gravity model in predicting migration counts is the following. The gravity model fails to capture the mechanistic relationship between migration out-flows to different targets from the same origin, for example, due to competition between different targets for a constant number of possible migrants from a given origin. This drawback increases as the share of a population that migrates relative to those that do not migrate increases. This is because the smaller the share of the people in an origin that do not migrate (hence the larger the share of those who migrate), the larger the (negative) correlation between migration flows to different destinations, and since the gravity model does not capture those correlations naturally, the lower should be the predictive accuracy of the gravity model.

To demonstrate the mechanism explained above and help understand the predictive performance of the gravity model vis-a-vis ours better, we carry out a Monte Carlo simulation, where we test the predictive performances of the Dirichlet-multinomial model and the gravity model on simulated data sets. In our simulation, we randomly generate migration flows using two separate processes. In the *first* process, we generate migration flows from the Dirichlet-multinomial model with 10 origins to those 10 destinations for 5 time periods. In generating those flows, we randomly simulate the origin populations (from the Poisson distribution with a mean of 1000, independently for each origin) and 3 $\rightarrow$, 3 $\leftarrow$ and 2 $\leftrightarrow$ factors, and draw migration counts from the respective Dirichlet-multinomial distributions as in (3). We fix $\theta_1 = (1,1,1), \theta_2 = (1,1,1)$, and $\theta_3 = (1,1)$ and control the baseline out-migration probability by varying a common θ_0 parameter for all origins. (For simplicity, we do not add sender random effects in the data generation process.) We vary θ_0 from -5 to 0 with increments of 0.5 and generate 10 datasets for each unique θ_0 value. Note that the parameter θ_0 controls the level of negative correlation between out-migration counts. As θ_0 increases, the negative correlation between the out-migration counts increases.

Note that in the simulation above our aim is to understand better how the predictive accuracy of the gravity mode depends on the baseline probability of out-migration, rather than comparing our model with the gravity model in all possible data generation scenarios. Nevertheless, because we generate the migration counts from a Dirichlet-multinomial model, one may argue that the data generation process does not do justice to the gravity model. Hence, in a *second* process we generate migration counts from a multivariate normal distribution obtained by randomly 'perturbing' the Dirichlet-multinomial model. Specifically, for every t and i, let the mean vector and covariance matrix of Dirichlet-Multinomial$(X_t(i); \alpha_t(i,1), \dots, \alpha_t(i,n))$ be $m_{t,i}$ and $S_{t,i}$. We sample the counts as

$$(H_t(i,1), \dots, H_t(i,n)) \sim \mathcal{N}(m_{t,i}, S_{t,i}), \quad \text{where} \quad S_{t,i} \sim \text{Wishart}(S_{t,i}/n, n), \tag{7}$$

that is, we distort the covariance matrix to deviate from the Dirichlet-multinomial model while still keeping the negative correlation among the counts. The counts drawn from the multivariate normal distribution are then rounded to the nearest integers and truncated at the zero lower bound. Finally, we readjust the population according to the final values of the counts. As in the first process, in this second process, too, θ_0 controls the level of negative correlation between out-migration counts.

For both processes, we then calculate mean squared errors as we did above, that is by leaving out one time period in training the gravity and the Dirichlet-multinomial model and predicting flows in the left-out period. For the gravity model, we also add the population as a factor.

Figure 8 shows the results of our Monte Carlo simulations. The figure confirms our conjecture as to the reasons for the gravity model's relatively poor predictive performance. As the baseline probability of migration increases (corresponding to larger θ_0), the predictive power of the gravity model vis-a-vis the Dirichlet-multinomial model decreases, for both of the two data generation processes we described above. These results imply that as the ratio of migrants relative to non-migrants from a given province approaches zero, the predictive accuracy of the Dirichlet-Multinomial and the gravity model converge. This is an important observation that we will return to in the final section, for the predictive drawbacks of the gravity model may be less problematic if, in a particular application, the migration counts (flows) relative to non-migrant counts (stocks) are low.

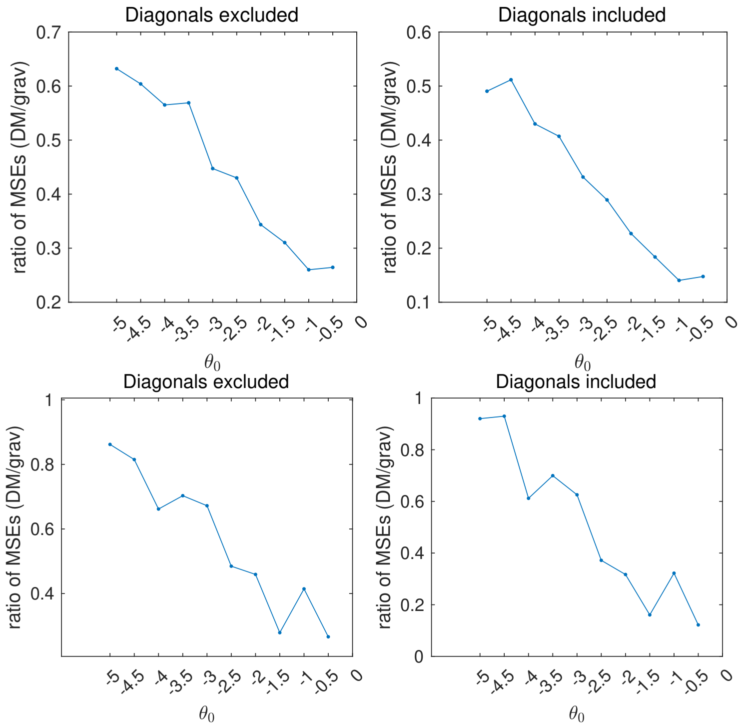

Figure 8: Predictive performance of the gravity model relative to the Dirichlet-multinomial model on simulated data. **Top row:** Data are simulated from the Dirichlet-multinomial model. **Bottom row:** Data are simulated from the normal distributions, as in (7). The parameter θ_0 captures the baseline probability of migration and as this probability increases so does the negative correlation between migration to different destinations, and the predictions of the gravity model worsen.

6 Discussion and conclusions

In this study, we propose a Dirichlet-multinomial model and a Bayesian inference method to analyse the interdependent flows. We apply the developed methodology to analyse the dynamic migration flows of the 81 provinces of Turkey from 2009 to 2018. Our study offers methodological and substantive contributions to the literature.

On the methodological front, our Dirichlet-multinomial model alleviates several shortcomings of the popular models applied to migration flows. The advantages of our model are as follows. *First*, it naturally captures systemic effects. That is, our model incorporates, unlike the gravity model, the effects of the characteristics of all alternative destinations on a decision to migrate to a specific destination. *Second*, in our model, non-migration is accounted for. The alternative to ignoring non-migration and conditioning the results on migration having taken place, which is a common practice in the literature, results in the so-called selection on the dependent variable. *Third*, our model adheres to the natural boundary of migration and the mechanistic relationship between migration and non-migration. That is, out-migration and in-migration change the populations of the origin and the destination, and the total number of out-migration cannot exceed the population of the origin in a given period. Our model naturally incorporates these phenomena. *Fourth*, our model lends itself to an estimation method that is computationally much simpler than many of its alternatives. The computational simplicity allows us to expand our model with a hierarchical specification that captures variations in the baseline of and potentially also longitudinal changes in migration probabilities.

To demonstrate those advantages, we carried out an out-of-sample prediction comparison of our model vis-a-vis the *gravity model*, one of the most common models used in the migration literature. The gravity model ignores systemic effects, and it typically ignores non-migration as well as the mechanistic relationship between populations, non-migration, and migration flows. We showed that our model consistently outperforms the gravity model in predicting migration flow counts.

We also showed through Monte Carlo simulations that the predictive accuracy of the gravity model worsens as the ratio of migrants to non-migrants increases. Conversely, as the ratio of migrants to non-migrants approaches zero, the predictive accuracy of the gravity model converges with that of our model. This implies that when migration is a much rarer event compared to non-migration, the gravity model would perform reasonably well.

Note that while we framed our model within a migration setting, and indeed we developed it to understand migration flows and their determinants, the model is more general. It can be applied to any other setting that involves discrete dynamic flows between a finite number of origins and destinations.

We should note that there are several models developed in the social networks literature, that are very general and flexible and, in principle, can also be applied to dynamic flows with some adjustments. These include Exponential Random Graph Models and their various extensions for count data or longitudinal data; relational event models, dynamic network actor models, stochastic actor-oriented models, and latent space and latent factor models. We discussed those models in detail and compared them with ours. While they are indeed very powerful and flexible, we argued that they may be too general for our problem at hand. Our method provides a straightforward approach to model dynamic flows while offering computational simplicity and at the same time taking into account key features of and dependencies in the data.

The computational simplicity of our model rests on the assumption that migrations from different origins at a given time point are independent, *conditional* on the covariates which can include those that are based on previous migration flows. Note that this conditional independence assumption is about decisions in different provinces, our model incorporates possible positive correlations between migration decisions within the same province. We believe that this assumption is not unrealistic in our longitudinal context. When an individual is deciding to migrate or not in a given time-point, they may not consider or indeed be aware of the migration decisions of others outside their provinces at the same time point. Moreover, dependencies among migration probabilities that may violate this assumption can be explained away, to a certain degree, by including in the model the predictors that are the likely causes of that dependence. For example, if one expects a 'herding' dynamic for migration (i.e. people in different provinces imitating each other in migrating to a specific destination) this factor can be added to the model, by adding a destination *popularity* factor based on previous migration flows into the destination, as we do in our analysis. In other words, because we have longitudinal data, we can condition current flows on earlier flows, which makes the conditional independence assumption more plausible.

We also note that our model allows further extensions. One particular direction is the following. In our hierarchical specification, we include random intercepts for out-migration (i.e. for sending provinces). One can include such random effects for the receiving provinces, too. This would result in a *cross-classified* model (Snijders and Bosker, 2011, Chapter 13). In fact, we have fitted this cross-classified version. The parameter estimates as well as the predictive performance of our model hardly changed. At the same time, however, the receiver random effects increased the computational complexity of the model substantially. This is because, in a cross-classified model, each of the n random intercepts of the receiving provinces would affect the Dirichlet-multinomial probabilities for all provinces. Thus, the acceptance probability in the estimation which is needed to accept or reject a proposed update for each of those parameters would require the computation of the whole product in (6). Note that this is in stark contrast with the update of the intercept of the sending provinces, for which we calculate a single Dirichlet-multinomial probability for each t in the numerator and the denominator of the acceptance ratio, see Algorithm 1. Because receiver random effects were inconsequential to our results and the predictive performance of our model, we decided to proceed without the receiver random effects. Nevertheless, they could be incorporated should the researcher sees a need to do so.

Overall, we contribute to the literature methodologically by providing a viable alternative for analysing migration and other types of flows.

Our study also offers substantive contributions to the migration literature. Far greater shares of populations are affected by internal migration than by international migration. Most past research, however, focuses on the latter. Using a new dataset and a model, we analyse inter-provincial migration in Turkey. Our results largely confirm the existing theories of migration: economic prospects, population, spatial distance, and network characteristics shape the flow of internal migration in Turkey. We also offer several novel findings.

Firstly, we register that political distance between provinces (measured using electoral results) is negatively associated with migration flows. This negative association is even stronger than that between spatial distance and migration. Moreover, we find that the strength of the AKP (Justice and Development Party) in a province is negatively associated with out- and in-migration. These findings together imply that political divisions in the country contribute to the sorting of province populations as well. This association between politics and migration, in turn, may accelerate further the political polarisation between the provinces of Turkey which is already rife.

We also provide, to our knowledge, the first systematic test of whether the proportion of Kurds in a province is associated with migration. Kurds are affected by higher levels of unemployment and poverty and have higher fertility. These factors are traditionally associated with migration. Indeed there has been large-scale Kurdish migration to western parts of Turkey. Our analysis, however, shows that conditional on the factors we include in our model, the share of Kurds in a province has a small negative association with migration. While these findings should be interpreted with caution, as the independent variable is the share of Kurds in a province while the outcome is migration across all ethnic groups, a lack of positive coefficient suggests that internal migration in Turkey is not a predominantly 'Kurdish phenomenon'.

References

Abramski, K., Katenka, N., and Hutchison, M. (2020). A network-based analysis of international refugee migration patterns using gergms. In Cherifi, H., Gaito, S., Mendes, J. F., Moro, E., and Rocha, L. M., editors, *Complex Networks and Their Applications VIII*, pages 387–400, Cham. Springer International Publishing.

Akın, D. and Dökmeci, V. (2015). Cluster analysis of interregional migration in Turkey. *Journal of Urban Planning and Development*, 141(3):05014016.

Aksoy, O. and Billari, F. C. (2018). Political islam, marriage, and fertility: evidence from a natural experiment. *American Journal of Sociology*, 123(5):1296–1340.

Aksoy, O. and Gambetta, D. (2021). The Politics behind the Veil. *European Sociological Review*, 37(1):67–88.

Alamá-Sabater, L., Alguacil, M., and Bernat-Martí, J. S. (2017). New patterns in the locational choice of immigrants in spain. *European Planning Studies*, 25(10):1834–1855.

Almquist, Z. W. and Butts, C. T. (2013). Dynamic network logistic regression: A logistic choice analysis of inter- and intra-group blog citation dynamics in the 2004 us presidential election. *Political analysis*, 21(4):430–448.

Almquist, Z. W. and Butts, C. T. (2014). Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. *Sociological Methodology*, 44(1):273–321. PMID: 26120218.

Azose, J. J. and Raftery, A. E. (2015). Bayesian probabilistic projection of international migration. *Demography*, 52(5):1627–1650.

Azose, J. J. and Raftery, A. E. (2018). Estimating large correlation matrices for international migration. *The annals of applied statistics*, 12(2):940.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1-3):1–101.

Bell, M. and Muhidin, S. (2009). Cross-national comparison of internal migration. *United Nations Development Programme Human Development Reports Research Paper*.

Benson, A. R., Kumar, R., and Tomkins, A. (2016). On the relevance of irrelevant alternatives. In *Proceedings of the 25th International Conference on World Wide Web*, pages 963–973.

Block, P., Koskinen, J., Hollway, J., Steglich, C., and Stadtfeld, C. (2018). Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks*, 52:180–191.

Block, P., Stadtfeld, C., and Robins, G. (2022). A statistical model for the analysis of mobility tables as weighted networks with an application to faculty hiring networks. *Social Networks*, 68:264–278.

Block, P., Stadtfeld, C., and Snijders, T. A. B. (2019). Forms of dependence: Comparing saoms and ergms from basic principles. *Sociological Methods & Research*, 48(1):202–239.

Borjas, G. J. (1999). *Economic research on the determinants of immigration: Lessons for the European Union*. The World Bank.

Borjas, G. J. and Bratsberg, B. (1996). Who leaves? The outmigration of the foreign-born. *The Review of Economics and Statistics*, 79(1):165–176.

Borjas, G. J., Bronars, S. G., and Trejo, S. J. (1992). Self-selection and internal migration in the united states. *Journal of Urban Economics*, 32(2):159–185.

Boyle, P., Halfacree, K., and Robinson, V. (1998). *Exploring contemporary migration*. Longman, Harlow.

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, 38:155–200.

Butts, C. T., Acton, R. M., Hipp, J. R., and Nagle, N. N. (2012). Geographical variability and network structure. *Social Networks*, 34(1):82–100.

Chan, T. W., Henderson, M., Sironi, M., and Kawalerowicz, J. (2020). Understanding the social and cultural bases of brexit. Technical report, Department of Quantitative Social Science-UCL Institute of Education.

Cheng, S. and Long, J. S. (2007). Testing for iia in the multinomial logit model. *Sociological methods & research*, 35(4):583–600.

Çoban, C. (2013). Different periods of internal migration in Turkey from the perspective of development. *American International Journal of Contemporary Research*, 3(10):58–65.

Damelang, A. and Haas, A. (2012). The benefits of migration. *European Societies*, 14(3):362–392.

Danchev, V. and Porter, M. A. (2018). Neither global nor local: Heterogeneous connectivity in spatial network structures of world migration. *Social Networks*, 53:4–19.

Desmarais, B. A. and Cranmer, S. J. (2012). Statistical inference for valued-edge networks: The generalized exponential random graph model. *PloS One*, 7(1):1–12.

Erdoğan, E. and Semerci, P. (2017). Dimensions of polarization in Turkey. *Report by Istanbul Bilgi University Center for Migration Research*.

Ergin, M. (2014). The racialization of Kurdish identity in Turkey. *Ethnic and Racial Studies*, 37(2):322–341.

Evcil, A. N., Kiroplu, G. B., and Dokmeci, V. (2006). Regional migration in turkey: Its directions and determinants. In *46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September 3rd, 2006, Volos, Greece*, Louvain-la-Neuve. European Regional Science Association (ERSA).

Expert, P., Evans, T. S., Blondel, V. D., and Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences*, 108(19):7663–7668.

Filiztekin, A. and Gökhan, A. (2008). The determinants of internal migration in Turkey. In *International Conference on Policy Modelling (EcoMod2008), Berlin, Germany*, volume 24.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41.

Gedik, A. (1997). Internal migration in Turkey, 1965–1985: Test of conflicting findings in the literature. *Review of Urban & Regional Development Studies*, 9(2):170–179.

Gezici, F. and Keskin, B. (2005). Interaction between regional inequalities and internal migration in turkey. In *ERSA Conference Papers*. European Regional Science Association.

Gimpel, J. G. and Schuknecht, J. E. (2001). Interstate migration and electoral politics. *The Journal of Politics*, 63(1):207–231.

Guimaraes, P. and Lindrooth, R. C. (2007). Controlling for overdispersion in grouped conditional logit models: A computationally simple application of dirichlet-multinomial regression. *The Econometrics Journal*, 10(2):439–452.

Hacettepe University Institute of Population Studies (2008-2013). Turkey demographic and health survey, 2008–2013.

Hanneke, S., Fu, W., and Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4(none):585 – 605.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.

Hoff, P. D. and Ward, M. D. (2004). Modeling dependencies in international relations networks. *Political Analysis*, 12(2):160–175.

Huang, P. and Butts, C. T. (2021). Parameter estimation procedures for exponential-family random graph models on count-valued networks: A comparative simulation study. *arXiv preprint arXiv:2111.02372*.

Huang, P. and Butts, C. T. (2022). Rooted america: Immobility and segregation of the inter-county migration networks. *arXiv preprint arXiv:2205.02347*.

İçduygu, A. and Sert, D. (2009). Country profile: Turkey. *Focus Migration*, 5.

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.

Keele, L. and Kelly, N. J. (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political analysis*, 14(2):186–205.

Koc, I., Hancioglu, A., and Cavlin, A. (2008). Demographic differentials and demographic integration of Turkish and Kurdish populations in Turkey. *Population Research and Policy Review*, 27(4):447–457.

Koramaz, T. K. and Dökmeci, V. (2017a). Türkīye'de īller arasi göç (1995-2000). In Koramaz, T. K., Dökmeci, V., and Özdemir, Z., editors, *Türkiye'de göç ve illerin demografik, ekonomik ve fiziksel dönüşümü*, pages 13–36. Hiperlink eğit. ilet. yay. san. tic. ve ltd. sti.

Koramaz, T. K. and Dökmeci, V. (2017b). Türkīye'de īller arasi göç tīpolojīsīnde uzaklik etkīsī. In Koramaz, T. K., Dökmeci, V., and Özdemir, Z., editors, *Türkiye'de göç ve illerin demografik, ekonomik ve fiziksel dönüşümü*, pages 213–236. Hiperlink eğit. ilet. yay. san. tic. ve ltd. sti.

Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic journal of statistics*, 6:1100.

Krivitsky, P. N. and Butts, C. T. (2017). Exponential-family random graph models for rank-order relational data. *Sociological Methodology*, 47(1):68–112.

Krivitsky, P. N. and Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46.

Krivitsky, P. N., Hunter, D. R., Morris, M., and Klumb, C. (2021). ergm 4: New features. *arXiv e-prints*, pages arXiv–2106.

Krivitsky, P. N., Koehly, L. M., and Marcum, C. S. (2020). Exponential-family random graph models for multi-layer networks. *psychometrika*, 85(3):630–659.

Kuhn, R. (2015). Internal migration: Developing countries. *International Encyclopedia of the Social & Behavioral Sciences*, 12:433–442.

Leal, D. F. (2021). Network inequalities and international migration in the americas. *American Journal of Sociology*, 126(5):1067–1126.

Levy, M. (2010). Scale-free human migration and the geography of social networks. *Physica A: Statistical Mechanics and its Applications*, 389(21):4913–4917.

Lusher, D., Koskinen, J., and Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge: Cambridge University Press.

Massey, D. S. (1990). Social structure, household strategies, and the cumulative causation of migration. *Population Index*, 56(1):3–26.

Massey, D. S. and España, F. G. (1987). The social process of international migration. *Science*, 237(4816):733–738.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444.

Minhas, S., Hoff, P. D., and Ward, M. D. (2019). Inferential approaches for network analysis: Amen for latent factor models. *Political Analysis*, 27(2):208–222.

Nelson, J. F. (1984). The dirichlet-gamma-poisson model of repeated events. *Sociological Methods & Research*, 12(4):347–373.

Newman, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701.

Palmer, J. R. and Pytlikova, M. (2015). Labor market laws and intra-european migration: The role of the state in shaping destination choices. *European Journal of Population*, 31(2):127–153.

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. *Investigaciones Regionales*, 2016(36Specialissue):63–86.

Schwartz, A. (1973). Interpreting the effect of distance on migration. *Journal of Political Economy*, 81(5):1153–1169.

Snijders, T. A. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*, 21(1-2):149–172.

Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395.

Snijders, T. A. and Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Stadtfeld, C. and Block, P. (2017). Interactions, actors, and time: Dynamic network actor models for relational events. *Sociological Science*, 4:318–352.

Stadtfeld, C., Hollway, J., and Block, P. (2017a). Dynamic network actor models: Investigating coordination ties through time. *Sociological Methodology*, 47(1):1–40.

Stadtfeld, C., Hollway, J., and Block, P. (2017b). DyNAMs and the grounds for actor-oriented network event models: A rejoinder to Snijders and Butts. *Sociological Methodology*, 47:56–67.

Strasser, S. (2008). Europe's other: Nationalism, transnationals and contested images of Turkey in Austria. *European Societies*, 10(2):177–195.

Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review*, 10(3):251–259.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1762.

Van Tubergen, F., Maas, I., and Flap, H. (2004). The economic incorporation of immigrants in 18 western societies: Origin, destination, and community effects. *American Sociological Review*, 69(5):704–727.

Westveld, A. H. and Hoff, P. D. (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics*, 5(2A):843–872.

Windzio, M. (2018). The network of global migration 1990–2013: Using ergms to test theories of migration between countries. *Social Networks*, 53:20–29.

Windzio, M., Teney, C., and Lenkewitz, S. (2021). A network analysis of intra-eu migration flows: how regulatory policies, economic inequalities and the network-topology shape the intra-eu migration space. *Journal of Ethnic and Migration Studies*, 47(5):951–969.

Yazgi, B., Dokmeci, V., Koramaz, K., and Kiroglu, G. (2014). Impact of characteristics of origin and destination provinces on migration: 1995–2000. *European Planning Studies*, 22(6):1182–1198.

A Prior and proposal distributions

In the hierarchical specification, θ_1, θ_2, θ_3, and θ_4 are assigned flat priors ($\eta_i(\theta_i) \propto 1$), while for the hyperparameters of the mean and covariance of the prior distribution of ϑ_i's we set $\mu_h = 0$ and $\sigma_h^2 = 100$, $\nu_h = 1$ and $\Psi_h = 1$. We use a Metropolis-Hastings update for $\theta_1, \theta_2, \theta_3$. The proposal distributions are set as $q_j(\theta_j'|\theta_j) = \mathcal{N}(\theta_j'; \theta_j, 0.025^2 \times C_j^{-1})$, where

- C_1 is the sample covariance matrix calculated from $\{u_t(i) : i = 1, \dots, n; t = 1, \dots, T\}$,
- C_2 is the sample covariance matrix calculated from $\{v_t(i) : i = 1, \dots, n; t = 1, \dots, T\}$,
- C_3 is the sample covariance matrix calculated from $\{z_t(i,j) : i, j = 1, \dots, n; t = 1, \dots, T\}$,

We use a random walk Metropolis-Hastings to update θ_4 and ϑ_is also, with $q_0(\theta_4'|\theta_4) = \mathcal{N}(\theta_4'; \theta_4, 10^{-4})$ and $q_0(\vartheta_i|\vartheta_i) = \mathcal{N}(\vartheta_i'; \vartheta_i, 10^{-4})$.

B Additional numerical results

B.1 Histogram of the parameters of the hierarchical Dirichlet-Multinomial model

Figure 9 shows the histograms for θ_1, θ_2, θ_3, and θ_4 (the 'scale' parameter) obtained from 5×10^5 samples for the hierarchical Dirichlet-multinomial model. Figure 10 shows the same for μ_0 and Σ_0, the mean and variance parameters for the province-specific intercepts.

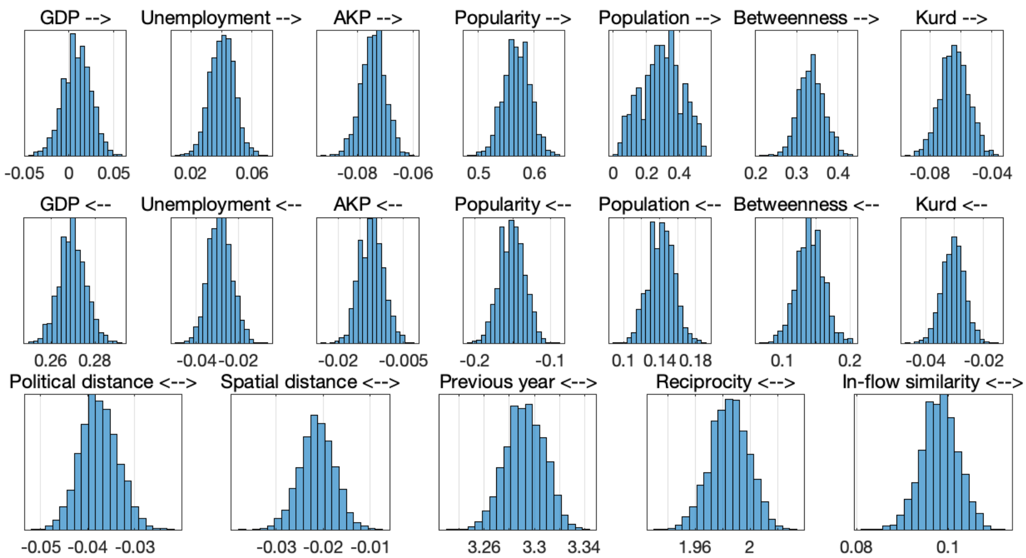

Figure 9: Histograms for θ_1, θ_2, and θ_3 for the hierarchical Dirichlet-multinomial model.

B.2 Excluding explanatory variables based on past migration

Figure 11 shows the box plots of marginal posterior distributions of factor coefficients in the hierarchical Dirichlet-multinomial model, after excluding variables that are based on lagged

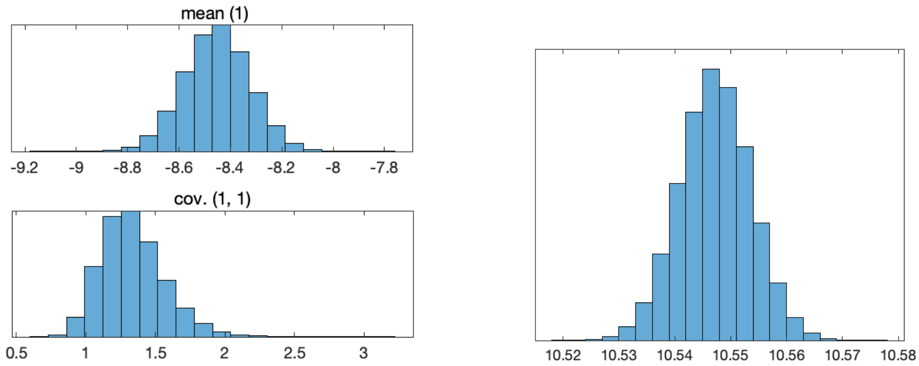

Figure 10: Left: Histograms for (μ_0, Σ_0), the mean and variance parameters for the province-specific intercepts, and θ_{sc}, the 'scale' parameter, for the hierarchical Dirichlet-multinomial model.

values of migration. These coefficients do not show a general suppression effect for those reported in the paper which are adjusted for lagged measures. Figure 12 shows the posterior correlations of these parameters. Figure 13 shows the posterior distributions of the intercepts per province. Figure 14 (left) shows the histograms for θ_1, θ_2, and θ_3. Figure 15 shows the histograms for (μ_0, Σ_0) (left) and θ_4, the 'scale' parameter (right).

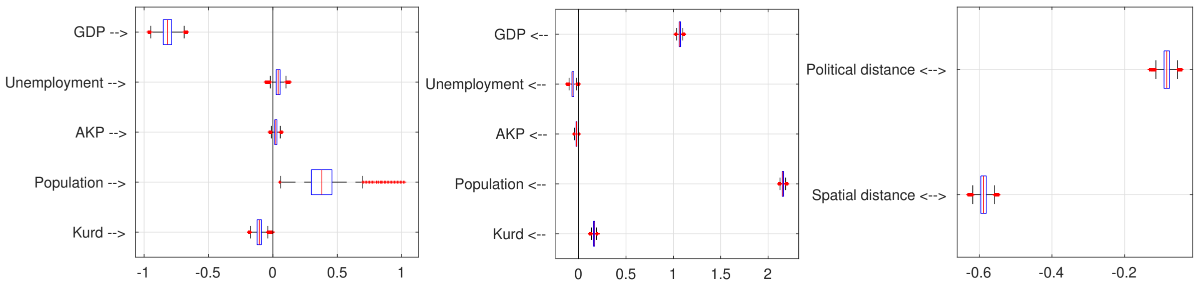

Figure 11: Box plots of marginal posterior distributions in the hierarchical Dirichlet-multinomial model described, after excluding variables that are based on lagged values of migration.

B.3 Predictive performance of the variants of the Dirichlet-multinomial model

Table 3 shows the prediction results for the hierarchical Dirichlet-multinomial model and its simpler variants. Those simpler variants are a non-hierarchical specification with a single common base parameter instead of random province bases, and a specification that excludes variables based on the previous year's migration flows. Comparing the errors in the first and the second columns with the last, we can see that not regressing on the past data considerably reduces the predictive performance. Also, comparing the errors in the first and the third columns, we can see that the random province-specific base parameters improve the performance. The latter observation suggests that the inclusion of province-specific base parameters does not lead to overfitting, hence capturing a genuine aspect of the true phenomenon.

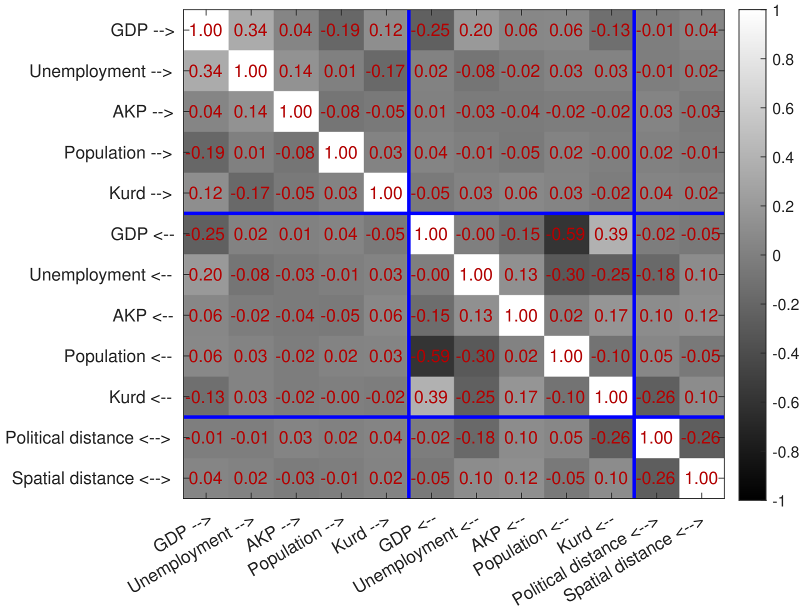

Figure 12: The posterior correlation matrix of factor coefficients in the hierarchical Dirichlet-multinomial model, after excluding variables that are based on lagged values of migration.

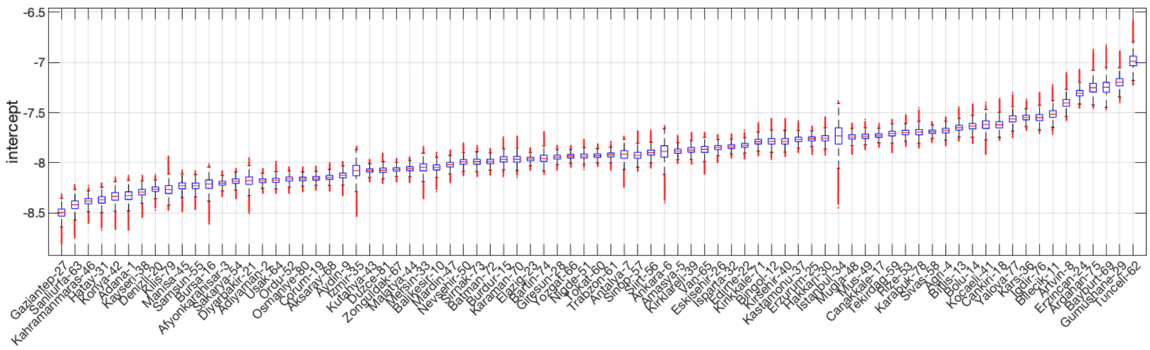

Figure 13: Box plots of the marginal posterior distributions of random intercepts for provinces in the hierarchical Dirichlet-multinomial model, after excluding variables that are based on lagged values of migration.

Table 3: Predictive performance (mean squared predictive errors) of the hierarchical Dirichlet-multinomial model and its simpler variants

Predicted value	**Diagonals**	**DM-hier**	**DM common θ_0**	**DM-hier wo past**
log counts	excluded	0.1273	0.1327	0.8566
log counts	included	0.1257	0.1311	0.8460
counts($\times 10^6$)	excluded	0.0927	0.1096	0.7395
counts($\times 10^6$)	included	1.0847	1.8661	3.0924

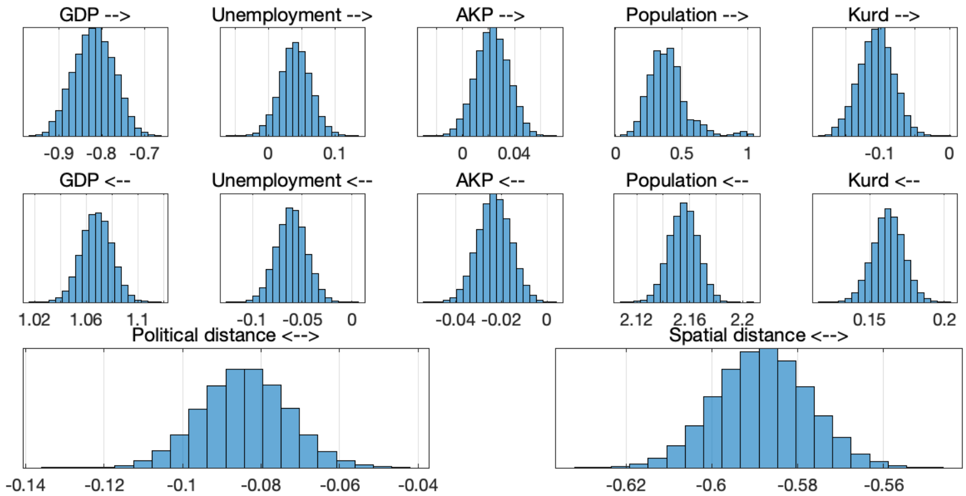

Figure 14: Histograms for θ_1, θ_2, and θ_3 obtained from 5×10^5 samples for the hierarchical Dirichlet-multinomial model, after excluding variables that are based on lagged values of migration.

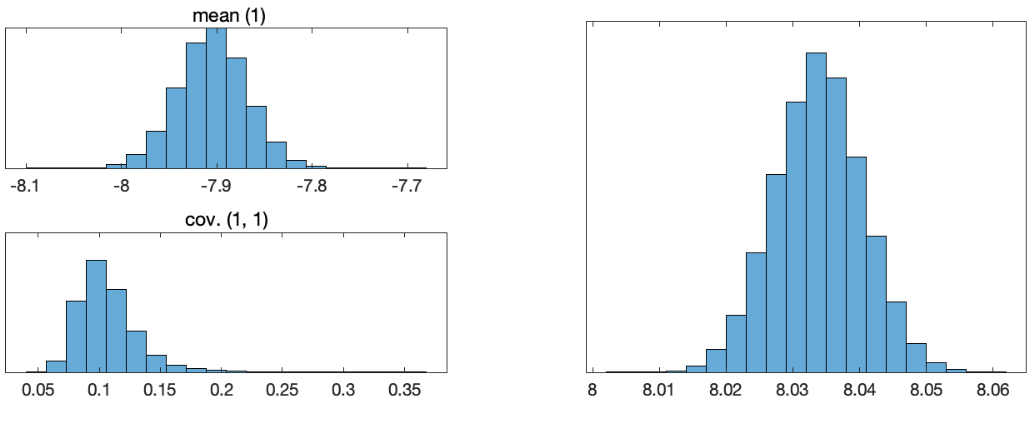

Figure 15: Histograms for (μ_0, Σ_0) (left) and θ_4 (right) for the hierarchical Dirichlet-multinomial model, after excluding variables.