Journal of Regional Science
Working paper

Urban exodus: housing market structure and interregional migration revisited

Thomas de Graaff[1]*

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 role on the migration of natives out of the larger and more popular Dutch cities. By applying a Bayesian multilevel gravity 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. Moreover, city specific in- and out-migration flows are highly correlated (0.88) just as regional specific dyad flows (0.8). Finally, I show that the probabilistic model proposed is able to accurately predict migration flows both within and out-of-sample.

Keywords
Gravity model — housing market — interregional migration — multilevel model — cities

[1] *Department of Spatial Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
***Corresponding author**: ✉ t.de.graaff@vu.n; thomasdegraaff.nl

1. Introduction

In the recent decade metropolitan areas 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 cities 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).

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.

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.

2. Data

I use inter-regional migration flows measured in individuals between all of the 40 Dutch COROP regions between 2012 and 2018. I use no information on within regional migration. So, I have 320 regional characteristics (or doubled when accounting for both origin and destination municipalities) and 10,902 aggregate migration flows ($7 \times (40 \times 40 - 40)$).

The histograms in Figure 2 show the distribution of migrant flows within my 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.[1] Thus, there are still observations quite far right in the distribution. Indeed, the sample mean is about 270, while the sample variance is around 290,000, leading to

[1] The largest migration flows are between the urban regions of Amsterdam and Utrecht and amount to about 6,555 migrants in 2018.

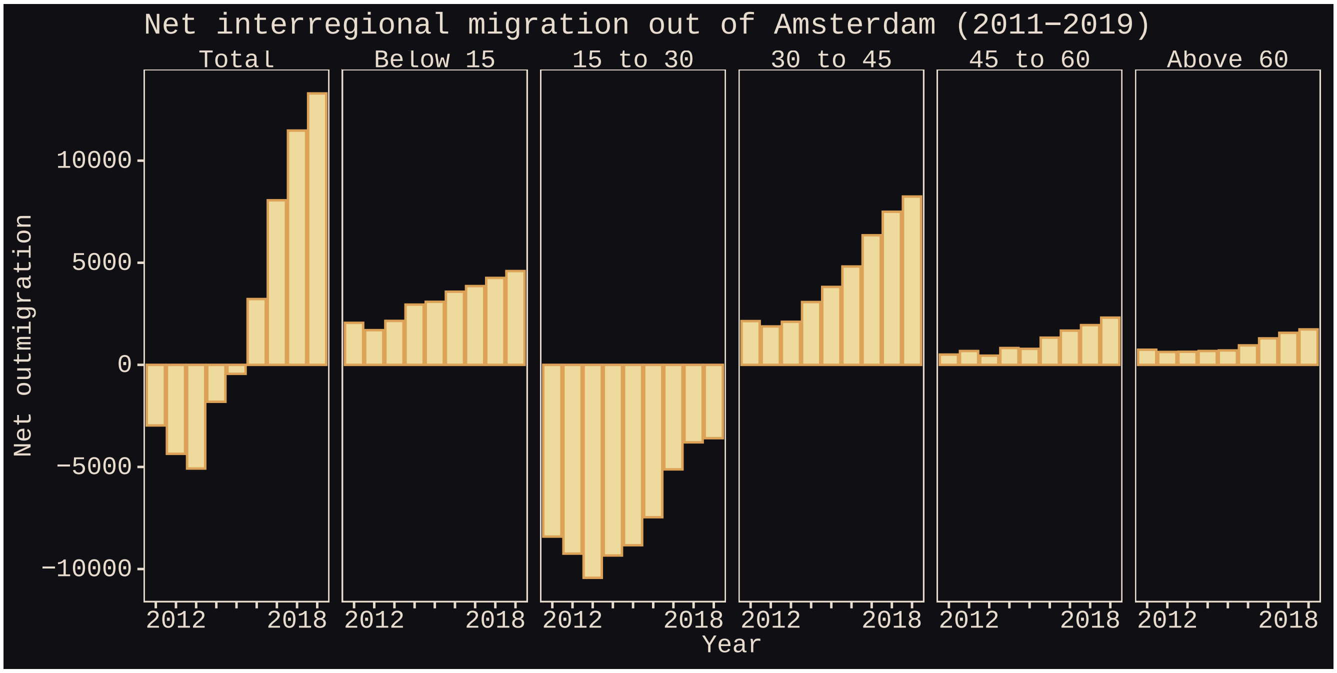

Figure 1. Interregional migration out of Amsterdam in the period 2011–2019 for various age cohorts (including total out-migration in the most left-panel)

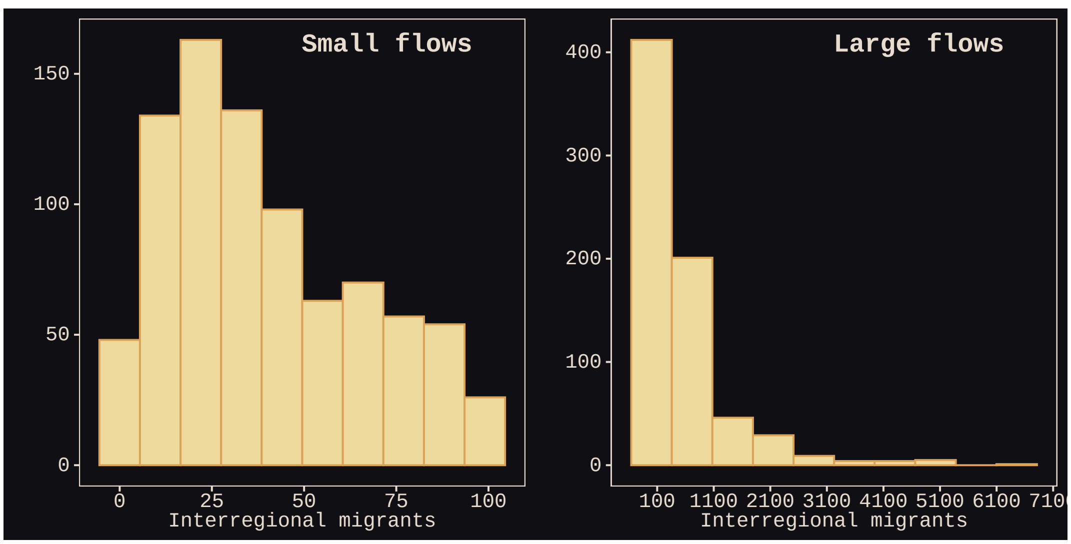

Figure 2. Histogram of inter-regional migrant flows. Left panel shows the histogram of small migrant flows ($0 \leq N < 100$) and the right panel shows the histogram of large migrant flows ($N \geq 100$). Note the different scale of the y-axes.

a 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 3 shows the distribution of social renting and homeownership across Dutch regions in 2018. Clearly, both home-

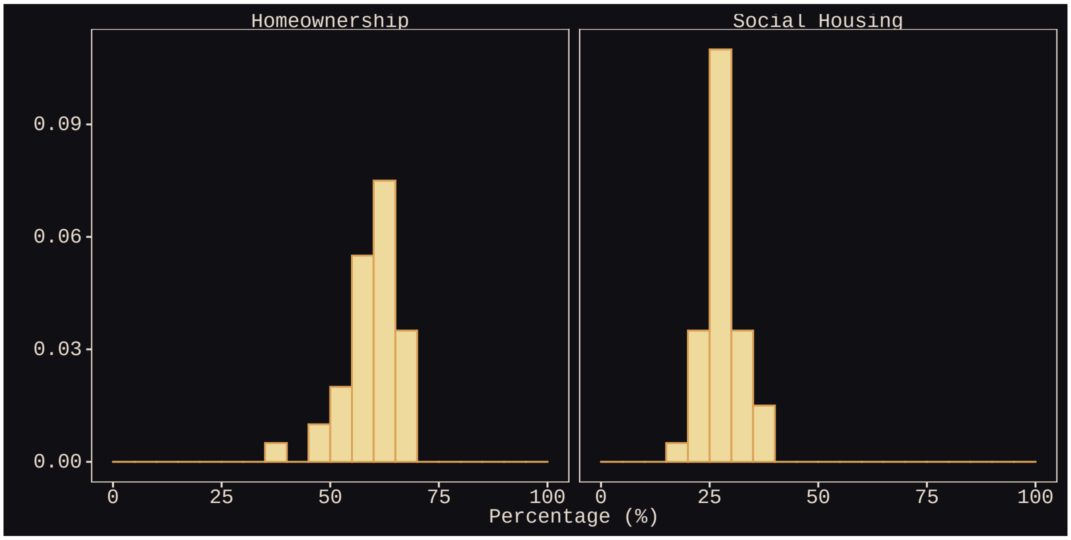

Figure 3. Histogram of social housing (left) and homeownership (right) percentages in COROP regions in 2018 (check)

ownership and social housing are prevalent across Dutch regions, with an average per city of 25% of social housing and around 60% of homeownership. Moreover, it is worthwhile to note that social renting is especially prevalent in the larger cities with a correlation of 0.46 between regional size and social renting (e.g., Amsterdam has about a 40% social renting rate). Also, more rural Dutch regions exhibit much less social renting. Homeownership and city size correlate negatively (-0.63). Finally, there is a large negative correlation between social renting and homeownership (-0.88) across regions.

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.[2][3] Figure 4 denotes the problem schematically in a

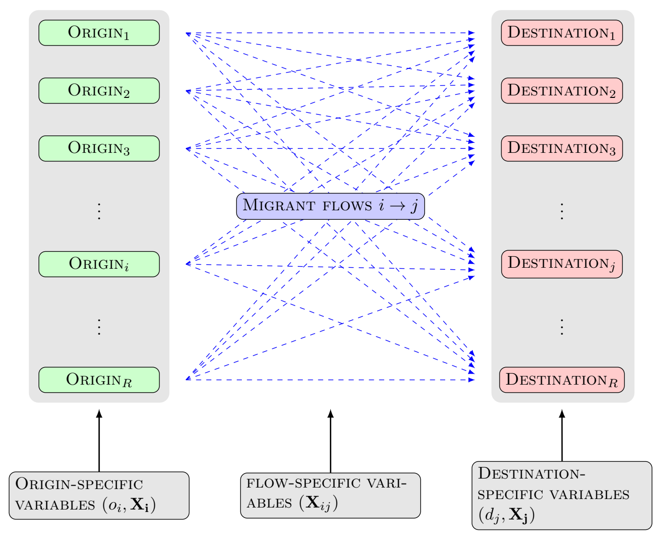

Figure 4. Decomposition of variables impacting migration flows from i to j ($\{i, j\} \in \{1, \ldots, R\}$)

[2] 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.

[3] 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.

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.

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).

Therefore, I opt in the next subsection for a different strategy, with which I can tackle simultaneously the two disadvantes of above: incorporating both city varying effects and city specific variables and modelling the distribution of migrants flows as they are displayed in Figure **??**—even when being zero.

3.2 A Bayesian multilevel gravity model

First, as municipal migrants flows are discrete, non-negative and relatively rare give the size of the population, theoretically the most appropriate way to go forward is to model migrant flows with a Poisson type of model. However, given that the sampling variance is much larger than the sampling mean of the migration flows (although not conditional on the covariates), we likely need to correct for overdispersion of heteroskedasticity (Silva and Tenreyro, 2006, states that heteroskedasticity (rather than the presence of too many zeros) is responsible for the main source of bias within gravity models.). An often used distribution to account for overdispersion is the Gamma-Poisson model (which is under re-parametrization similar to the perhaps better known negative binomial model). So, we use that for our outcome variable.

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.

Finally, to adopt both municipality effects and variables I adopt a multilevel model with partial pooling. This entails that the municipal varying 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 municipalities are partially pooled indicating that (statistical) information between municipalities 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 municipality 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.

The complete model now looks as follows:[4]

$$
\begin{aligned}
\text{Migrants}_{ij} \sim & \,\text{Gamma-Poisson}(\lambda_{ij}, \tau) && \text{(3a)}\\
\log(\lambda_{ij}) = & \,\alpha + o_{\text{mun}[i]} + d_{\text{mun}[j]} + \\
& \beta_1 \log(\text{pop}_i) + \beta_2 \log(\text{pop}_j) + \\
& \beta_3 \log(\text{home}_i) + \beta_4 \log(\text{home}_j) + \\
& \beta_5 \log(\text{soc}_i) + \beta_6 \log(\text{soc}_j) + \\
& \beta_7 \log(\text{dist}_{ij}) && \text{(3b)}\\
d_{\text{mun}} \sim & \,\text{Normal}(0, \sigma_d) && \text{(3c)}\\
o_{\text{mun}} \sim & \,\text{Normal}(0, \sigma_o) && \text{(3d)}\\
\beta_1, \ldots, \beta_7 \sim & \,\text{Normal}(0, 2) && \text{(3e)}\\
\alpha_o, \alpha_d \sim & \,\text{Normal}(0, 2) && \text{(3f)}\\
\sigma_o, \sigma_d \sim & \,\text{HalfCauchy}(0, 1) && \text{(3g)}\\
\tau \sim & \,\text{Gamma}(0.01, 0.01) && \text{(3h)}
\end{aligned}
$$

The first part (3a) models the outcome variable, being the number of migrants, using a Gamma-Poisson distribution (with parameter λ_{ij}) allowing for overdispersion by using an additional parameter τ. The being elasticities. Equations (3d) and (3c) constitute the linear part of the model is given by (3b) and states that the poisson outcome variable is on a log-scale and that most explanatory variables are on a log-scale as well, allowing for direct comparison of the parameters multilevel part, where parameters σ_o and σ_d measure the amount of pooling. If they go to zero, then the data exhibits complete pooling. If they become very large (go to infinity) there is no pooling (which is the fixed effects case). Equations (3e)–(3h) denote priors for all parameter involved. These priors are chosen that they are rather conservative. Namely, we know from previous empirical literature that the β-parameters typically are not lower than -2 or higher than 2. But given the amount of data these priors are of little influence. The only structure I impose is that the standard deviations σ_o and σ_d are assumed to be non-negative with relatively probability in the their right tails. The Gamma prior for τ is a standard and as well a very conservative prior.

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 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.

[4] I adopt here the model structure from McElreath (2020).

References

Anderson, J. E. (2011). "The gravity model". In: *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". In: *American economic review* 93.1, pp. 170–192.

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". In: *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". In: *Urban Studies* 56.4, pp. 760–777.

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". In: *Spatial Economic Analysis* 4.2, pp. 167–190.

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

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". In: *Urban Geography* 39.1, pp. 26–53.

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.

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

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". In: *Regional science and urban Economics* 39.3, pp. 332–342.

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". In: *Investigaciones Regionales* 2016.36, pp. 63–86.

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

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

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". In: