HOUSING MARKET AND MIGRATION REVISITED

A multilevel gravity model for Dutch regions

Thomas de Graaff

August 25, 2020

Vrije Universiteit Amsterdam
Tinbergen Institute Amsterdam

Housing market and interregional migration: why bother?

- Dutch housing market: tight and regulated
  - large shortage of housing
  - large yearly prices increases ($\approx 5\% - 9\%$ annually)
  - decrease in housing transactions
  - large regional variation

Housing market and interregional migration: why bother?

- Dutch housing market: tight and regulated
  - large shortage of housing
  - large yearly prices increases ($\approx 5\% - 9\%$ annually)
  - decrease in housing transactions
  - large regional variation

- Regional Dutch population projections (PEARL) $\longrightarrow$ difficulties with interregional migration
  - especially long-distance migration
  - changes in local housing supply as input

Housing market and interregional migration: why bother?

- Dutch housing market: tight and regulated
    - large shortage of housing
    - large yearly prices increases ($\approx 5\% - 9\%$ annually)
    - decrease in housing transactions
    - large regional variation

- Regional Dutch population projections (PEARL) $\longrightarrow$ difficulties with interregional migration
    - especially long-distance migration
    - changes in local housing supply as input

- Large literature of external effects of home-ownership (Dietz and Haurin, 2003)
    - negative: migration (by increased moving costs) and on aggregate labour market performance (Oswald, 1996, 1999)

My contributions to the literature

- Large empirical (economic) literature on impact home-ownership as drivers of interregional migration, but:
  - usually concerns marginal effect of home-ownership
  - less attention to predictions for the whole network
- Literature on impact of social renting on migration flows is scarce (De Graaff et al., 2009)
  - In the Netherlands social renting is a large phenomenon ($\approx$ 24% of total housing stock)
  - Social renting rights only valid within city
  - Social renting is an urban phenomenon (e.g. $\approx$ 40–50% in Amsterdam)

So, this paper

Does what? Revisits the impact of housing market structure (with focus on social renting) on Dutch interregional migration flows using a multilevel gravity model

- UK context by Congdon (2010)
- social relations model *cf.* Koster and Leckie (2014)
- *Statistical Rethinking* from McElreath (2020)
- `ggplot2` code from Solomon Kurz (2020)

Aim To model the impact of housing market structure on the whole network of interregional migration flows

Why a multilevel approach for migration?

There are at least two levels in migration (I use three)

Why a multilevel approach for migration?

There are at least two levels in migration (I use three)

Observed migration flows Migration between i and j with friction (e.g., distance) attributes (obs $= R^2 - R$)

$$\text{REGION}_i \xrightarrow{\mathbf{X}_{ij}} \text{REGION}_j$$

Why a multilevel approach for migration?

There are at least two levels in migration (I use three)

Observed migration flows Migration between i and j with friction (e.g., distance) attributes (obs $= R^2 - R$)

Observed push & pull factors Attributes of i and j (obs $= R$)

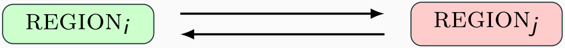
Why a multilevel approach for migration?

There are at least two levels in migration (I use three)

Observed migration flows Migration between i and j with friction (e.g., distance) attributes (obs $= R^2 - R$)

Observed push & pull factors Attributes of i and j (obs $= R$)

Observed flows between region dyads migration from $i \to j$ is correlated with migration from $j \to i$ (obs $= \frac{R^2 - R}{2}$)



Why a Bayesian multilevel approach?

- Hierarchical, mixed effects, varying intercept/parameter, shrinkage, partial pooling models

Why a Bayesian multilevel approach?

- Hierarchical, mixed effects, varying intercept/parameter, shrinkage, partial pooling models
- Increasingly used for model performance and flexibility

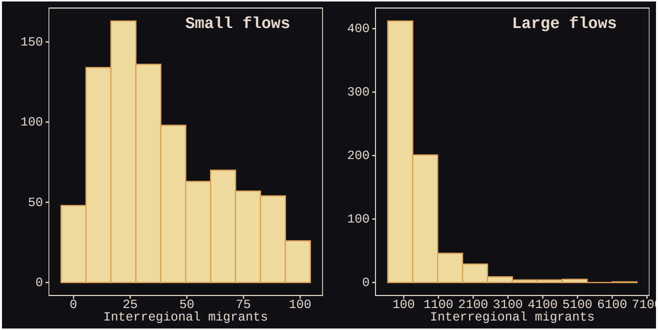
Why a Bayesian multilevel approach?

- Hierarchical, mixed effects, varying intercept/parameter, shrinkage, partial pooling models
- Increasingly used for model performance and flexibility
- Simultaneous modeling at various levels (e.g., cities, regions, flows, individuals)
  - no two-stage models anymore
  - precision (standard errors) is correct

Why a Bayesian multilevel approach?

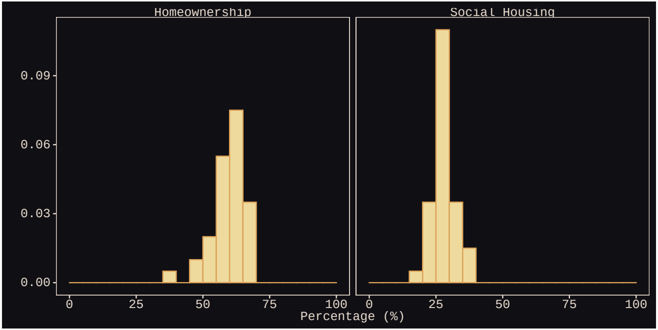
- Hierarchical, mixed effects, varying intercept/parameter, shrinkage, partial pooling models
- Increasingly used for model performance and flexibility
- Simultaneous modeling at various levels (e.g., cities, regions, flows, individuals)
    - no two-stage models anymore
    - precision (standard errors) is correct
- Partial pooling: For example, origin specific effects are drawn from a distribution: usually $\sim$ Normal(α, σ)
    - $\sigma \longrightarrow 0$: complete pooling
    - $\sigma \longrightarrow \infty$: no pooling (fixed effects)

Data: migrations flows in 2018



- Panel for the period 2012–2018
- Migration flows between 40 Dutch regions ($1,560$ flows per year)
- Variance $\gg$ mean: over-dispersion

Data: regional housing structure in 2018



- Positive correlation between population and share social renting (0.46)
- Negative correlation between share social renting and share home-ownership (−0.88)

Data: regional housing structure in 2018 (cont.)

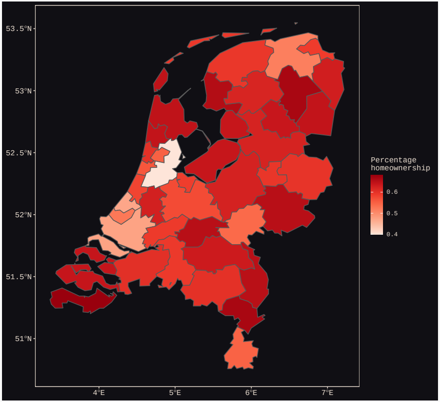


Figure 1: Share of homeownership

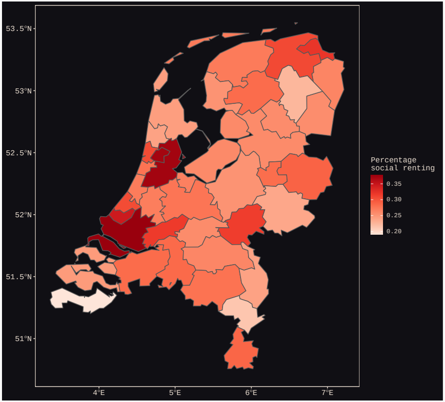


Figure 2: Share of social renting

Modeling framework: traditional gravity modeling

$$\log(\text{Migrants}_{ij}) = o_i + d_j + \gamma \log(\text{dist}_{ij}) + \epsilon_{ij}$$

Origin and destination specific fixed effects for multilateral resistance (Anderson and Van Wincoop, 2003), but:

- what about zeros in Migrants_{ij}?
- how to incorporate housing structure in the presence of o_i and d_j?
- over-dispersion and heteroskedasticity (Silva and Tenreyro, 2006)

Poisson versus negative binomial[1]

- Counts of migrants

- With Poisson & regional effects of origin and destination the following origin and destination constraints automatically hold

$$\sum_{j=1}^{R} \widehat{\text{Migrants}}_{ij} = O_i \qquad \sum_{i=1}^{R} \widehat{\text{Migrants}}_{ij} = D_j$$

- Does not apply with negative binomial

- Multilevel model accounts for dispersion

[1] *We urge researchers to resist the siren song of the Negative Binomial (Head and Mayer, 2014)*

Modeling framework: multilevel gravity modeling

$$\text{Migrants}_{ij} \sim \text{Poisson}(\lambda_{ij})$$

$$\text{Migrants}_{ji} \sim \text{Poisson}(\lambda_{ji}) \qquad \text{(flow of migrants)}$$

Modeling framework: multilevel gravity modeling

$$\begin{aligned}
\text{Migrants}_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\text{Migrants}_{ji} &\sim \text{Poisson}(\lambda_{ji}) && \text{(flow of migrants)} \\
\log(\lambda_{ij}) &= \alpha + o_i + d_j + \text{dyad}_{ij} + && \text{(varying effects)}
\end{aligned}$$

Modeling framework: multilevel gravity modeling

$$
\begin{aligned}
\text{Migrants}_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\text{Migrants}_{ji} &\sim \text{Poisson}(\lambda_{ji}) && \text{(flow of migrants)} \\
\log(\lambda_{ij}) &= \alpha + o_i + d_j + \text{dyad}_{ij} + && \text{(varying effects)} \\
&\quad \mathbf{X}_i \beta_i + \mathbf{X}_j \beta_j && \text{(explanatory variables)}
\end{aligned}
$$

Modeling framework: multilevel gravity modeling

$$\begin{aligned}
\text{Migrants}_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\text{Migrants}_{ji} &\sim \text{Poisson}(\lambda_{ji}) && \text{(flow of migrants)} \\
\log(\lambda_{ij}) &= \alpha + o_i + d_j + \text{dyad}_{ij} + && \text{(varying effects)} \\
&\quad \mathbf{X}_i \beta_i + \mathbf{X}_j \beta_j && \text{(explanatory variables)} \\
\log(\lambda_{ji}) &= \alpha + o_j + d_i + \text{dyad}_{ji} + \\
&\quad \mathbf{X}_j \beta_i + \mathbf{X}_i \beta_j
\end{aligned}$$

Modeling framework: multilevel gravity modeling

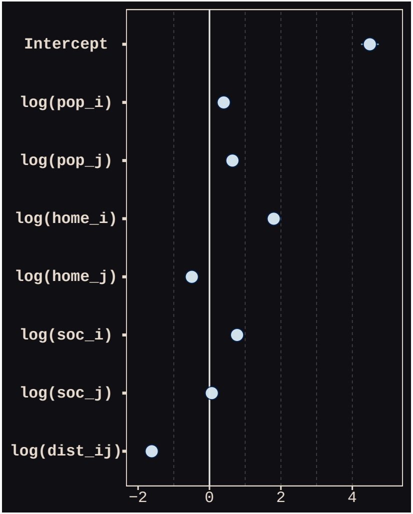
$$
\begin{aligned}
\text{Migrants}_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\text{Migrants}_{ji} &\sim \text{Poisson}(\lambda_{ji}) && \text{(flow of migrants)} \\
\log(\lambda_{ij}) &= \alpha + o_i + d_j + \text{dyad}_{ij} + && \text{(varying effects)} \\
&\quad \mathbf{X}_i \beta_i + \mathbf{X}_j \beta_j && \text{(explanatory variables)} \\
\log(\lambda_{ji}) &= \alpha + o_j + d_i + \text{dyad}_{ji} + \\
&\quad \mathbf{X}_j \beta_i + \mathbf{X}_i \beta_j \\
\begin{pmatrix} o_i \\ d_i \end{pmatrix} &\sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_o^2 & \sigma_o \sigma_d \rho \\ \sigma_o \sigma_d \rho & \sigma_d^2 \end{pmatrix} \right) \\
& && \text{(varying region effects)}
\end{aligned}
$$

Modeling framework: multilevel gravity modeling

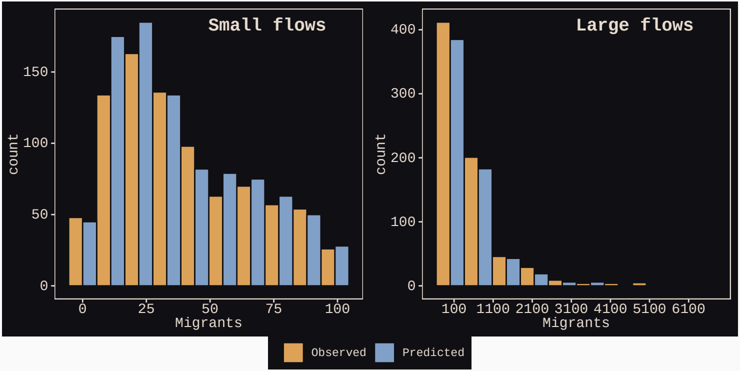
$$
\begin{aligned}
\text{Migrants}_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\
\text{Migrants}_{ji} &\sim \text{Poisson}(\lambda_{ji}) && \text{(flow of migrants)} \\
\log(\lambda_{ij}) &= \alpha + o_i + d_j + \text{dyad}_{ij} + && \text{(varying effects)} \\
&\quad \mathbf{X}_i \beta_i + \mathbf{X}_j \beta_j && \text{(explanatory variables)} \\
\log(\lambda_{ji}) &= \alpha + o_j + d_i + \text{dyad}_{ji} + \\
&\quad \mathbf{X}_j \beta_i + \mathbf{X}_i \beta_j \\
\begin{pmatrix} o_i \\ d_i \end{pmatrix} &\sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_o^2 & \sigma_o \sigma_d \rho \\ \sigma_o \sigma_d \rho & \sigma_d^2 \end{pmatrix} \right) \\
& && \text{(varying region effects)} \\
\begin{pmatrix} \text{dyad}_{ij} \\ \text{dyad}_{ji} \end{pmatrix} &\sim \text{MVNormal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\text{dyad}}^2 & \sigma_{\text{dyad}}^2 \rho_{\text{dyad}} \\ \sigma_{\text{dyad}}^2 \rho_{\text{dyad}} & \sigma_{\text{dyad}}^2 \end{pmatrix} \right) \\
& && \text{(varying dyad effects)}
\end{aligned}
$$

Estimation results

Parameter	mean	sd
Intercept (α)	4.49	0.15
$\log(\text{pop}_i)$	0.40	0.04
$\log(\text{pop}_j)$	0.64	0.03
$\log(\text{home}_i)$	1.80	0.10
$\log(\text{home}_j)$	-0.50	0.09
$\log(\text{soc}_i)$	0.77	0.07
$\log(\text{soc}_j)$	0.06	0.07
$\log(\text{dist}_{ij})$	-1.62	0.03

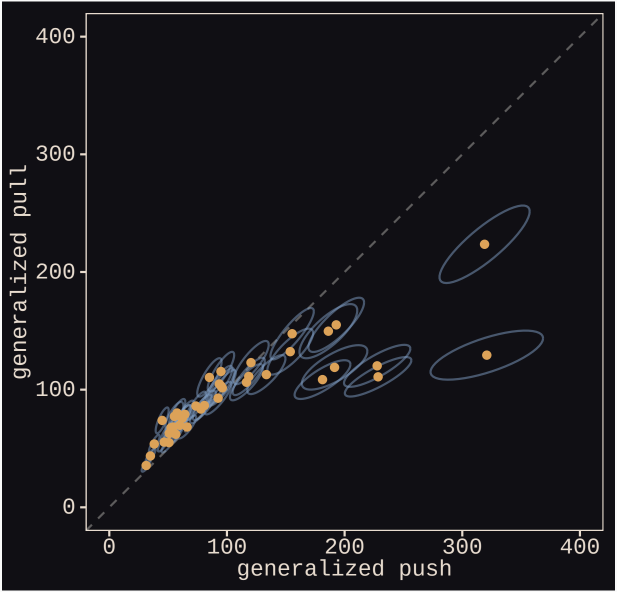


Observed versus predicted flows (correlation ≈ 0.99)

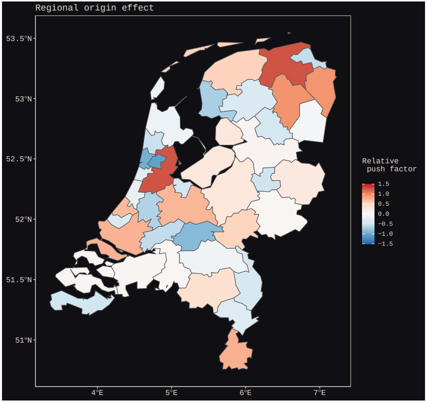


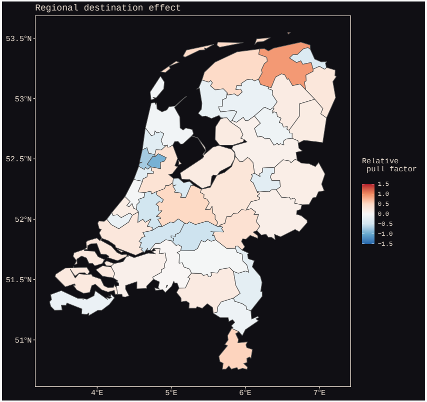
- maximum observed flow: 6,555
- maximum predicted flow: 4,704

Correlation between origin and destination $\rho = 0.88$

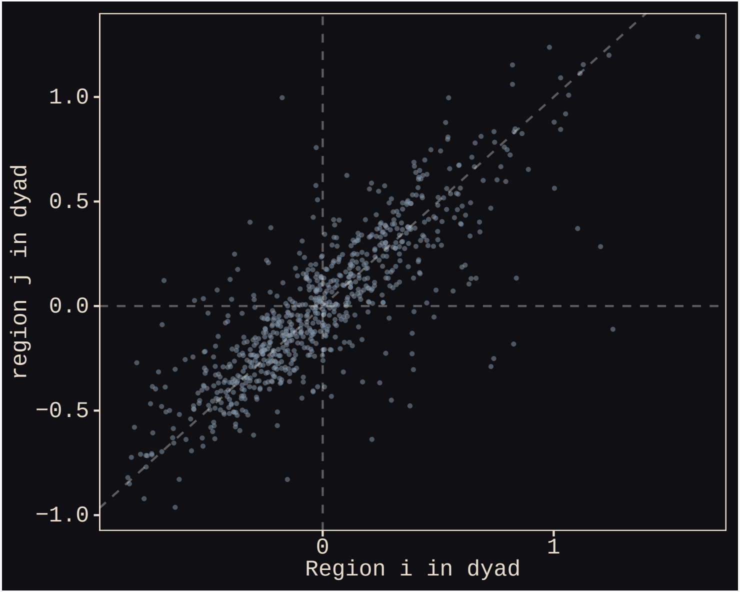


Asymmetric push and pull factors





Dyad specific effects $\rho = 0.8$



Sensitivity checks

Results are robust to

- year

Sensitivity checks

Results are robust to

- year
- interaction effects with population

Sensitivity checks

Results are robust to

- year
- interaction effects with population
- inclusion of household size

Sensitivity checks

Results are robust to

- year
- interaction effects with population
- inclusion of household size
- spatial autocorrelation in regional effects:

$$\begin{aligned} o_i, d_j &\sim \text{MVNormal}(0, \mathbf{K}) \\ \mathbf{K}_{ij} &= \eta^2 \exp(-\rho^2 \mathbf{D}_{ij}) \end{aligned}$$

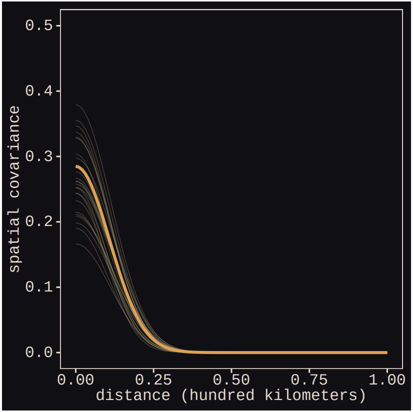
Sensitivity checks

Results are robust to

- year
- interaction effects with population
- inclusion of household size
- spatial autocorrelation in regional effects:

$$
\begin{aligned}
o_i, d_j &\sim \mathrm{MVNormal}(0, \mathbf{K}) \\
\mathbf{K}_{ij} &= \eta^2 \exp(-\rho^2 \mathbf{D}_{ij})
\end{aligned}
$$

Modest spatial autocorrelation



Conclusions

Flexibel and powerful Bayesian multilevel gravity model:

- housing structure asymmetric impact on migration
  - positive on push/negative on pull
  - push factor large in large cities
- impact social renting smaller than homeowership (Boyle, 1998)
  - social housing is like a different ball game
- tight housing market

Now what?

- model performance is quite good
  - out-of-sample prediction
  - long-distance migration (dyad effects)

Supplementary materials

Paper, presentation, data and code can be retrieved from the project's GitHub page:

https://github.com/Thdegraaff/migration_gravity

Thank you!

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.

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

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

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.

References ii

Head, K. and T. Mayer (2014). "Gravity equations: Workhorse, toolkit, and cookbook". In: *Handbook of international economics*. Vol. 4. Elsevier, pp. 131–195.

Koster, J. M. and G. Leckie (2014). "Food sharing networks in lowland Nicaragua: an application of the social relations model to count data". In: *Social Networks* 38, pp. 100–110.

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

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 Europes unemployment: a non-technical paper". In: *Homeownership and the labour Market in Europe*.

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