Urban exodus or rural shrinkage? Regional migration and attractiveness in a tight Dutch housing market*

Thomas de Graaff[1,2]

[1]Department of Spatial Economics, Vrije Universiteit Amsterdam, The Netherlands
[2]Tinbergen Institute, Amsterdam, The Netherlands

21st September 2021

Abstract

In this paper, I address the impact of home-ownership and social renting rates on interregional migration in the Netherlands. I focus especially on their relation with natives' migration out of the larger and more popular Dutch urban regions. By applying a multilevel social relations model I am able to control simultanously for (*i*) both region-specific effects of origin and destination, (*ii*) dyad (regional pair) specific effects, and (*iii*) the impact of the housing market structure in both the region of origin and the region of destination. I find positive and high elasticities of social renting (0.8) and homeownership (1.8) rates on out-migration, while homeownership rates have a smaller and negative impact (-0.5) on in-migration—pointing to the significant role the housing market has on the crowding out of natives in the larger urban regions. On top of that, I find that regional specific in- and out-migration flows are highly correlated (0.9) just as regional specific dyadic flows (0.8), showing each region's and dyad's ideosyncratic migration pattern. Finally, I show that the probabilistic model proposed is able to accurately predict migration flows both within and out-of-sample.

* Corresponding author: t.de.graaff@vu.nl.

Keywords: gravity model, housing market, interregional migration, multilevel social relations model, Regional Attractiveness
JEL-classification: .

1 Introduction

In the recent decade cities have been proclaimed to be the overall “winners” within the regional socio-economic landscape (Glaeser, 2012). Indeed, there is an abundant empirical literature that finds that especially large metropolitan areas exhibit—on average—relatively more employment, more innovation and produce overall more added value (see, e.g., Balland et al., 2020). Most of this success of (large) cities can be attributed to positive regional and urban agglomeration economies (see for recent overviews of the size, scope and nature of these urban economies Melo et al., 2009; Duranton and Puga, 2020; Rosenthal and Strange, 2020)

Arguably, however, urban benefits do not accrue to everyone equally and recent empirical research has highlighted the negative sides of the proclaimed urban success. For example, there is ample empirical evidence of rising levels of economic segregation witin cities (Tammaru et al., 2015), of suburbanization of poverty (Hochstenbach and Musterd, 2018), and crowding out of the housing market by short-term rentals (Koster et al., 2018) and by the increasing influx of high-skilled migrants to the most popular (inner) cities (Beckers and Boschman, 2019).

Indeed, many native residents of large and popular cities are moving out (cite . . .)

Figure 1 illustrates this by showing out-migration rates of the urban region of Amsterdam for various age cohorts in the period 2011–2019.

To anticipate the results of this paper, I find strong negative effects of home-ownership rates on on both in- and out-migration flows. Further, social renting rates also affect regional migration flows negatively, but only for out-migration. So, regions accociated with . . .

This paper reads as follows. The next section describes the data and focuses especially on the distribution of regional migration flows and housing market structure. Section 3 describes the modelling approach, where starting from traditional gravity model and using the descriptives of the migration flows, a Bayesian multilevel gravity model is constructed. Section 4 gives both the model results and interprets them by providing as well predictions within and out-of-sample. The last section concludes.

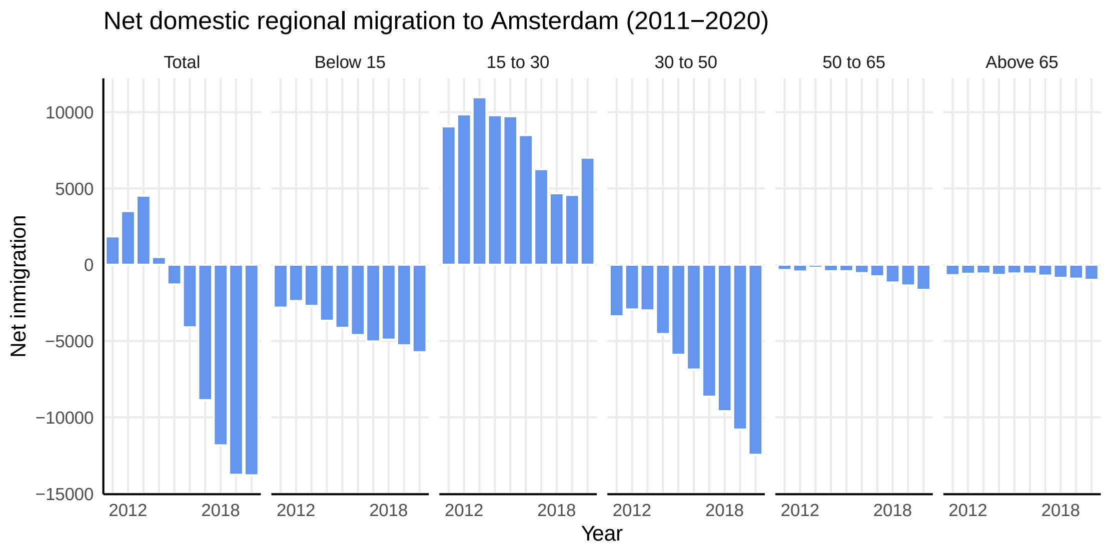

Figure 1: Net domestic regional domestic migration into Amsterdam in the period 2011–2020 for various age cohorts (including total population in the most left-panel)

2 Literature

The current Dutch housing market is characterized by a low housing supply elasticity, high housing demand, and only a small segment (10%) that belongs to the private rental market (Michielsen, Groot and Maarseveen, 2017). Because of the resulting large shortage of housing, many policy recommendations have been proposed, including plans for 700,000 new houses to be built in the coming decade. Unfortunately, the construction of a large amount of new dwellings is problematic in the Netherlands due to large and restrictive spatial constraints (Michielsen, Groot and Veenstra, 2019). Therefore, policy makers consider as well to change the housing structure, e.g., by converting rent controlled housing to home-ownership properties, to tackle at least the high housing prices by enlarging the home-ownership segment.

There is already a large amount of literature looking into the positive and negative effects of home-ownership—both from a private and from a social perspective (see for an overview Dietz and Haurin, 2003). It is argued that home-ownership leads to, e.g., better maintenance (of the own dwelling and the neighbourhood), more savings, higher education outcomes, higher individual labor supply, and even better health. One prominent negative effect of home-ownership is that it leads to less incentives to move residence because of higher moving costs vis-à-vis private renting.

This paper revisits the role of housing market structure as impediment for inter-municipality migration and specifically focuses on the role of home-ownership and social renting rates. In addition, it gives a framework to predict changes in the whole network of migration flows, when housing market structure change locally. To this end, I adopt a Bayesian multilevel gravity model which is not frequently encountered in the traditional gravity literature.[1] Traditional gravity modelling has the disadvantage that either municipality fixed effects of origins and destinations can be flows. This paper circumvents this disadvantage by adopting a multilevel incorporated or the municipalities' characteristics when not varying over approach with partial pooling[2], where the latter terms indicates that I do not impose fixed effects to control for origin and destination specific effects, but that I "draw" them from a distribution, hence the name partial pooling (where complete pooling states no group effects and no pooling fixed effects).

This papers adds two main elements to the literature. First, it does not only consider home-ownership but as well municipal social renting structure, which can be argued (see, e.g., Hughes and McCormick, 1981; Boyle and Shen, 1997; Boyle, 1998) to have a large effect on regional mobility as well as social renting rights are usually only valid locally (within municipality) and are lost when moving residence between municipalities.

Second, a partial pooling approach has another advantage, namely the municipal varying effects are completely probabilistic, making it feasible to predict both within and out-of-sample. In other words, with the results at hand I can predict migration flows between existing *and* hypothetical cities. The former might be used for looking at counterfactuals; for example, the changes in in-migration for all municipalities, when one municipality changes its housing structure. The latter is useful when one wants to assess new migrations flows between one or even two new municipalities outside the sample.[3]

That housing market structure has a sizeable effect on migration decisions is empirically well-established, especially at the micro-level, where it is widely accepted that home-ownership has a negative effect on municipal mobility (Dietz and Haurin, 2003; Dohmen, 2005). For example, Palomares-Linares and van Ham (2018) find that home-ownership

[1] That is, in the economic literature; a notable exception is Ranjan and Tobias (2007) in the economic trade literature. In the geographical literature this approach is more commonly adopted (see within a migration context Congdon, 2010; Congdon and Lloyd, 2012)

[2] There is a whole variety of names for these types of models, including hierarchical modeling, varying effects, mixed effects and shrinkage models. I use the more generic multilevel description as municipality and flows are by definition measured at a different level (scale) (see Gelman, Carlin et al., 2013, for an indepth discussion).

[3] See for probabilistic predictions of internation migration Azose and Raftery, 2015.

has a very strong immobility effect on internal migration in Spain during the period 2001–2011.

In the literature, less attention has been given to inter-city migration on the aggregate level with respect to the housing market as a specific barrier.[4] For the UK, Congdon (2010) found within a multilevel gravity model that social rented housing had little effect on the attractivity of a region, although it had a small positive effect on preventing people from moving residence. For the Canadian case, Amirault et al. (2016) looked at the impact of home-ownership on migration flows within a gravity model using a Poisson pseudo maximum likelihood estimator and found an elasticity around -1.

One of the main reasons to look into housing market structure and migration is that higher moving costs are detrimental to the aggregate labor market (Oswald, 1996; Oswald, 1999). There is a large empirical literature (see, e.g., Munch et al., 2006; Munch et al., 2008; De Graaff and Van Leuvensteijn, 2013) looking at the impact of individual and aggregate home-ownership on labour market performance, where seemingly paradoxically at the aggregate level home-ownership is indeed harmful for labour market behaviour where at the individual level it is correlated with positive labour market performance.

This difference between individual and aggregate level is explained by sorting. Home-owners are indeed less mobile than private renters because of higher fixed and sunk moving costs which has a negative *aggregate* effect on labour market performance. However, home-owners are different from renters as they do *individually* better on the labour market (due to individual unobserved heterogeneity). So home-owners in countries with high home-ownership rates perform worse on the labour market vis-à-vis home-owners in countries with low home-ownership rates; but they still perform better than private renters. For social renters, the effect is different from private renters. On the individual level they are less mobile than renters at the free market as well, but their labor market performance is also worse than that of private renters (Hughes and McCormick, 1981; De Graaff, Van Leuvensteijn and Van Ewijk, 2009).

[4] See Cushing and Poot (2004) for a historical overview of common themes within migration research.

3 Modeling framework

3.1 The traditional gravity model

In most disciplines, the workhorse model to study aggregate empirical migration flows has been the gravity model (see Anderson (2011) for a generic survey of the use of gravity models and Poot et al. (2016) for an overview of migration applications). I therefore start by adopting the basic gravity model specification pioneered by Tinbergen (1962), so:

$$\text{migrants}_{ij} = \text{M}_i^{\beta_1} \text{M}_j^{\beta_2} \text{dist}_{ij}^{\gamma}, \tag{1}$$

where migrants_{ij} are the number of migrants moving from i to j, M_i (M_j) denotes the 'mass' of i (j), and dist_{ij} the distance between i and j. Usually, the 'mass' variables are proxied by population, gross domestic product, density, etcetera. Moreover, the variable dist_{ij} may represent in general all sorts of frictions, not only physical distance.

Crucially, Anderson and Van Wincoop (2003) argue that origin and destination specific variables should be incorporated to take into account multilateral resistance terms. Most often, this is done by log-linearising model (1 and incorporating fixed effects for origins and destinations, as follows:

$$\log(\text{migrants}_{ij}) = o_i + d_j + \gamma \log(\text{dist}_{ij}) \tag{2}$$

Note that now all origin and destination specific variables are absorbed by the fixed effects o_i and d_j and that only variables affecting the frictions (dist_{ij}) can be incorporated, which could be cumbersome if one is especially interested in those variables.[5,6]

Figure 2 denotes the problem schematically in a generic dyadic type of network. Typically, one wants to model migration flows between i and j, whilst taken into account both the regional specific effects (o_i and d_j) and the regional variables ($\mathbf{X}_i$ and $\mathbf{X}_j$) one is interested in, such as housing market, population structure or cultural variables.

[5] If there is another variable dimension—say, repeated observations over time—then this problem might be circumvented. However, this requires enough variation in the data as time-invariant variables can still not be taken into account.

[6] An often applied strategy to overcome this problem is to use differences between origin and destination specific variables. Take for example Δh_{ij} as the difference in home-ownership rates between i and j. A disadvantage of this approach is that the difference between 10% and 20% home-ownership rates and the difference between 80% and 90% home-ownership rates would be valued as the same.

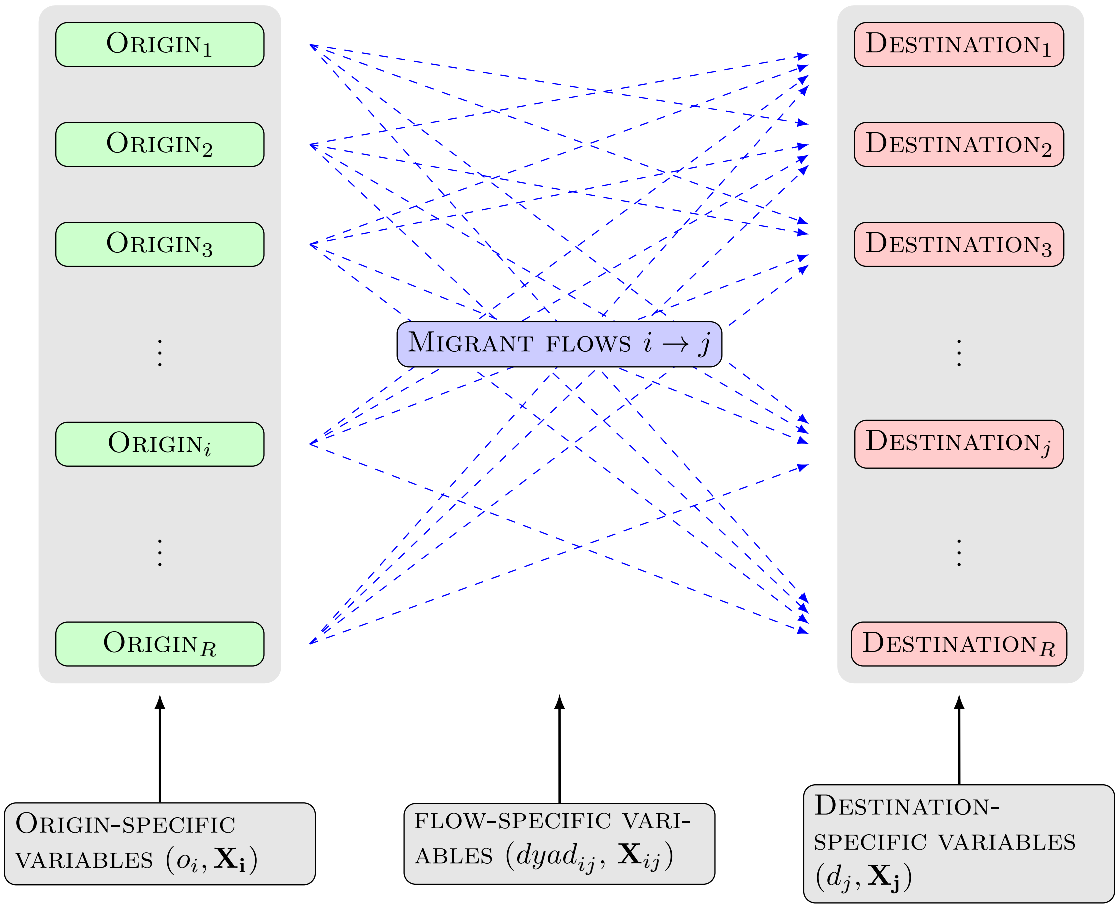

Figure 2: Decomposition of variables impacting migration flows from i to j ($\{i, j\} \in \{1, \ldots, R\}$)

Moreover, equation (2) is typically estimated with linear regression type of models, which is often very cumbersome given the large amount of zeros migrants flows. 'Quick and dirty' remedies as adding a small amount to the flow variable or removing all zeros have been proven to seriously bias the results (Linders and De Groot, 2006; Burger et al., 2009). An additional problem is that migrant flow data usually suffers from heteroskedasticity—with distributions skewed to the right (Silva and Tenreyro, 2006, actually states that heteroskedasticity (rather than the presence of too many zeros) is responsible for the main source of bias within gravity models.). For continuous outcome variables a strategy to overcome this is to make use of Pseudo Poisson Maximum Likelihood (PPML) models. If the outcome variable is a count variable one often resorts to negative-binomial models.

Note that Figure 2 as well denotes, apart from $\mathbf{X}_{ij}$, country-specific of dyad effects (dyad_{ij}) able to model unobserved region-pair specific effects. Although now frequently used in the trade literature (Baier, Bergstrand et al., 2018; Baier, Yotov et al., 2019), incorporating them in a model such as in model (2) requires a large amount of (variation in) panel-data.

To overcome these hurdles, I opt in the next subsection for an alternative strategy which can tackle simultaneously issues above: incorporating both city and dyad varying effects and region and dyad specific variables, controlling for heteroskedasticity and modelling the distribution of migrants flows as they are displayed in Figure 3—even when being zero.

3.2 A multilevel social relations model for domestic regional migrant flows

As regional migrant flows are discrete, non-negative and relatively rare given the size of the population, theoretically the most appropriate way to go forward is to model migrant flows with a Poisson type of model. To account for the multiplicative nature of the theoretical model as in (1), I adopt a log link for the expected number of migrants λ_{ij} in the Gamma-Poisson model. Apart from the theoretical model, note that this log link ensures as well that the expected number of migrants is always positive. Further, I assume that $\log(\lambda_{ij})$ is a linear function of the municipal specific variables and the distance between i and j.

To adopt both municipality effects and variables I adopt a multilevel model with partial pooling. This entails that the regional specific effects (unlike fixed effects) are now drawn from a, in this case Normal, distribution, where the parameters of this distribution are estimated as well (in the Bayesian literature they are known as well as hyper-parameters). Intuitively, this entails that regions are partially pooled indicating that (statistical) information between regions is shared. This is an attractive feature, as fixed effects assume no pooling. In that case, the model only learns from the information contained in that specific region whereas with partial pooling it is ensured that outliers (very high or low effects) are effectively *shrunk* towards the mean. Indeed, this is a further extension of that best feature of linear regression: regression towards the mean.

Likewise, I adopt a similar strategy for regional pair effects, where I assume that there is a specific dyad effect for flows from i to j *and* from flows j to i, drawn again from a normal distribution. Together with the other variables for origins, desinations and flows, the complete model now looks as follows:

$$\text{Migrants}_{ij} \sim \text{Poisson}(\lambda_{ijt}), \tag{3a}$$

$$\log(\lambda_{ijt}) = \alpha + o_i + d_j + t_t + \text{dyad}_{ij} + \ln(\mathbf{X}_{it})\beta_1 + \ln(\mathbf{X}_{jt})\beta_2 + \ln(\mathbf{X}_{ijt})\gamma. \tag{3b}$$

The first part (3a) models the outcome variable, being the number of migrants moving from region i to region j in year t, using a Poisson distribution (with parameter λ_{ijt}). The second part (3b) states that the poisson outcome variable is on a log-scale and that the explanatory variables are on a log-scale as well, allowing for the parameter (vectors) β_1, β_2 and γ to be elasticities.

As it may well be that, because of unobserved factors, regions are more or less open for both in- and outmigration, I assume that the regional specific origin, o_i, and destination, d_j, variables are correlated as follows:

$$\begin{pmatrix} o_i \\ d_j \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_{ij} \\ \rho_{ij} & \sigma_j^2 \end{pmatrix} \right\}. \tag{4}$$

Here, σ_i and σ_j control the amount of pooling. If they converge to zero, there is complete pooling and regions of origin and destination come from the same distribution. If both deviation parameters tend to be very large (relative to the overall varation in the data), there is no pooling of regions. Every region of origin and destination is then a separate identity which should be modeled with the use of fixed effects. When σ_i and σ_j are positive and relatively small (again, relative to the overall variation in the data) there is partial pooling yielding shrinkage: outliers in the data are effectively shrunk by the model to the overall mean—where the amount of shrinkage is governed by the size of σ_i and σ_j. The correlation parameter ρ_{ij} models the correlation between in- and outflows of regions, as I hypothesize that regions with relatively large amounts of immigration tend as well to have large amounts of outmigration. Equation (4) indicates as well that the region specifi c effects are mean centered around zero, where the final effect can be deduced as deviations around the grand mean α.

The dyad specific effects are drawn as well from a multivariate normal distribution as follows:

$$\begin{pmatrix} \text{dyad}_{ij} \\ \text{dyad}_{ji} \end{pmatrix} \sim \mathcal{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\text{dyad}} & \rho \\ \rho & \sigma^2_{\text{dyad}} \end{pmatrix} \right\}, \tag{5}$$

where σ^2_{dyad} denotes again the amount of partial pooling at the dyad level. Note that for dyadic relationship it does not matter whether flows go from i to j of from j to i. Moreover, parameter ρ measures whether the size of flows from i to j correlates with the size of flows from j to i. Given the tight Dutch housing market in the Netherland, this is more than likely as houses tend not to be vacant very long.

Finally, to control for year specific effects, we include a year varying effect, t_t, drawn from a normal distribution as follows: $t_t \sim \mathcal{N}(0, \sigma_t)$.

4 Data and methods

4.1 Data

I use interregional migration flows measured in individuals migrating between all of the 40 Dutch COROP regions between 2012 and 2020. I use the period 2012–2019 for estimation and the year 2020 for out-of-sample forecasting.[7] I use no information on within regional migration. So, I have 320 time-varying regional characteristics (or doubled when accounting for both origin and destination municipalities) and 12,480 aggregate migration flows ($8 \times (40 \times 40 - 40)$).

The histograms in Figure 3 show the distribution of migrant flows within the sample. The left panel deals with migrant flows below 100, the right panel with migrant flows of 100 and larger. Two main observations can be made. First, there is a strong but consistent decay pattern in migration flow size in both panels, which points to a persistent underlying pattern. However, the right 'tail' in this distribution is rather thick.[8] In effect, this entails that there are still observations quite far right in the distribution. Indeed, the sample mean is about 273, while the sample variance is around 300,000, leading to a

[7] The year 2020 is obviously a strange year containing the start of the COVID-19 outbreak in March 2020. I delibarately choose not to incorporate this year at it might affect the results, and at the same time out-of-sample prediction might contain evidence whether COVID-19 truly changed migration patterns.

[8] The largest migration flows are between the urban regions of Amsterdam and Utrecht and amount to 7,327 migrants in 2019.

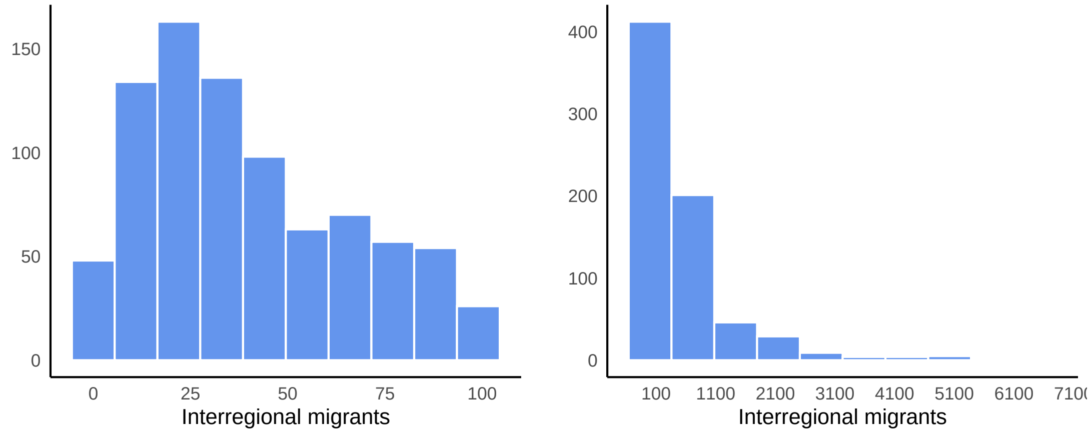

Figure 3: Histogram of domestic interregional migrant flows. Left panel shows the histogram of small migrant flows ($0 \leq N < 100$) and the right panel shows the histogram of large migrant flows ($N \geq 100$). Note the different scale of the y-axes.

strong presence of *overdispersion* (unconditional on other explanatory variables). Second, a small part of the dataset consists of zero observations. Although they do seem to be genuine observations and not caused by another process, I check in a robustness analysis whether the occurrence of zeros does need to be taken specifically into account.

Seven explanatory variables are added to the model. First, to account for spatial distance decay between origin i and destination j, distance between all regions are calculated as Eucledian distance between regional centroids (dist_{ij}). Secondly, as regional mass we use population size for both region of origin and region of destination (so pop_i and pop_j). Finally, for housing market structure we use variables indicating percentage of homeownership (home_i and home_j) and percentage of social renting (soc_i and soc_j), again in both regions of origin and destination. Social renting in the Netherlands includes all kinds of rent controlled housing but typically involves local housing corporations offering housing to lower income households, where eligibility is based on (local—within region) waiting lists.

Figure 4 shows the distribution of social renting and homeownership across Dutch regions in 2018. Clearly, both homeownership and social housing are prevalent across Dutch regions, with an average per region of 25% of social housing and around 60% of homeownership. In the last decade, these numbers have been quite robust. Percentage homeownership grew in the period 2012–2020 with only a 0.5 percentage point, while social renting decreased with 1.5 percentage point. Within regions these numbers are less stable. For example, in the region of Amsterdam social renting decreased from 0.39

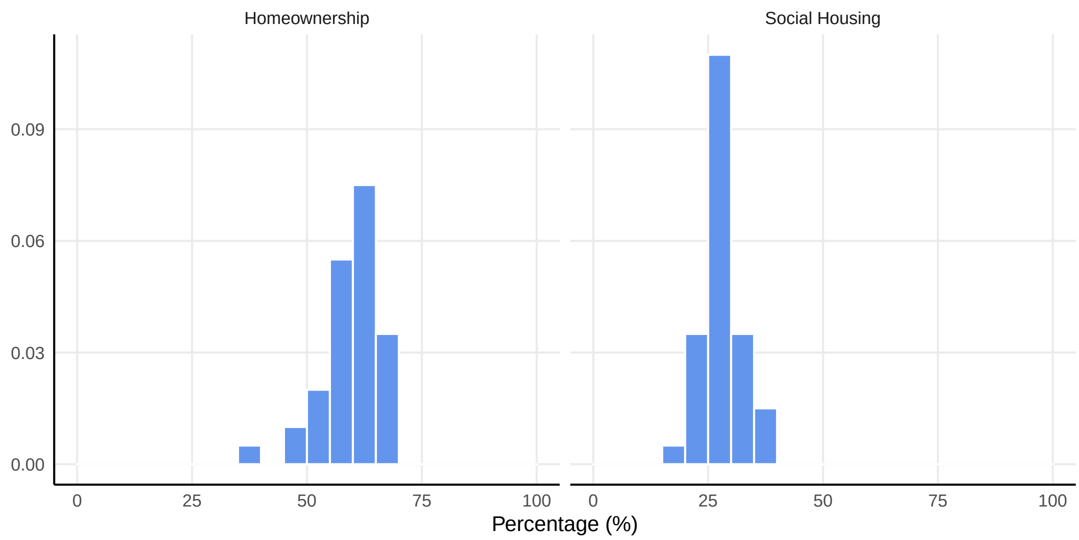

Figure 4: Histogram of homeownership (left) and social housing (right) percentages in Dutch COROP regions in 2018

percent to 0.35 percent of total housing stock.

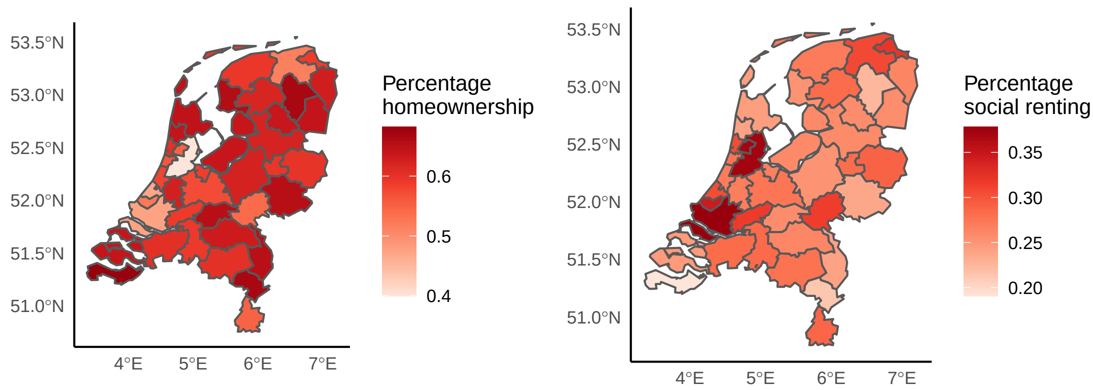

Figure 5: Maps of homeownership (left) and social housing (right) percentages in Dutch COROP regions in 2018

Figure 9 shows the regional distribution of social renting and homeownership in 2018. Clearly, social renting is especially prevalent in the larger urban regions with a correlation of 0.46 between regional size and social renting (e.g., Amsterdam has about a 37% social renting rate in 2018). Also, more rural Dutch regions exhibit much less social renting. Homeownership and regional size correlate negatively (-0.63). Finally, the maps show as well that there is a large negative correlation between social renting and homeownership

(-0.88) across regions.

4.2 Modeling strategy

For my main estimation, I construct four baseline models, where the first is the most simple gravity model as presented in equation (1) but then with a log-link and using a Poisson distribution on the outcome and the fourth the preferred and most complete specification as presented in equation (3).

So, the first model is specified as:

$$\begin{aligned} \text{Migrants}_{ij} \sim & \text{Poisson}(\lambda_{ijt}), \\ \log(\lambda_{ijt}) = & \alpha + \beta_1 \ln(\text{pop}_{it}) + \beta_2 \ln(\text{pop}_{jt}) + \gamma \ln(\text{dist}_{ijt}). \end{aligned} \tag{model 1}$$

the second model allows for regional housing market structure:

$$\begin{aligned} \text{Migrants}_{ij} \sim & \text{Poisson}(\lambda_{ijt}), \\ \log(\lambda_{ijt}) = & \alpha + \beta_1 \ln(\text{pop}_{it}) + \beta_2 \ln(\text{pop}_{jt}) + \gamma \ln(\text{dist}_{ijt}) + \\ & \beta_3 \ln(\text{home}_{it}) + \beta_4 \ln(\text{home}_{jt}) + \beta_5 \ln(\text{soc}_{it}) + \beta_6 \ln(\text{soc}_{jt}). \end{aligned} \tag{Model 2}$$

The third model adds origin and destination regional varying effects to model (Model 2), as well as time-varying effects:

$$\begin{aligned} \text{Migrants}_{ij} \sim & \text{Poisson}(\lambda_{ijt}), \\ \log(\lambda_{ijt}) = & \alpha + o_i + d_j + t_t + \\ & \beta_1 \ln(\text{pop}_{it}) + \beta_2 \ln(\text{pop}_{jt}) + \gamma \ln(\text{dist}_{ijt}) + \\ & \beta_3 \ln(\text{home}_{it}) + \beta_4 \ln(\text{home}_{jt}) + \beta_5 \ln(\text{soc}_{it}) + \beta_6 \ln(\text{soc}_{jt}). \end{aligned} \tag{Model 3}$$

Finally, the fourth and preferred model adds dyad specific effects to model (Model 3):

$$\begin{aligned}
\text{Migrants}_{ij} \sim & \text{Poisson}(\lambda_{ijt}), \\
\log(\lambda_{ijt}) = & \alpha + o_i + d_j + t_t + \text{dyad}_{ij} + \\
& \beta_1 \ln(\text{pop}_{it}) + \beta_2 \ln(\text{pop}_{jt}) + \gamma \ln(\text{dist}_{ijt}) + \\
& \beta_3 \ln(\text{home}_{it}) + \beta_4 \ln(\text{home}_{jt}) + \beta_5 \ln(\text{soc}_{it}) + \beta_6 \ln(\text{soc}_{jt}).
\end{aligned} \tag{Model 4}$$

As models (model 1)–(Model 4) are estimated with Bayesian Markov Chain Monte Carlo methods, all parameters can be given priors.[9]. Based on previous literature and data descriptives we assume α to be $\mathcal{N}(4, 2)$, γ to be $\mathcal{N}(-1.5, 1)$, β_1 and β_2 to be $\mathcal{N}(1, 0.5)$, and β_3–β_6 to be $\mathcal{N}(0, 1)$ distibuted. All variation parameters (σ_i, σ^2_{dyad} and σ_t) are assumed to be exponentially distributed with parameter 1, mainly to ensure a positive standard deviation whilst the tails are still relativily flat. The correlation parameters, ρ_{ij} and ρ, are assumed to follow a LKJ(2) distribution which ensures a relatively low probability mass close to the extremes (-1 and 1) (Lewandowski et al., 2009).[10]

4.3 Estimation

I fit models (model 1)–(Model 4) with Markov Chain Monte Carlo sampling provided by the STAN platform for statistical modeling (Gelman, Lee et al., 2015), more notably its implementation in the statistical software package R (R Core Team, 2021; Stan Development Team, 2020). For data wrangling, coding and visualisation I follow McElreath (2020) and Kurz (2021). All models are run with 4 chains, each with 4,000 iterations and a warm-up sample of 1,000. Model diagnostics for all models are deemed satisfactorily as each parameter has a large effective sample size (usually more than a 1,000) and has a $\widehat{R}$ at least smaller than 1.05 (but usually very close to 1) indicating convergence (Vehtari et al., 2019).

[9] It is also possible to give parameters flat priors, essentially reducing the models to maximum likelihood models.

[10] Although the priors on the standard deviation and correlation parameters are important as they ensure that restrictions on those parameters are satisfied, the priors on the α, β and γ parameters are in this case less essential given the relatively large number of observations.

5 Results

Table 1: Estimation results—Benchmark models

Variable	Model 1	Model 2	Model 3	Model 4
b_dist	-1.345^{***}	-1.385^{***}	-1.686^{***}	-1.628^{***}
	(0.001)	(0.001)	(0.001)	(0.024)
b_popb	0.827^{***}	0.843^{***}	0.516^{***}	0.549^{***}
	(0.001)	(0.001)	(0.031)	(0.032)
b_popa	0.790^{***}	0.770^{***}	0.319^{***}	0.318^{***}
	(0.001)	(0.001)	(0.033)	(0.033)
cons	4.793^{***}	4.777^{***}	4.595^{***}	4.490^{***}
	(0.001)	(0.001)	(0.176)	(0.164)
b_sB		-1.474^{***}	0.867^{***}	0.867^{***}
		(0.007)	(0.054)	(0.054)
b_hB		-1.143^{***}	0.169^{**}	0.166^{**}
		(0.008)	(0.086)	(0.087)
b_sA		-1.821^{***}	-0.245^{***}	-0.259^{***}
		(0.007)	(0.055)	(0.054)
b_hA		-1.670^{***}	1.575^{***}	1.601^{***}
		(0.008)	(0.088)	(0.086)
sigma_y			0.125^{***}	0.125^{***}
			(0.046)	(0.042)
sigma_gr.1			0.656^{***}	0.673^{***}
			(0.073)	(0.082)
sigma_gr.2			0.466^{***}	0.441^{***}
			(0.053)	(0.053)
Rho_1.3			0.803^{***}	0.781^{***}
			(0.060)	(0.070)
sigma_d				0.385^{***}
				(0.009)
Rho_2.3				0.804^{***}
				(0.014)
PSIS_loo	$849,398$	$793,476$	$384,369$	$123,396$
R2	0.67	0.70	0.79	0.84

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

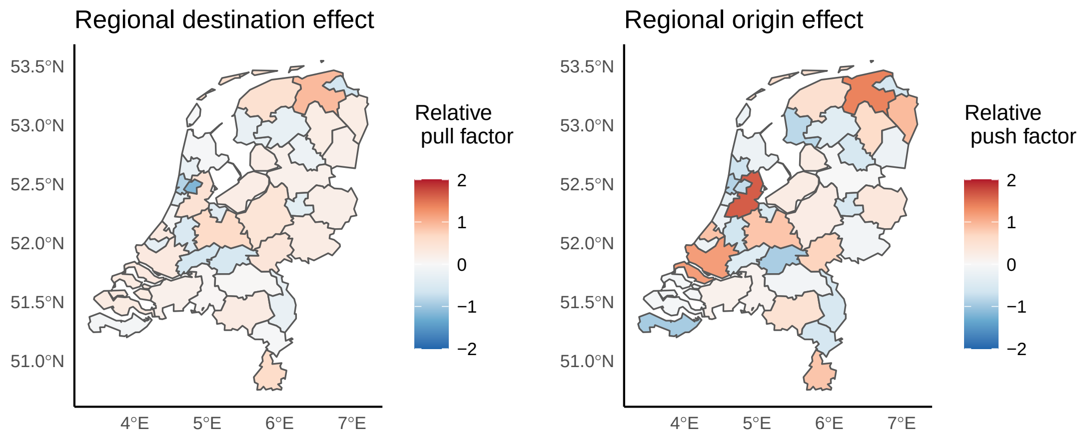

Figure 6: Maps of regional pull/destination (left panel) and push/origin (right panel) factors in Dutch COROP regions in the period 2012–2019.

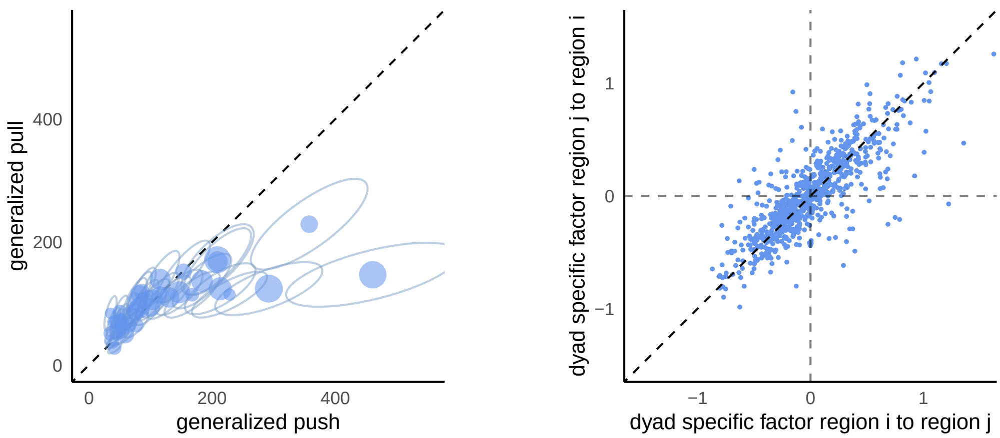

Figure 7: Maps of regional pull/destination (left panel) and push/origin (right panel) factors in Dutch COROP regions in the period 2012–2019.

6 Discussion

7 Conclusion

Acknowledgments

I would like to thank Wim Bernasco, Thor Husby, Viktor Venhorst, Maureen Lanhuizen, participants at the 15th biennial NECTAR conference in Helsinki, participants at the 59th

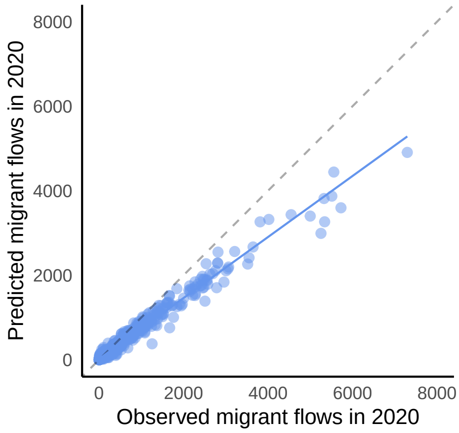

Figure 8: Maps of regional pull/destination (left panel) and push/origin (right panel) factors in Dutch COROP regions in the period 2012–2019.

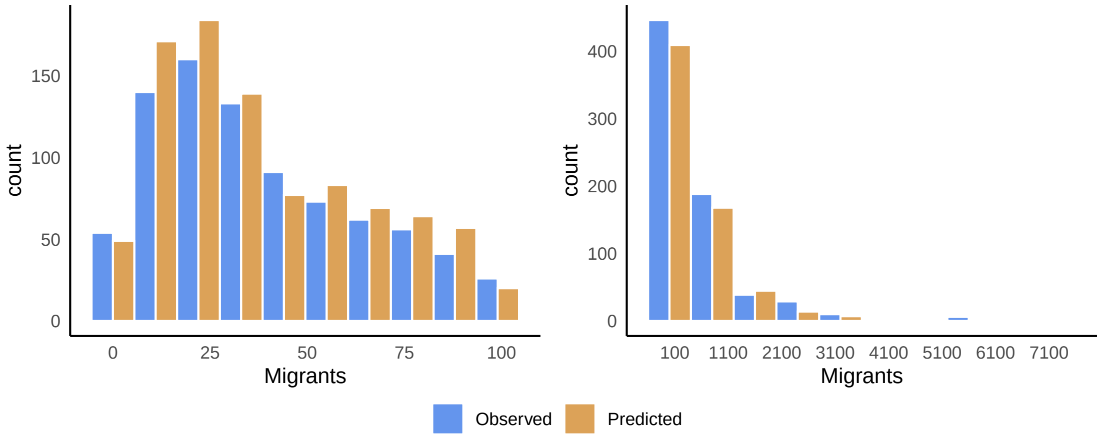

Figure 9: Maps of regional pull/destination (left panel) and push/origin (right panel) factors in Dutch COROP regions in the period 2012–2019.

ERSA conference in Lyon and seminar participants at the VU University Amsterdam for valuable comments on a first draft of this paper. Paper, data and code can be retrieved from the following project's GitHub page:

https://github.com/Thdegraaff/migration_gravity.

Acknowledgements

References

Amirault, D., D. de Munnik and S. Miller (2016). ‘What drags and drives mobility? Explaining Canada’s aggregate migration patterns’. *Canadian Journal of Economics/Revue canadienne d’économique* 49.3, pp. 1035–1056.

Anderson, J. E. (2011). ‘The gravity model’. *Annu. Rev. Econ.* 3.1, pp. 133–160.

Anderson, J. E. and E. Van Wincoop (2003). ‘Gravity with gravitas: a solution to the border puzzle’. *American economic review* 93.1, pp. 170–192.

Azose, J. J. and A. E. Raftery (2015). ‘Bayesian probabilistic projection of international migration’. *Demography* 52.5, pp. 1627–1650.

Baier, S. L., J. H. Bergstrand and M. W. Clance (2018). ‘Heterogeneous effects of economic integration agreements’. *Journal of Development Economics* 135, pp. 587–608.

Baier, S. L., Y. V. Yotov and T. Zylkin (2019). ‘On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration’. *Journal of International Economics* 116, pp. 206–226.

Balland, P.-A., C. Jara-Figueroa, S. G. Petralia, M. P. Steijn, D. L. Rigby and C. A. Hidalgo (2020). ‘Complex economic activities concentrate in large cities’. *Nature Human Behaviour* 4.3, pp. 248–254.

Beckers, P. and S. Boschman (2019). ‘Residential choices of foreign highly skilled workers in the Netherlands and the role of neighbourhood and urban regional characteristics’. *Urban Studies* 56.4, pp. 760–777.

Boyle, P. (1998). ‘Migration and housing tenure in South East England’. *Environment and Planning A* 30.5, pp. 855–866.

Boyle, P. and J. Shen (1997). ‘Public housing and migration: a multi-level modelling approach’. *International Journal of Population Geography* 3.3, pp. 227–242.

Burger, M., F. Van Oort and G.-J. Linders (2009). ‘On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation’. *Spatial Economic Analysis* 4.2, pp. 167–190.

Congdon, P. and C. Lloyd (2012). ‘A spatial random-effects model for interzone flows: commuting in Northern Ireland’. *Journal of Applied Statistics* 39.1, pp. 199–213.

Congdon, P. (2010). ‘Random-effects models for migration attractivity and retentivity: a Bayesian methodology’. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173.4, pp. 755–774.

Cushing, B. and J. Poot (2004). ‘Crossing boundaries and borders: Regional science advances in migration modelling’. *Papers in regional science* 83.1, pp. 317–338.

De Graaff, T. and M. Van Leuvensteijn (2013). ‘A European cross-country comparison of the impact of homeownership and transaction costs on job tenure’. *Regional Studies* 47.9, pp. 1443–1461.

De Graaff, T., M. Van Leuvensteijn and C. Van Ewijk (2009). ‘Homeownership, social renting and labor mobility across Europe’. *Homeownership and the labour market in Europe*, pp. 53–81.

Dietz, R. D. and D. R. Haurin (2003). ‘The social and private micro-level consequences of homeownership’. *Journal of urban Economics* 54.3, pp. 401–450.

Dohmen, T. J. (2005). ‘Housing, mobility and unemployment’. *Regional Science and Urban Economics* 35.3, pp. 305–325.

Duranton, G. and D. Puga (2020). ‘The economics of Urban Density’. *Journal of Economics Perspectives* 34.3, pp. 3–26.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A., D. Lee and J. Guo (2015). 'Stan: A probabilistic programming language for Bayesian inference and optimization'. *Journal of Educational and Behavioral Statistics* 40.5, pp. 530–543.

Glaeser, E. L. (2012). *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Penguin.

Hochstenbach, C. and S. Musterd (2018). 'Gentrification and the suburbanization of poverty: Changing urban geographies through boom and bust periods'. *Urban Geography* 39.1, pp. 26–53.

Hughes, G. and B. McCormick (1981). 'Do council housing policies reduce migration between regions?' *The Economic Journal* 91.364, pp. 919–937.

Koster, H., J. van Ommeren and N. Volkhausen (2018). *Short-term rentals and the housing market: Quasi-experimental evidence from Airbnb in Los Angeles*. Tech. rep.

Kurz, A. S. (2021). *Statistical rethinking with brms, ggplot2, and the tidyverse: Second edition*. version 0.2.0.

Lewandowski, D., D. Kurowicka and H. Joe (2009). 'Generating random correlation matrices based on vines and extended onion method'. *Journal of multivariate analysis* 100.9, pp. 1989–2001.

Linders, G.-J. and H. L. De Groot (2006). 'Estimation of the gravity equation in the presence of zero flows'.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

Melo, P. C., D. J. Graham and R. B. Noland (2009). 'A meta-analysis of estimates of urban agglomeration economies'. *Regional science and urban Economics* 39.3, pp. 332–342.

Michielsen, T., S. Groot and R. Maarseveen (2017). 'Prijselasticiteit van het woningaanbod'. *CPB Notitie*.

Michielsen, T., S. Groot and J. Veenstra (2019). 'Het bouwproces van nieuwe woningen'. *CPB Boek*.

Munch, J. R., M. Rosholm and M. Svarer (2006). 'Are homeowners really more unemployed?' *The Economic Journal* 116.514, pp. 991–1013.

— (2008). 'Home ownership, job duration, and wages'. *Journal of Urban Economics* 63.1, pp. 130–145.

Oswald, A. J. (1996). *A conjecture on the explanation for high unemployment in the industrialized nations: Part I*. Tech. rep.

— (1999). 'The housing market and Europe's unemployment: a non-technical paper'. *Homeownership and the labour Market in Europe*.

Palomares-Linares, I. and M. van Ham (2018). 'Understanding the effects of homeownership and regional unemployment levels on internal migration during the economic crisis in Spain'. *Regional Studies*, pp. 1–12.

Poot, J., O. Alimi, M. P. Cameron and D. C. Maré (2016). 'The gravity model of migration: The successful comeback of an ageing superstar in regional science'. *Investigaciones Regionales* 2016.36, pp. 63–86.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ranjan, P. and J. L. Tobias (2007). 'Bayesian inference for the gravity model'. *Journal of Applied Econometrics* 22.4, pp. 817–838.

Rosenthal, S. C. and W. C. Strange (2020). 'How Close Is Close? The Spatial Reach of Agglomeration Economies'. *Journal of Economics Perspectives* 34.3, pp. 27–49.

Silva, J. S. and S. Tenreyro (2006). 'The log of gravity'. *The Review of Economics and statistics* 88.4, pp. 641–658.

Stan Development Team (2020). *RStan: the R interface to Stan*.

Tammaru, T., M. Van Ham, S. Marcińczak and S. Musterd (2015). *Socio-economic segregation in European capital cities: East meets West*. Routledge.

Tinbergen, J. J. (1962). 'Shaping the world economy; suggestions for an international economic policy'.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter and P.-C. Bürkner (2019). ‘Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC’. *arXiv preprint arXiv:1903.08008*.