**Conference paper**
**Prepared for ERSA 2019**

# Housing market and migration revisited: a multilevel gravity model for Dutch municipalities

Thomas de Graaff[1]*

**Abstract**
This paper revisits the impact of the housing market structure on interregional migration, but adopts an alternative modeling approach to migration flows between cities. The starting point is a gravity model, but instead of using fixed effects for cities of origin and destination, I use a multilevel mixed effects approach allowing me to simultaneously model migration flow characteristics and the cities of origin and destination characteristics. This approach has two main advantages. First, it allows for simultaneous estimation of the impact of city characteristics on migration flows, where the impact is not necessarily symmetrical for cities of origin and destination. Second, it allows for prediction of migration flows between cities both in and out of sample. Preliminary results show that homeownership decrease migration flows significantly with an elasticity below $-1$. Municipal social renting rate has a negative impact as well, but its elasticity is close to zero.

**Keywords**
Gravity model — housing market — migration — multilevel model — partial pooling — prediction

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

## 1. Introduction

In the 1990s, Andrew Oswald wrote two famous working papers (Oswald, 1996, 1999) postulating that homeownership rates would have a negative impact on labor market behavior, as the high costs of moving residence associated with homeownership would impede regional mobility. These two working papers evoked a large empirical literature (see, e.g., Munch et al., 2006, 2008; De Graaff and Van Leuvensteijn, 2013) looking at the impact of individual and aggregate homeownership on labor market performance, where seemingly paradoxically at the aggregate level homeownership is indeed harmful for labor market behavior where at the individual level it is correlated with positive labor market performance.

That housing market structure has an effect on migration decisions is well-established, especially at the micro-level, where it is widely accepted that homeownership has a negative effect on regional mobility (Dietz and Haurin, 2003). For example, Palomares-Linares and van Ham (2018) find that homeownership has a very strong immobility effect on in internal migration in Spain during the period 2001–2011.

On an aggregate level, Amirault et al. (2016) already looked at the impact of homeownership on migration flows within a gravity model using a Poisson pseudo maximum likelihood estimator and found an elasticity around $-1$. However, traditional gravity modeling has the disadvantage that either regional fixed effects of origins and destinations can be incorporated or the regions' characteristics when not varying over flows. Moreover, theoretically, regional effecs should be incorporated leaving no room in the traditional approach to incorporate regional characteristics

This paper circumvents this disadvanted by adopting a multilevel approach with partial pooling[1], where the latter terms indicates that I adopt regions of 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.

A partial pooling approach has another advantege, namely the regional specific 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 flows between hypothetical regions.

This paper reads as follows. The next section describes the data and focuses especially on the distribution of regional migration flows and regional labour market structure. Section 3 describes the modeling approach, where starting from traditional gravity model and using the descriptives of migration flows I argue for a specific type of model. Section 4. gives both the model results and their analysis. By the latter I mean that this sections deals as well with interpretation by giving prediction both within and out-of-sample. The last section concludes.

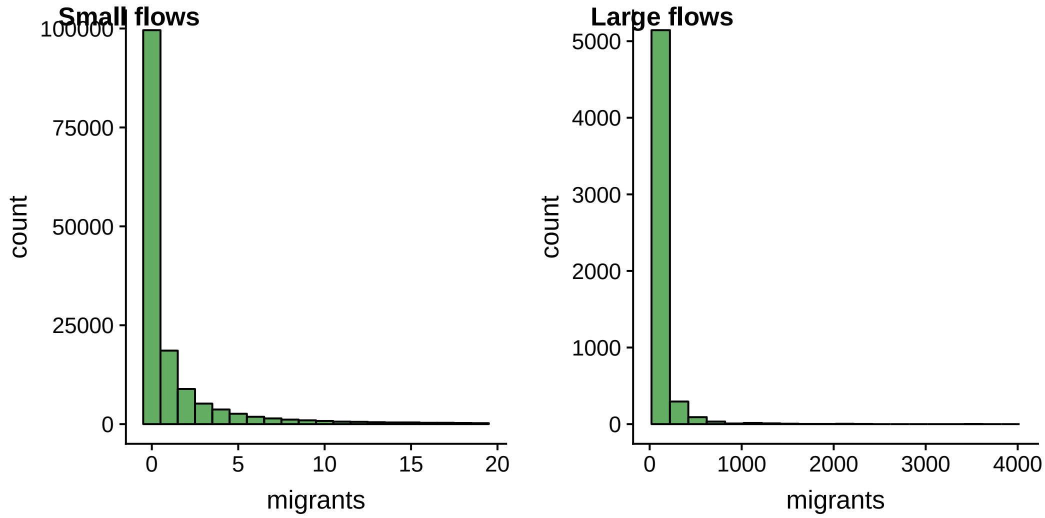


**Figure 1.** Histogram of migrant flows. Left panel shows the histogram of small migrant flows ($N < 0$) and the right panel shows the histogram of large migrant flows ($N \geq 20$). Note the different scale of the y-axis.

## 2. Data

I model migration flows measured in individuals between the 393 Dutch municipalities in 2015. There is no information available about within municipality residential migration. So, I have 393 regional characteristics (or doubled when accounting for both regions of origin and destination) and 154,056 flows ($393 \times 393 - 393$).

Figure 1 shows the distribution of migrant flows within my sample. The left panel deals with migrant flows below 20, the right panel with migrant of 20 and larger. Two main observations can be made.

First, there is strong but consistent decay in both panels, which points to a strong underlying pattern. However, the 'tail' in this distribution is rather thick.[2] There are still observations quite far right in the distributution. Indeed, the sample mean is about 10, while the sample variance is around 40, leading to a strong presence of *overdispersion*.

Secondly, most of my dataset consists of zero observations. Although they do seem to be genuine observations and not caused by another proces (we will check for this later), they do need to be taken specifically into account.

I include 7 other variables in my model. First, to account for spatial distance decay between origin $i$ and destination $j$, distance between all municipalities are calculated as Eucledian distance between centroids ($\text{dist}_{ij}$). Secondly, as municipality mass we use population size (so $\text{pop}_i$ and $\text{pop}_j$). Finally, for housing market structure we use variables indicating percentage of homeownership ($\text{home}_i$ and $\text{home}_j$ and percentage of social renting ($textsoc_i$ and $\text{soc}_j$). Social renting in the Netherlands includes all kinds of renting but typically involves local housing corporations offering housing to lower income households, where eligibility is based on (local) waiting lists. Both social renting and homeownership are assumed to impede regional mobility (De Graaff et al., 2009).

Figure 2 shows the distribution of social renting and homeownership across Dutch municipalities in 2015. Clearly, both types of housing structures are important for the Netherlands, with an average of 25% of social housing and around 60% of homeownership. Moreover, it is worthwhile to note that social renting is especially prevalent in the the larger cities with a correlation of 0.4 between city size and social renting (e.g., Amsterdam has about 40% social renting rate). Also, some smaller dutch municipalities do not exhibit any social renting. Homeownership and city size correlate negatively ($-0.51$). Finally, there is a large negative correlation between social renting and homeownership ($-0.84$) across municipalities.

## 3. Modeling framework

### 3.1 The traditional gravity model

We adopt the basic gravity model specification pioneered by Tinbergen (1962), so:

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

Note, that in model (1) the variable $\text{dist}_{ij}$ may represent all sorts of frictions, not only physical distance. Thus, in my case we incorporate variables for homeownership and social renting to account for frictions on the housing market that may impede regional mobility.

Importantly, Anderson and Van Wincoop (2003) argued that origin and destination specific variables should be incorporated to take into account multilateral resitance terms. Most

[1] There is a whole variety of names for these types of models, including varying effects, mixed effects and shrinkage models. I use the more generic multilevel description as regions and flows are by definition measured at a different level (scale).

[2] The largest migration flows are between the municipalities of Amsterdam and Amstelveen and amount to about 3,500 migrants.

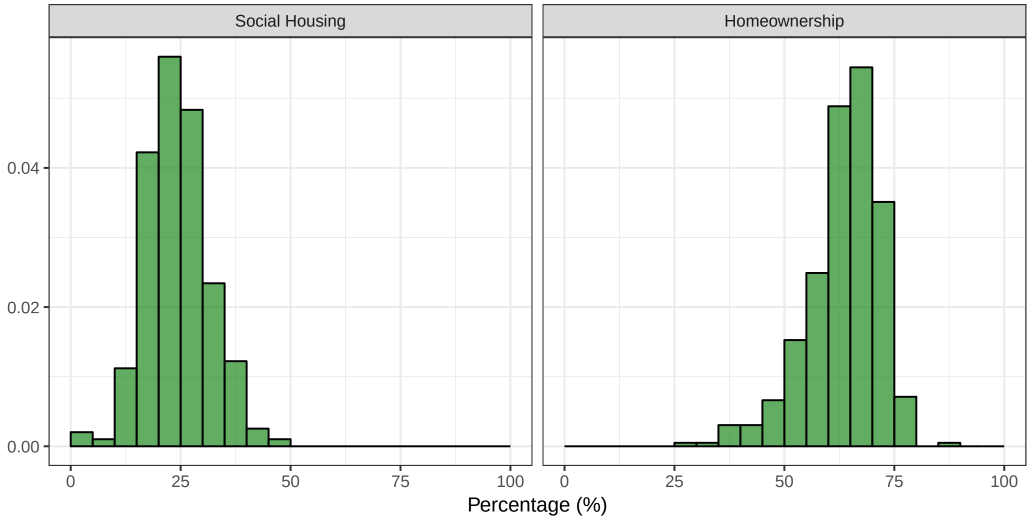


**Figure 2.** Histogram of social housing (left) and homeownership (right) percentages in Dutch municipalities 2015

oftens, this is done by log-linearizing model (1)[3] and incorporating fixed effects for origins and destinations, as follows:

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

Unfortunately, this approach does not allow for municipality specific variables; so, population and housing market variables drop out of this model. But those are exactly the variables we are interested in! Moreover, equation (2 is typically estimated with regression type of models, which is often very cumbersome given the large amount of zeros present flows of migrants.

Therefore, I next allow for a different strategy, where I would like to tackle both the disadvantes of above: incorporating both region specific effects and variables and modeling the distribution of migrants flows as they are displayed in Figure 1.

### 3.2 A multilevel gravity model

Firstly, as regional migrants flows are discrete and relatively rare give the size of the population, the most appropriate way to go forward is to model number of migrants with a Poisson type of model. However, given that the sampling variance is four times 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 differences.). An often used distribution to account for overdispersion is the gamma-poisson model (also known as the 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 expectation variable in the Poisson model.

Finally, to adopt both region effects and variables I adopt a multilevel model with partial pooling. This entails that our regional specific effects (the formerly fixed effecs) are drawn from a, in this case Normal, distribution, where the parameters of this distrbution are estimated as well (in the literature they are known as well as hyper-parameters). Intuitively, this entails that regions are partially pooled indicating that information between regions is shared. This is very attractive, as fixed effects for no pooling. The model only learns from the information contained in that specific region. The partial pooling also ensures 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 total model looks now as follows:

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

The first line (3a) models the outcome variable, being the number of migrants, using a Poisson distribution (with param-

[3] In our case, note that zeros are present in our social renting variable. We therefore add a small number to this variables (0.0001). Doing this only on the *right-hand side* does not affects our results

eter $\lambda_{ij}$) allowing for overdispersion by using an additional parameter $\tau$. The linear part of the model is given by (3b) and states that the poisson outcome space in on a log-scale and that most parameters are on a log-scale as well, allowing for direct comparison of the parameters being elasticities. Equations (3c) and (3d) constitute the multilevel part, where parameters $\sigma_o$ and $\sigma_d$ measure the amount of pooling. If they tend to zero, there the data exhibits complete pooling. If they become very large (go to infinity) there is no pooling (thus fixed effects). All other parameters are priors (chosen such that they are rather conservative but given the amount of data they are of little influence).

## 4. Results

### 4.1 Parameter estimates

I estimate model (3) by using the *No U-Turn Sampler* (NUTS) in Stan.[4] NUTS is a relatively recent developed Hamiltonian Monte Carlo (a specific form of Markov Chain Monte Carlo simulation) method, able to draw samples efficiently from large multilevel models (Hoffman and Gelman, 2014). Parameter estimates and probability intervals of the main parameters (so not the region specific effects: there are 786 of them) are given in Table 1. Perhaps more insightful, there are graphically depicted in Figure 3.

**Table 1.** Parameter estimates with 95% probability intervals (group speficic origin and destination estimates are not presented)

| Parameter | mean | sd | 2.5% | 97.5% |
|---|---|---|---|---|
| b_Intercept | -0.74 | 0.04 | -0.82 | -0.66 |
| b_pop_d | 0.89 | 0.03 | 0.83 | 0.96 |
| b_pop_o | 0.88 | 0.04 | 0.79 | 0.97 |
| b_hom_d | -1.48 | 0.19 | -1.86 | -1.10 |
| b_hom_o | -1.27 | 0.25 | -1.75 | -0.78 |
| b_soc_o | -0.04 | 0.04 | -0.11 | 0.03 |
| b_soc_d | -0.06 | 0.03 | -0.12 | -0.01 |
| b_log_distance | -1.96 | 0.01 | -1.97 | -1.95 |
| sd_destination__Intercept | 0.45 | 0.02 | 0.42 | 0.49 |
| sd_origin__Intercept | 0.61 | 0.02 | 0.57 | 0.66 |
| shape | 1.22 | 0.01 | 1.20 | 1.24 |

As most important conclusions in this stage can we say that housing structure indeed impedes regional mobility, but that it is mainly homeownership rates and not social renting rates that have a negative effect. The homeownership elasticities are slightly larger in absolute size than what Amirault et al. (2016) reported. Furthermore, estimations for parameters $\sigma_o$ and $\sigma_d$ point to more pooling than less, so fixed effects in this case might lead to substantial overfitting.

### 4.2 Model predictions

[4] As interface to Stan (see for an overview article of Stan Carpenter et al., 2017) I used the `brms` R-package (Bürkner, 2017).

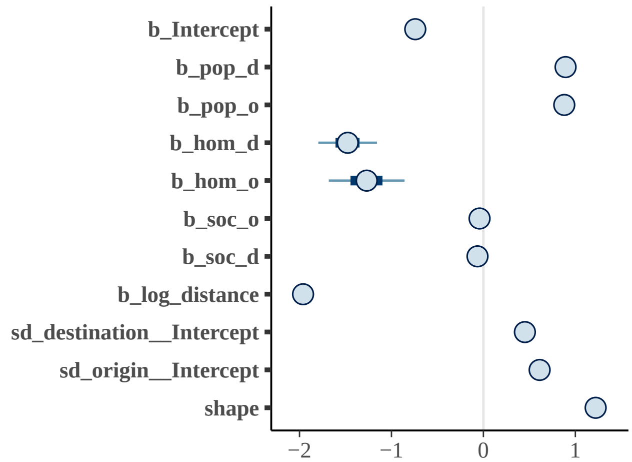


**Figure 3.** Forest plot of parameter means and 95% probability intervals (group speficic origin and destination estimates are not presented)

## 5. In conclusion

## Acknowledgments

I would like to thank Wim Bernasco for valuable comments on a first draft of this paper. Paper, data and code can be retrieved from the project's GitHub page:
https://github.com/Thdegraaff/migration_gravity.

## References

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

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.

Bürkner, P.-C. (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan". In: *Journal of Statistical Software* 80.1, pp. 1–28.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). "Stan: A probabilistic programming language". In: *Journal of statistical software* 76.1.

De Graaff, T. and M. Van Leuvensteijn (2013). "A European cross-country comparison of the impact of homeownership and transaction costs on job tenure". In: *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". In: *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". In: *Journal of urban Economics* 54.3, pp. 401–450.

Hoffman, M. D. and A. Gelman (2014). "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." In: *Journal of Machine Learning Research* 15.1, pp. 1593–1623.

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

— (2008). "Home ownership, job duration, and wages". In: *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". In: *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". In: *Regional Studies*, pp. 1–12.

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

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