TI 2016-062/VIII
Tinbergen Institute Discussion Paper

Ethnic Drift and White Flight: A Gravity Model of Neighborhood Formation

Jessie Bakens[a,c]
Raymond Florax[b,c,d]
Peter Mulder[c]

[a] *OIS, Municipality of Amsterdam, the Netherlands;*
[b] *Purdue University, United States;*
[c] *Faculty of Economics, VU University Amsterdam, the Netherlands;*
[d] *Tinbergen Institute, the Netherlands.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Ethnic Drift and White Flight: A Gravity Model of Neighborhood Formation*

Jessie Bakens [a,c,†], Raymond J.G.M. Florax [b,c,d] and Peter Mulder [c]

[a] *Research, Information and Statistics (OIS), Municipality of Amsterdam, Amsterdam, The Netherlands*
[b] *Department of Agricultural Economics, Purdue University, West Lafayette, IN, USA*
[c] *Department of Spatial Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
[d] *Tinbergen Institute, Amsterdam, The Netherlands*

August 16, 2016

Abstract

Ethnicity has become an increasingly important factor in neighborhood formation in many developed economies. We specify a gravity model for neighborhoods to assess the role of ethnicity in intra-urban residential relocations. Migration patterns of different ethnic groups are hypothesized to depend on bilateral socioeconomic, demographic and ethnic differences between origin and destination neighborhoods. We account for heterogeneous and interdependent location preferences of natives and several immigrant groups. In addition, we incorporate friction measures of ethnic population shares and a diversity indicator to allow for nonlinear and asymmetric effects of the population composition on ethnic sorting and spatial clustering. We utilize a unique micro data set of place–to–place migrants across neighborhoods in the urban agglomerations of Amsterdam and The Hague, in The Netherlands. Our results provide evidence of ethnic drift leading to clustering of ethnic minority groups and "white flight" of native Dutch residents. Taken together, our findings suggest a preference for living among people of one's own ethnic group, but in a sufficiently diverse neighborhood. We discuss ways to extend and apply our gravity approach to further analyze intra-urban residential relocation flows.

Keywords: neighborhood formation, ethnicity, diversity, immigrants, gravity model
JEL-classification: C21, F22, J15, J61, R23

*We have profited substantially from comments and suggestions of Henri L.F. de Groot and Brigitte S. Waldorf. Sergio Rey and other colleagues have contributed through discussions of previous versions of this paper at various seminars and conferences. The first author notes that the views expressed are those of the author and should not be attributed to the Department for Research, Information and Statistics of the municipality of Amsterdam.

† Corresponding author. Tel. +31–20–2510443. E-mail: j.bakens@amsterdam.nl.

1 Introduction

Societies in most developed countries are becoming increasingly ethnically diverse. The spatial distribution of ethnic minority groups is, however, frequently all but random: the largest concentrations of ethnic groups are usually found in (large) cities, and within cities segregation by ethnicity often defines the invisible lines that create neighborhoods (Borjas, 1995). These spatial patterns are the result of individual residential mobility choices made within the context of a pre-set spatial structure. This observation is a well known stylized fact in the literature on aggregate migration patterns at the regional level (Mueser, 1989; Rogers and Raymer, 1998; Peeters, 2012). However, this is equally true at the city level. Within the spatial structure of a city, decisions with respect to neighborhood choice and housing are interrelated: the demand for housing materializes in the context of endogenous and contextual neighborhood effects (Waldorf, 1993; Brock and Durlauf, 2001; Ioannides and Zabel, 2008). Typically, households make a nested choice of residential location by choosing an (urban) region first, and then a neighborhood based on a comparison of alternatives (McFadden, 1978). As a matter of fact, most migrant flows actually occur at a disaggregated spatial scale; people choose a housing property within the same neighborhood or between different neighborhoods within the same urban area (Cadwallader, 1992).

Research on (ethnic) relocation dynamics at the neighborhood level includes studies on ethnic sorting and tipping of neighborhoods through "white flight" (Schelling, 1971; Becker and Murphy, 2000; Card et al., 2008), racial housing price differentials (Chambers, 1992; Kiel and Zabel, 1996; Myers, 2004), behavioral models of household location under influence of race and social interaction (Gabriel and Rosenthal, 1989; Bayer et al., 2004; Ioannides and Zabel, 2008; Gabriel and Painter, 2012), neighborhood gentrification processes (McKinnish et al., 2010; Ellen and O'Regan, 2011), and the potential negative effects of segregation on economic outcomes for individuals (Cutler and Glaeser, 1997; Kling et al., 2007). Most of these studies focus on either inmigration or outmigration of people, taking into account characteristics, such as ethnic population shares and housing prices, of either the origin or the destination neighborhood.

In contrast, this paper provides an empirical analysis of bilateral migration patterns of different ethnic groups at the neighborhood level as a function of bilateral socioeconomic, demographic and ethnic differences between origin and destination neighborhoods. More specifically, we aim to identify to which extent ethnic heterogeneity across space affects both the size and ethnic composition of place–to–place migration flows within cities, given the intra-urban spatial heterogeneity that originates from differences in price, quality and accessibility of housing and a range of other socioeconomic and demographic characteristics (Waldorf, 1993). Arguably, given individual preferences for neighborhoods with relatively large populations of one's own ethnicity, socioeconomic neighborhood characteristics and the degree of neighborhood segregation impact location decisions of both the immigrant population as well as natives (Schelling, 1971; Becker and Murphy, 2000; Krysan and Farley, 2002; Vigdor, 2003; Bolt et al., 2008; Card et al., 2008; Saiz and Wachter, 2011; Bayer et al., 2014). Therefore, in our analysis we explicitly take into account that preferences of natives and immigrants for the own ethnic group are heterogeneous but interdependent, and we explicitly incorporate the link between individual choices and associated changes at the neighborhood level in a suitably specified gravity equation.

Gravity models are the standard workhorse for explaining international or interregional migration flows, but they are seldom used to explain residential flows at lower spatial scales.[1] Similarly to

[1] An exception is Saiz and Wachter (2011), who use a gravity pull measure to instrument the endogeneity of

Newtons original gravity equation in physics, we hypothesize that the magnitude of migration flows between two neighborhoods is positively related to their respective weights in terms of population and impeded negatively by their distance (Greenwood, 1975; Karemera et al., 2000; Lewer and Van den Berg, 2008; Mayda, 2010; Peeters, 2012; Falck et al., 2012; Ortega and Peri, 2013). We extend the concept of distance to encompass not only physical distance but also ethnic differences between areas of origin and destination. The underlying idea is that ethnic differences may act as additional barriers to migration because they increase the cost of moving goods or people in addition to the costs of bridging physical distance (Karemera et al., 2000; Guiso et al., 2009; Grosjean, 2011; Caragliu et al., 2013). Indeed, Falck et al. (2012) show that similarity in cultural identity between German regions affects the size of mover flows between these regions.

Likewise, in our gravity approach we assess the extent to which ethnic friction between neighborhoods in terms of an ethnic group's own population share, and diversity among other ethnic groups, impacts mover flows of this ethnic group between these neighborhoods. We allow for nonlinearity and asymmetry in the association between the ethnic friction measure and the size of the bilateral mover flows at the neighborhood level (Schelling, 1971) through the use of a higher-order polynomial of the ethnic friction measure for positive and negative differences between neighborhoods in terms of ethnic concentration. This way of measuring the role of ethnicity in neighborhood dynamics is rather unique. In addition to providing a simple way of incorporating the potential asymmetric effect of ethnic distance on the size of mover flows, this method uses the information available in the data on the ethnic connectedness between neighborhoods within cities. Larger mover flows between neighborhoods based on the ethnic friction measure indicate a higher ethnic connectedness (or lower ethnic barriers for mover flows) between those neighborhoods.

We estimate the gravity equation using a modified version of the Poisson maximum likelihood (PML) estimator rather than applying ordinary least squares (OLS) to a log-linearized version of the nonlinear relationship between distance and the size of the origin and destination populations. Santos Silva and Tenreyro (2006) have shown that log-linearized gravity models only comply with the homoskedasticity assumption under very specific conditions and that the OLS estimator is therefore in general inconsistent and inefficient. An additional advantage of the Poisson approach is that it offers a natural and consistent way of dealing with zero-counts of the dependent variable.

We apply our method to a unique micro data set that encompasses information of place–to–place migrants between neighborhoods in the larger agglomerations of the Dutch cities Amsterdam and The Hague, for the period 2004–2008. Amsterdam and The Hague are respectively the largest and third-largest city in The Netherlands, as well as the capital city and the seat of the national government. The population of both cities comprises almost 50 percent ethnic minorities, with The Hague being considerably more segregated than Amsterdam. The data include information on individual ethnic, socioeconomic and demographic characteristics of all residents, their location, individual housing prices, and a range of neighborhood characteristics. The urban agglomerations are defined in such a way that they largely encompass the labor market area in order to exclude people moving for job market reasons. The cities are analyzed separately as a system of neighborhoods, and the mover flows are aggregated to a cross-section over the period from 2004 to 2008.

Focusing on the mover flows of the four largest ethnic groups (Dutch, Turks, Moroccans, and Carribeans), our results clearly show that in both cities ethnic drift leads to clustering of ethnic

immigrant location patterns. Waldorf (1993) develops a gravity model to measure segregation incorporating additional relocation barriers such as income differences. Bolt et al. (2008) and Schaake et al. (2014) study mover flows between neighborhoods, but they do not use a gravity model of neighborhoods.

minorities as well as to a "white flight" of native Dutch residents, with the ethnic drift being relatively stronger than the white flight. Except for one minority group, we also find that mover flows of all ethnic groups are higher into more diverse neighborhoods. We find clear evidence that origin–destination differences in the concentration of the own minority group (i.e., ethnic distance) is asymmetric and nonlinearly related to the size of mover flows of that specific ethnic group. Taken together these findings suggest a preference for living with the own ethnic group in an otherwise diverse neighborhood. A generalization of our results would indicate that ethnic groups prefer some level of local dominance, although not all ethnic groups can enjoy numerical superiority (Schelling, 1971).

The next section describes the specification of the gravity equation and the estimation procedure for count data models. In Section 3 we give a detailed description of the neighborhood and flow characteristics of Amsterdam and The Hague, and we describe the data used in our analysis. Section 4 gives the results of the estimation of different ethnic neighborhood mover flow models for Amsterdam and The Hague. Several robustness checks are presented as well. Section 5 concludes and provides ideas to further explore the use of a gravity approach to analyze intra-urban residential relocation flows.

2 Econometric Model

To identify the relevance of ethnic drift and white flight in relocation decisions across neighborhoods, we start from a traditional gravity model equation. In addition to the usual geographical distance measure, we include a measure pertaining to ethnic distance between neighborhoods. This friction measure is operationalized as the difference in population share of an ethnic group between the neighborhoods of origin and destination:

$$\Delta s_{ij} = s_j - s_i, \tag{1}$$

where s is the population share of an ethnic group, and i and j are the origin and destination neighborhoods belonging to a set of n disjoint neighborhoods of a metropolitan area. We use the Δ operator to define the difference in attribute values between the origin and destination neighborhoods. For ease of interpretation, equation (1) is defined as the difference between the neighborhood of destination and the neighborhood of origin. A positive value, then, corresponds to a move to a neighborhood with a higher attribute value, i.e., a higher population share of an ethnic group. Obviously, the order of the terms does not materially affect the analysis.

Nonlinearity in the association between the ethnic friction measure and the size of mover flows can be achieved by including a higher-order polynomial of the ethnic friction measure. Asymmetry can be incorporated by relaxing the assumption that the effect is symmetric for mover flows into neighborhoods with higher ethnic shares and neighborhoods with lower ethnic shares. This leads to a flexible functional form in which a linear and quadratic terms are split into Δs_{ij+} and Δs_{ij-}, and Δs^2_{ij+} and Δs^2_{ij-}, respectively, depending on whether $\Delta s_{ij} \geq 0$ or $\Delta s_{ij} < 0$.

We include a friction measure of neighborhood diversity, ΔH_{ij}. This index is calculated for a specific ethnic group and constructed by including all ethnic groups in the neighborhood except for the specific ethnic group for which the index is calculated; for both the origin and the destination neighborhood. The index measures whether the composition of a neighborhood is diverse in terms of ethnic groups other than the ethnic group under consideration. The diversity index for neighborhood i is calculated as:

$$H_i^e = 1 - \sum_{e=1}^{m-1} (s_i^e)^2 \,, \tag{2}$$

where s_i^e is the share of people from ethnic group e among the residents of neighborhood i, with m for the total number of ethnic groups. An index value of 0 indicates that all residents in the neighborhood are of the same ethnic group, whereas a value of $1 - 1/(m-1)$ is the maximum value of the index, which indicates that all ethnic groups are equally large.

Inter-neighborhood flows for a specific ethnic group are subsequently modeled as:

$$\begin{aligned} f_{ij} = F\big(&\beta_0 + \beta_1 \Delta s_{ij+} + \beta_2 \Delta s_{ij-} + \beta_3 \Delta s_{ij+}^2 + \beta_4 \Delta s_{ij-}^2 + \beta_5 \Delta H_{ij} \\ &+ \Delta x'_{ij}\beta_6 + \beta_7 d_{ij} + \delta_i + \delta_j\big), \ \forall i \neq j, \end{aligned} \tag{3}$$

where f_{ij} is the size of the mover flow between neighborhoods i and j, $\Delta x'_{ij}$ is a set of control variables capturing socioeconomic and demographic differences between neighborhoods, d_{ij} is the geographical distance between neighborhoods, δ_i and δ_j are unobserved origin and destination neighborhood effects, respectively, and F is an as of yet unspecified functional form.

The control variables in the regression analysis include the distance between the origin and destination neighborhoods, the number of dwellings in the destination neighborhood, and the growth in housing stock in the destination neighborhood. These variables are the basic explanatory variables regarding size and distance in a gravity equation. The economic neighborhood characteristics included are the mean per capita neighborhood income and the share of social rent and owner-occupied housing in the total housing stock. To control for life-cycle-related residential relocations, the share of children in the total population of the neighborhood is included, which serves as a proxy for whether the neighborhood is family oriented and whether the houses in the neighborhood are suitable for families. We control for unobserved neighborhood characteristics of the origin and destination neighborhoods by including origin and destination fixed effects at the district level. The neighborhoods of Amsterdam and The Hague are aggregated into 15 district levels, and the neighborhoods outside of Amsterdam and The Hague that belong to the metropolitan area of the cities are aggregated into one district for each city.

Equation (3) is given for inter-neighborhood flows, which is the case where $i \neq j$. The case where $i = j$ concerns intra-neighborhood mover flows. Following LeSage and Pace (2008), intra-neighborhood mover flows can be included in the model using a simple re-parameterization:

$$\begin{aligned} f_{ij} = F\big(&\beta_0 + \beta_1 \Delta s_{ij+} + \beta_2 \Delta s_{ij-} + \beta_3 \Delta s_{ij+}^2 + \beta_4 \Delta s_{ij-}^2 + \beta_5 \Delta H_{ij} \\ &+ \Delta x'_{ij}\beta_6 + \beta_7 d_{ij} + x'_i \beta_8 + \delta_i + \delta_j\big), \end{aligned} \tag{4}$$

and creating a block-diagonal design matrix by defining $\Delta s_{ij-} = \Delta s_{ij+} = \Delta s_{ij+}^2 = \Delta s_{ij-}^2 = \Delta x'_{ij} = d_{ij} \equiv 0$ if $i = j$, and $x'_i \equiv 0$ for all $i \neq j$, where x'_i is a vector of neighborhood characteristics. The sub-model for intra-neighborhood flows depends only on the levels of the characteristics in each neighborhood, while the differences are by definition zero. For the sub-model for inter-neighborhood flows, only the differences between the characteristics of the neighborhoods, and hence not the level of the characteristics, determine the size of the flows.

The gravity model for origin–destination flows is originally a nonlinear relationship expressing the magnitude of mover flows as a multiplicative function of the origin and destination populations, taking into account the friction of geographical distance. A straightforward log-linearization makes it feasible to estimate the linearized version of the model with ordinary least squares (OLS). However, the validity of log-linearization and estimation with OLS crucially depends on the homoskedasticity of the error terms. Santos Silva and Tenreyro (2006) show that log-linearized gravity models only comply with the homoskedasticity assumption under very specific conditions and that the OLS estimator is therefore in general inconsistent and inefficient. They propose that the multiplicative form be estimated directly with a Poisson pseudo-maximum likelihood (PML) estimator used in count data models. Through simulations and an empirical example, they show that standard approaches lead to substantial bias in the estimated coefficients. An additional advantage of the Poisson PML approach is that it offers a natural and consistent way of dealing with zero-counts of the dependent variable. Although different ways of incorporating zero-flows in a log-linearized version of the gravity model have been suggested, these solutions remain unsatisfactory, especially in cases with many zero-flows (Linders and De Groot, 2006; Burger et al., 2009).[2]

We therefore estimate the gravity equation using the Poisson PML estimator rather than applying OLS to appropriately log-linearized versions of equations (3) and (4). In general terms, the structural equation of the regression model shows that the probability of observing a realization y_i for a random (count) variable Y is given by the Poisson probability function (Long, 1997; Winkelmann, 2008):

$$\Pr\left(Y = y_i\right) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!},\ \text{with } \lambda_i \in \mathbb{R}^+,\ y_i \in \mathbb{N}_0, \tag{5}$$

where λ_i is the mean of the Poisson distributed variable for observation i. The expected value λ_i is always positive and therefore often modeled as conditional on a set of explanatory variables using the logarithmic link function, $\log\left(\mathrm{E}(y_i)\right) = x_i'\beta$, and therefore $\mathrm{E}(y_i) = \lambda_i = \exp\left(x_i'\beta\right)$.

The defining characteristic of the Poisson distribution is equidispersion, which implies that the conditional mean and the conditional variance are equal, $\mathrm{E}\left(y_i|x_i\right) = \mathrm{Var}\left(y_i|x_i\right) = \lambda_i$. The Poisson distribution entails that the occurrences of events are independent and that the rate at which an event occurs is constant, but in most cases, even conditionally, socioeconomic data will not be Poisson distributed due to contagion and unobserved heterogeneity. In practice, socioeconomic data are often overdispersed, as is the case with our data, in which case a Poisson model underestimates the magnitude of the dispersion as well as the standard errors (Long, 1997; Long and Freese, 2006). We therefore prefer to estimate a negative binomial (NB) model that allows for overdispersion by incorporating unobserved heterogeneity through an unobserved unit-specific error:

$$\lambda_i = \exp\left(x_i'\beta + \varepsilon_i\right). \tag{6}$$

The typical assumption for the error term is that it follows a gamma distribution with mean unity, so that the expected value λ_i for the negative binomial distribution is the same as for the Poisson distribution. In contrast to the Poisson model, the conditional variance for the NB model is assumed to follow the function $\mathrm{Var}\left(y_i|x_i\right) = \lambda_i + \alpha\lambda_i^2$, which, for a dispersion parameter $\alpha > 1$,

[2] Santos Silva and Tenreyro (2006) convincingly claim that heteroskedasticity rather than truncation of the data is the more severe problem in estimating gravity model equations. Extended simulations in Santos Silva and Tenreyro (2011) show that the Poisson PML estimator also performs satisfactorily in cases where the proportion of zero-flows in the sample is large.

clearly shows the allowance for overdispersion.[3]

So far we have ignored differences in the population at risk across observations. In our empirical analysis, the size of the mover flow obviously depends on the size of the different ethnic groups in the neighborhoods. In order to correct for these exposure differences, the count variable can be scaled by the size of the population at risk, but this scaling implies that instead of a count variable we end up with a fraction to be explained.[4] In the negative binomial model, an exposure variable on the right-hand side of equation (6) can be included to control for the potential population that can be observed to move out of a neighborhood:

$$\begin{aligned} \lambda_i &= S_i \exp\left(x_i' \beta + \varepsilon_i\right) \\ &= \exp\left(x_i' \beta + \ln S_i + \varepsilon_i\right), \end{aligned} \tag{7}$$

where S_i is the population at risk, of which the coefficient is restricted to unity.[5]

Excess zero's for the dependent variable can also be the cause of overdispersion in the data. Although an NB model can deal with zero counts, it is assumed in the model that a count of zero is generated by the same process as a count of one, two, etc. There are cases in which a zero count can stem from two different processes, one in which the probability of a zero count is 1, and one in which the probability of a zero count is smaller than 1, in which case the observation has a probability of being a positive count. Count models of this type, Zero-Inflated Negative Binomial Models (ZINB), are estimated using probabilities of being in either of the two groups. The expected count in an ZINB model is given by:

$$E\left(y_i \mid X_i, Z_i\right) = \left[0 \times \psi_i\right] + \left[\lambda_i \times \left(1 - \psi_i\right)\right] = \lambda_i \left(1 - \psi_i\right), \tag{8}$$

where ψ_i is the probability of being in the definite-zero group, $1 - \psi_i$ is the probability of not being in the definite-zero group, and Z_i are the inflation variables determining the zero-generating process of the definite-zero group.

Finally, the main method for evaluating the impact of ethnic distance on the estimated regression is to calculate the predicted count of y_i using fixed values of the ethnic friction measure. We calculate the mean predicted value[6] of the ethnic friction measures and keep the values of the covariates as observed in the dataset using:

$$\lambda = E\left[y\right] = \exp\left(X_i \beta\right). \tag{9}$$

Following Winkelmann (2008), the standard errors of these predicted counts are calculated using

[3] In the literature, this is known as the 'NB2' or 'Negbin II' version of the negative binomial model. The NB1 model utilizes $\text{Var}\left(y_i|x_i\right) = \lambda_i + \alpha\lambda_i$ as the specification for the conditional variance (Cameron and Trivedi, 2009, 2013). Blackburn (2014) provides an overview of the performance of the Poisson, NB1, and NB2 estimators.

[4] This is feasible using, for instance, the fractional logit estimator developed by Papke and Wooldridge (1996).

[5] In an operational sense, S_i can be replaced by $S_i + c$ if the population variable is not strictly positive. The adjustment factor c is a sensibly defined constant (Long, 1997). We use $c = 1$.

[6] The predicted count at specific values of Δs_{ij} can be calculated at the mean of all observed values, i.e., the predicted values at the mean, but also by finding the mean after calculating the predicted value for each observation, i.e., the mean predicted value. Because the mean of most variables in our dataset is zero and we include dummy variables for which the mean does not have a sensible interpretation, we calculate the mean predicted values.

the variance matrix of $\hat{\beta}$ according to the maximum likelihood Negbin II estimator:

$$\operatorname{Var}\left(\hat{\beta}\right) = \left(\sum_{i=1}^{n} \frac{\exp\left(x_i'\beta\right)}{1 + \frac{1}{\alpha}\exp\left(x_i'\beta\right)} x_i x_i'\right)^{-1}. \tag{10}$$

We use marginal effects to compare the sizes of the different estimated coefficients in the model. These can be calculated for different observed values, i.e., scenarios, using:

$$\frac{\partial E\left[y\right]}{\partial x_l} = \lambda \beta_l, \tag{11}$$

which refers to any of the explanatory variables x_l.

3 Data

We utilize a unique micro data set that encompasses information of place–to–place migrants between neighborhoods in the larger agglomerations of the Dutch cities of Amsterdam and The Hague, for the period 2004 to 2008. Together these two agglomerations house about one-third of the immigrant population in The Netherlands. More precisely, in 2004, 19 percent and 10 percent of the immigrants in The Netherlands lived in the Metropolitan Areas (MA) of Amsterdam and The Hague, respectively. The Hague is considered to be relatively segregated while Amsterdam's neighborhoods are more ethnically diverse. Amsterdam and The Hague are geographically proximate, but they constitute different metropolitan areas. For the Metropolitan Area of Amsterdam we use the NUTS3-regions Greater Amsterdam, Haarlem agglomeration, and Zaanstreek, as well as the city of Almere.[7] For the Metropolitan Area of The Hague we use the NUTS3-regions Greater The Hague and Delft&Westland. Figure 1 shows the areas and Table A.1 lists the municipalities included in the analysis. Neighborhoods are the smallest statistical spatial units used by Statistics Netherlands and consist of, roughly, a couple of thousand dwellings.

⟨ Figure 1 about here ⟩

Because each single neighborhood in a gravity model has equal weight in the analysis, including all of the individual neighborhoods of the MAs would shift the weight of the analysis away from residential relocations within the cities of Amsterdam and The Hague. We therefore aggregate the neighborhoods outside of the city boundaries into a single large "neighborhood" (see Table A.1). This allows us to include suburbanization patterns without focusing on intra-suburban relocations. We exclude neighborhoods with less than 25 residential dwellings. The analysis is performed on 99 neighborhoods for the metropolitan area of Amsterdam (of which 93 are in the city of Amsterdam) and on 113 neighborhoods for the metropolitan area of The Hague (of which 106 are in the city of The Hague). All other relocations outside of the above defined system are not included in the analysis. We use a closed system of flows so that all possible neighborhood flows within the system are included in the analysis, even if these are zero.

A gravity model predicts that most mover flows occur between neighborhoods that are geographically close. Of all relocations within the municipality boundaries of Amsterdam between

[7] Our definition of the Metropolitan Area of Amsterdam differs from the official Metropolitan Region of Amsterdam (MRA).

2004 and 2008, 54 percent are between neighborhoods within these city boundaries. For The Hague, this number is 56 percent. By including the neighborhoods of the MAs, we can cover about 75 percent of all moves into or out of neighborhoods in both cities. Thus, 25 percent of the relocations in Amsterdam and The Hague are from or to a neighborhood outside of the area included in our analysis. Of all the relocations in our analysis, 83 percent are inter-neighborhood relocations in Amsterdam, and 86 percent are inter-neighborhood relocations in The Hague. Because we aggregate the neighborhoods of the MAs into large neighborhoods, the inter-neighborhood flows in these areas become much lower by construction, namely 51 percent for the Amsterdam MA and 62 percent for the The Hague MA.

Intra-urban mover flows are calculated from the administrative municipality data (GBA) from Statistics Netherlands, which contain the residential location of all residents of The Netherlands as well as changes in location. We select all residents who moved between any of the neighborhoods in our study area between 2004 and 2008.[8] Because our analysis focuses on mover flows at the neighborhood level, all neighborhood variables are totals or means of individuals.[9]

Each flow from one neighborhood to another is the total number of residential moves between 2004 and 2008. This five-year period is selected because during this time span mover flows are in general sufficiently small to not substantially change the ethnic composition of neighborhoods. In this way, we reduce as much as possible the dynamic processes in our data and the possibility that mover flows are compounded by changing neighborhood characteristics that have themselves resulted from the size of the mover flows that we are trying to explain.

We focus on the largest ethnic groups in the MAs of Amsterdam and The Hague: native Dutch, Turks, Moroccans, and Caribbeans.[10] Ethnicity is based on the country of birth of the parents. The country of birth of the mother is the leading determinant of ethnicity, unless she is born in the Netherlands, in which case the country of birth of the father is the leading determinant. Table 1 shows the ethnic population composition of the two cities and MAs. The four ethnic groups in our analysis constitute 81 percent of the total population of the MA of Amsterdam and 82 percent of the total population of the MA of The Hague.

⟨ Table 1 about here ⟩

The location patterns of non-native groups and the native Dutch differ. Ethnic minority groups are overrepresented within the city boundaries of Amsterdam and The Hague, while the native population is relatively overrepresented outside the cities, i.e., the suburbs. The standard dissimilarity index in Table 1 shows that the neighborhoods of Amsterdam are somewhat less segregated than the neighborhoods of The Hague.[11] When we calculate the index over the whole MA for either

[8] Individuals from institutional households such as penitentiaries or retirement homes, individuals who have resided at one location less than 180 days, and individuals who are registered at one dwelling with more than 100 other individuals are excluded from the analysis.

[9] Although households rather than individuals make the residential location choices, it are the aggregate characteristics of all individuals, rather than household heads, that determine the neighborhood's characteristics.

[10] Throughout the study, we refer to the Surinamese, Antilleans, and Arubans as Caribbeans.

[11] We calculate the dissimilarity index as follows:

$$G_c^e = \frac{\sum_{i=1}^{n} T_i \left| s_i^e - s_c^e \right|}{2 T_c s_c^e \left(1 - s_c^e\right)}, \tag{12}$$

where G_c^e is the dissimilarity index of ethnic group e in city c, T_i is the total population in neighborhood i, s_i^e is the share of ethnic group e in neighborhood i, s_c^e is the share of ethnic group e in city c, and T_c is the total population city, segregation increases in general. According to Massey and Denton (1988), segregation is low, with an index below 0.3, and moderate, with an index between 0.3 and 0.6. This categorizes the neighborhoods of Amsterdam and The Hague as mostly moderately segregated as far as Moroccans, Turks, Caribbeans, and Dutch are concerned. Compared to, for example, the segregation of blacks in US cities, segregation is rather low in Amsterdam and The Hague. Cutler and Glaeser (1997), Cutler et al. (1999), Vigdor (2003), and Bayer et al. (2014) calculate dissimilarity indices with mean values of about 56 percent for blacks in the US.

⟨ Figures 2 and 3 about here ⟩

Native Dutch are the majority ethnic group in most neighborhoods, as shown in Figure 2. Although a substantial share of the neighborhoods has a majority population of foreigners, there are no neighborhoods with a majority of Moroccans, Turks, or Caribbeans. The highest clustering of any of these three groups is around 30 percent of the population in a neighborhood. When we deal with multiple ethnic groups that differ greatly in size, clustered and mixed neighborhoods are not mutually exclusive. Neighborhoods in The Hague whose population are about 20 percent Caribbean represent relatively high clustering for Caribbeans, but the overall composition of such neighborhoods predominantly shows that they include 80 percent other ethnic groups. Because we are interested in the role of the own ethnic group as a location choice factor, we focus on the former while the latter is controlled for by the diversity index of the other ethnic groups in the neighborhood.

Ethnic neighborhood segregation is often closely related to the economic characteristics of a neighborhood. The mean incomes of the different ethnic groups in this study are shown in the lower part of Table 1. The economic position of the different ethnic groups in this study is representative of the position of ethnic minorities in many countries: natives have a higher per capita income than ethnic minority groups. For the ethnic group longest present in The Netherlands, the Caribbeans, the income gap between this group and the native population is lower than the gap between Turks or Moroccans and the native population.[12] Another explanation for the observed pattern of segregation and income clustering is partly related to the role of housing markets. Housing market policies determine the owner structure of dwellings in a neighborhood. In this case, the city of Amsterdam chooses to mix types of housing more extensively than the city of The Hague, i.e., social housing is less spatially clustered in Amsterdam than in The Hague, as shown in Figure 3. Assuming that the lowest income groups are relatively dependent on social housing, the general distribution of these groups should be more equal in Amsterdam than in The Hague, which is what we find in the descriptive statistics.

⟨ Tables 2 and 3 about here ⟩

Given the closed system of the gravity model, all zero-flows are included in the regression, and the first lines of Table 2 show that for all non-Dutch ethnic groups, over half of the possible mover flows between neighborhoods have a zero count. Our analysis also includes a few neighborhoods, in both Amsterdam and The Hague that have neither Moroccans nor Turks; the mover flows of

in the city. The index is calculated for each ethnic group against the rest of the population. If the index is close to 0, there is no segregation, while an index close to 1 indicates a fully segregated distribution (Waldorf, 1993).

[12] As income is measured per capita, ethnic differences in family sizes impact the results because ethnic groups with many children have lower per capita incomes.

Moroccans and Turks from these neighborhoods are zero by definition. This is the case for 5 percent of the Moroccan flow and 3 percent of the Turkish flow in Amsterdam and 7 percent and 4 percent in The Hague, respectively. The mean size of the non-zero mover flows is given in Table 2. We include a "rest" group that contains the total of residential moves between neighborhoods for all other ethnic groups, but we generally do not report results for this rest group.

The explanatory variables used in the regression analysis are given in Table 3. January 1, 2004 is used as the base date for measuring the individual neighborhood characteristics. We therefore explain mover flows between 2004 and 2008 by the differentials from neighborhood characteristics in 2004. The demographic characteristics of the residents (date of birth, gender, country of birth, country of birth of the mother, and country of birth of the father) are taken from the GBA (Gemeentelijke Basis Administratie) dataset. Yearly income data are taken from the Sociaal Statistisch Bestand (SSB) from Statistics Netherlands and include all sources of individual income, such as labor income, social welfare, and unemployment benefits. Because the total income of all neighborhood residents is averaged over the total population of the neighborhood, we obtain a per capita income that is lower than the average household income would be. For the aggregated neighborhood characteristics, individuals from institutional households are excluded. Information on the ownership composition of the housing stock in a neighborhood is taken from the Housing register of Statistics Netherlands. Because these data are unavailable for 2004, the owner composition in 2006 (the first available year) is used instead.

Table 3 gives the mean and standard deviations of the variables that are given as the friction measures used in the regressions, whereas the minimum and maximum values are given as neighborhood stock values. The latter gives insight into the base levels that result in the values of the friction measures. Because the setup of the gravity model intrinsically warrants perfectly symmetrical distribution of the variables if they are expressed as difference from the mean (except for increases in distance and the number of dwellings), the mean of all the variables equals zero, while the range is the difference between the minimum and maximum values. For example, for Amsterdam, the ethnic friction measure of s_{ij} (which we label Δs_{ij}) for the Dutch population falls between -0.75 $(0.14 - 0.89)$ and $+0.75$ $(0.89 - 0.14)$.

4 Results

Our gravity model of intra-urban mover flows is defined for each ethnic group individually, as described in Section 2, but we estimate the model for all groups together. The ethnic distance and distance in terms of diversity variables are interacted with an ethnic group-specific dummy. The other covariates are homogeneously estimated for all ethnic groups. We thus impose on the model that there are no differences between groups in terms of the effects of these variables on the size of mover flows. This assumption has the main advantage that results are comparable across the different ethnic groups. The results of the estimated negative binomial regressions cannot be easily interpreted because the underlying model specification is nonlinear, especially for the effect of ethnic peers. We therefore first discuss the direction of the results for the control variables. We then discuss the results of the ethnic friction measures by focusing on the predicted count of the mover flows. Lastly, we compare the size of the effects of all explanatory variables by calculating their marginal effects.

⟨ Table 4 about here ⟩

The results of the base specification of equation (7) are given in Table 4. The estimated α shows overdispersion in the data, which indicates that the use of a negative binomial model is warranted. The rest group is not reported in the results and serves as the reference group. We include ethnic fixed effects to account for ethnic group-specific unobserved characteristics that impact mobility or the tendency to move.[13] Only in Amsterdam are all the ethnic dummy variables statistically significant, indicating differences in mobility relative to the rest group.

First, we find in accordance with the gravity model that the physical distance between neighborhoods impacts the size of mover flows negatively, while mover flows are higher into larger and growing neighborhoods. However, the implied extra costs of migrating if the distance between areas increases, are likely to be marginal, and the results might also signal, for example, that social networks can be maintained more easily over shorter distances.

As regards economic neighborhood characteristics, we find a negative correlation between income differences and mover flows, indicating that mover flows from low-income neighborhoods into high-income neighborhoods and vice versa are small. Most mover flows are between neighborhoods that are alike in terms of average income. Nonetheless, it is to be noted that income differentials do not play a significant role in explaining the size of mover flows in Amsterdam although the results between Amsterdam and The Hague are broadly comparable in terms of the directions of all of the other effects. Our data and analysis cannot provide a conclusive reason for these differences, but we hypothesize that the income distribution and the owner structure of the housing stock in a neighborhood are two sides of the same coin because the latter influences income distribution. If the owner structure of houses is used as a policy instrument to mix incomes within neighborhoods—which is the case in Amsterdam—one might expect that the owner structure emerges as a more important determinant for mover flows than income per se.[14]

The friction measure of the share of children in a neighborhood is positive. This indicates that the life-cycle related characteristics of neighborhoods are related to the size of mover flows. Higher mover flows are observed into neighborhoods with a higher share of children. Finally, the results for the friction measure of the diversity index show that in both Amsterdam and The Hague mover flows are higher into neighborhoods that are overall more diverse in terms of ethnic groups other than the own. Mover flows of especially the Dutch and to a lesser extent Moroccans and Turks are higher if the ethnic population composition in a neighborhood is more diverse in terms of the other ethnic groups. The mover flows of Caribbeans are higher into neighborhoods where the population composition in terms of other ethnic groups is more homogeneous. This might indicate that neighborhoods with higher shares of Dutch, which form the most homogeneous neighborhoods, are attractive for Caribbean mover flows. However, the interpretation of these results also depends on the results found for the friction measure of a specific ethnic group, which we will discuss below.

⟨ Figure 4 about here ⟩

The predicted mover flows over the distribution of the ethnic friction measures are given in Figure 4. Using equation (9), Figure 4 is calculated with all other variables as observed.[15] We

[13] Van Ham and Clark (2009) and Schaake et al. (2010) have looked into residential mobility. Differences between the mobility of ethnic groups seem to be relevant vis-à-vis the presence of the same ethnic group in the neighborhood.

[14] Note that this still does not provide an explanation as to why mover flows in The Hague are higher into neighborhoods with less social rent, while mover flows in Amsterdam are higher into neighborhoods with more social rent and more owner-occupied houses.

[15] The 10th to 90th percentiles of the distributions of the ethnic friction measures are used for this calculation. The results are obtained using Stata's post estimation margins command.

clearly find evidence for the hypotheses that, for all ethnic groups, the distance of the own ethnic group has an asymmetrical and nonlinear correlation with mover flows of that ethnic group. From a gravity point of view, the largest mover flows are expected to be centered around zero between neighborhoods that are alike, i.e., over small ethnic distances. This results in a concave relationship between the ethnic friction measure and the predicted flow. Only for the Caribbeans in The Hague do we observe this concave relationship. A convex relationship between the ethnic differentials and mover flows is apparent for the other groups. All else being equal, the highest ethnic mover flow is predicted into neighborhoods that have a much higher share of that ethnic group. In the opposite direction, the ethnic mover flow is close to zero for that ethnic minority group and relatively small for the Dutch, indicating ethnic drift and white flight into neighborhoods with a substantial population of the own ethnicity.

In Amsterdam and The Hague, the predicted mover flow of Dutch into a neighborhood with a 20 percent larger share of Dutch residents is almost two times higher than the flow into a neighborhood with a 20 percent smaller share of Dutch residents. The mover flows of the ethnic minority groups in Amsterdam into a neighborhood with a 20 percent larger share of these ethnic groups are four to seven times higher than the ethnic flow of these groups into neighborhoods where the share of that ethnic group is 20 percent smaller. In The Hague, this is two times higher. The confidence intervals of the Moroccans and Turks overlap, indicating that the effects do not significantly differ between these groups. We thus find that white flight and ethnic drift lead to spatial clustering in neighborhoods with higher shares of the peer group, but the ethnic drift of minority groups is stronger than the white flight of the Dutch, specifically in Amsterdam. Given the number of Dutch relative to the number of ethnic minorities, Dutch preferences will most likely prevail in determining the overall diversity of neighborhoods. All other things equal, our empirical results point to the fact that ethnic spatial clustering in neighborhoods in Amsterdam and The Hague increases over time.

If we combine these results with the overall pattern that mover flows are higher into more ethnically diverse neighborhoods, we thus find that ethnic drift leads to clustering of ethnic groups in neighborhoods that are diverse in terms of other ethnic groups. These results point to the described tendency of people to live in neighborhoods where their own ethnic group is present, even if this group does not have a numerical majority in the neighborhood. The fact that we find this pattern for almost all groups indicates that the composition of neighborhoods depends on ethnic preferences for the own ethnic group relative to the presence of other ethnic groups. However, as noted earlier, in a city with many minority groups, the relative clustering of groups does not necessarily tell us anything about the overall neighborhood diversity or segregation.

⟨ Table 5 about here ⟩

To determine whether the role of ethnic distance prevails over other explanations such as income, we look at the marginal effects of each variable at different values of the distribution while keeping all the other variables as observed.[16] The results for the marginal effects at different percentiles of the distribution of each variable using equation (11) are given in Table 5. The 50th percentile is zero for all variables measured as differences between the characteristics of neighborhoods. Logically it follows that the difference in the observed values of the 25th percentile and the 75th percentile are mirrored values. The Dutch mover flow at the 25th percentile is into neighborhoods with about

[16] The results are obtained using Stata's post estimation margins command.

15 percent fewer Dutch residents than the neighborhood that was left, and at the 75th percentile the flow is into neighborhoods with 15 percent more Dutch. For the covariates with positive coefficients, the correlation with mover flows increases with distance in a positive direction, i.e., a positive difference between the neighborhoods of destination and origin. The exact opposite is true for the covariates with negative coefficients.

Table 5 shows that distance is in general the most important factor explaining the size of mover flows both in Amsterdam and the Hague. For neighborhoods in Amsterdam that are about 1.4 kilometers apart (the 5th percentile), a one-kilometer increase in distance between neighborhoods is associated with a decrease in the expected size of the mover flows between these neighborhoods of 3.2, which is more than half the standard deviation of the total mover flow. For The Hague, the effect of distance is in general lower; for neighborhoods that are 1.3 kilometers apart, an increase in distance of one kilometer is associated with a decrease in the size of the mover flow of about 2.2, or almost two-thirds of the standard deviation of the total mover flow.

The marginal effects of the ethnic presence are substantial too. A ten-percent increase in the friction measure of the share of Dutch between neighborhoods that have the same share of Dutch (the 50th percentile) increases the Dutch mover flow by 2.5 for both Amsterdam and The Hague. For Caribbeans, a ten-percent increase in the difference of the share of Caribbeans between neighborhoods with the same share increases mover flows by 2 in Amsterdam and by 0.66 in The Hague. For Moroccans and Turks, this increase is lower. The size of the marginal effects for the diversity index is substantial for the Dutch, especially in Amsterdam.

The growth of the housing stock has a constant effect for both Amsterdam and The Hague. This indicates that, irrespective of the extent of the increase in the number of houses in a neighborhood, the incoming mover flow increases proportionally, as is expected if the housing market is in equilibrium and the market clears. The importance of the share of children for explaining the size of mover flows increases substantially for flows into neighborhoods that have relatively high shares of children. If the difference in the share of children increases by 10 percent for neighborhoods that have a difference in this variable at the 75th percentile, mover flows increase by about 2 for Amsterdam and 1 for The Hague. In The Hague, an increase in the income difference by €1000 for neighborhoods that have an income difference of about €5000, decreases mover flows by 0.117. For social rent in Amsterdam, an increase in the share of social rent by 10 percent for neighborhoods that have a difference in the share of social rent of about 22 percent increases mover flows by 1.3. Overall, the marginal effects of ethnic distance and physical distance on estimated mover flows are the largest. Differences between neighborhoods in the share of an ethnic group are more important for the size of mover flows than the distance in ethnic diversity, although the latter is strong, in particular for the Dutch.

Given the role of income in location patterns in general and the differences in income per capita between the different ethnic groups, we further explore the role of income in determining the size of mover flows by performing two robustness checks. The relationship between income and mover flows might not be homogeneous, but differ between ethnic groups. We have therefore estimated the base regression including heterogeneous effects of income frictions for the ethnic groups. However, there are no differences between ethnic groups in the correlation between income distance and mover flows, and the estimated results for ethnic distance do not change.

Additionally, it can be argued that the role of ethnicity in influencing the size of mover flows differs between people of different income strata. For example, high-income Turks or Moroccans might move into neighborhoods with more high-income Turks and Moroccans and away from their low-income peers. Distinguishing between the mover flows of high- and low-income ethnic groups does not change our base results, and we do not find significant differences between high- or low-income ethnic groups and the role of the ethnic distance in influencing ethnic mover flows.[17] One potential caveat of these results is that the number of non-zero flows becomes very small for some income groups when we divide the ethnic mover flows into high- and low-income groups.[18]

As mentioned above, homeowner composition relates to economic characteristics, and from Table 5 it can be seen that the role of available housing, either owner-occupied or social rent and the composition of both types, is much more important than income differences for the size of mover flows in Amsterdam, while the effect of especially owner-occupied housing is not significantly different from zero for The Hague.

To test the robustness of our econometric specification, we estimate the gravity model with a ZINB regression. We assume that the zero-generating process is explained by the same covariates as the non-zero-generating process, i.e., the covariates in the NB regression. Because we run the risk of overspecifying the ZINB by including ethnic fixed effects and the ethnic and diversity friction measures for each ethnic group, we estimate the ZINB with only ethnic fixed effects and homogeneous ethnic and diversity friction measures. The results from the base specification are robust against the ZINB specification because the Poisson estimation given in Tables A.2 and A.3 in the Appendix is not generally different from the results in Table 4.[19] The binary logit regression shows that the probability of a zero count increases when the specific ethnic population share is low and if the ethnic distance for that specific ethnic group is large.

The analysis so far ignores intra-neighborhood mover flows. It can be argued that the underlying mechanisms from those of inter-neighborhood flows are different because these moves cannot be motivated by relative neighborhood characteristics and should therefore not be included in the analysis. However, LeSage and Pace (2008) provide a submodel for including the so-called diagonal flows. The results of equation (4), which include intra-neighborhood flows, are given in Tables A.4 and A.5 in the Appendix. Including the intra-neighborhood mover flows does not influence our results regarding the inter-neighborhood mover flows. For the intra-neighborhood flows, the measured differences are zero by definition, while the neighborhood stock values of the covariates used for the intra-neighborhood flows are included. The presence of the own ethnic group does not play a role in intra-neighborhood moves for the Dutch, Moroccans, and Turks, both in Amsterdam and in The Hague. The results for Caribbeans shows that mover flows within the same neighborhood are lower if the share of Caribbeans in that neighborhood is higher. We find the opposite effects for the control variables between the inter- and intra-neighborhood flows, which is not surprising because the factors that induce moving within the same neighborhood will probably counteract the factors that induce moving out of that neighborhood.

[17] We have estimated equation (7) for the mover flows of each ethnic group using €40000 household income per year as the cutoff point. The results show no statistically significant differences between the high- and low-income groups of each ethnic group and the size of mover flows. Results are available upon request.

[18] This is also the reason we cannot analyze potential differences in the effect of ethnicity on the size of ethnic mover flows between first- and second-generation migrants, although this issue is viable. There are simply not enough adult second-generation migrants in our sample.

[19] Although a Vuong test of the ZINB against the NB regression shows that the ZINB would be preferred, we consider the NB regression to be the preferred model. Our zero-generating process is easily observable from the data, in that the only cause of a zero-flow with probability 1 is the absence of a specific ethnic group in the neighborhood of origin. So as to preserve data, these flows are included.

5 Conclusion

We developed a gravity model to determine the role of ethnicity in intra-urban residential relocation dynamics. Subsequently, we utilized a unique micro data set of residential relocations across neighborhoods in the urban agglomerations of Amsterdam and The Hague, in The Netherlands. In most developed countries, ethnicity has become an increasingly important factor in neighborhood formation. As a matter of fact, most migrant flows occur at a disaggregate spatial level, i.e. within or between neighborhoods of the same urban area (Cadwallader, 1992). To the best of our knowledge, existing empirical gravity analyses of migration flows have focused merely on aggregate spatial units, such as countries or regions (Greenwood, 1975; Mueser, 1989; Rogers and Raymer, 1998; Karemera et al., 2000; Lewer and Van den Berg, 2008; Grosjean, 2011; Peeters, 2012; Falck et al., 2012; Caragliu et al., 2013). In contrast, we modeled bilateral intra-urban migration patterns at the neighborhood level as a function of bilateral socioeconomic, demographic and ethnic differences between the origin and destination neighborhoods. In doing so, we accounted for heterogeneous and interdependent preference structures of natives and various immigrant groups, as well as for the link between neighborhood choice at the individual level and change at the neighborhood level. We incorporated ethnic friction measures at the neighborhood level into our analysis, using population shares and diversity. This allows us to incorporate Schelling's observation that the effect of ethnic neighborhood composition on subsequent ethnic sorting and clustering is likely to be asymmetrical (Schelling, 1971). Following Santos Silva and Tenreyro (2006), we estimated the gravity equation using a modified version of the Poisson maximum likelihood estimator rather than applying ordinary least squares to a log-linearized version of the nonlinear relationship between distance and size of the origin and destination populations that underlies the standard gravity model.

Our results clearly show that in both cities ethnic drift leads to the clustering of ethnic minority groups and a white flight of native Dutch residents, which is in line with findings of Bolt et al. (2008) for major cities in The Netherlands. The predicted mover flow of native Dutch residents into a neighborhood with a 20 percent larger share of native Dutch residents is almost two times higher than the flow into a neighborhood with a 20 percent smaller share of Dutch residents. For the mover flows of the ethnic minority groups, this effect is four to seven times in Amsterdam, and eight to fourteen times in The Hague. Hence, the ethnic drift of the minority groups is relatively stronger than the white flight. However, the preferences of the majority group disproportionally impact the final composition of neighborhoods. Except for the Caribbean group, mover flows of all ethnic groups are higher into more diverse neighborhoods. We clearly find evidence for the hypotheses that, for all ethnic groups, the distance of the own ethnic group has an asymmetrical and nonlinear relationship with mover flows of that ethnic group. From a gravity point of view, the largest mover flows are expected to be centered around zero between neighborhoods that are alike, i.e., over small ethnic distances. This results in a concave relationship between the ethnic friction measure and the predicted flow. Together, these findings suggest a preference for living with the own ethnic group in an otherwise diverse neighborhood.

Our empirical findings therefore lend support to a modified version of the so-called ethnic enclave theory in the literature on residential mobility and migration in the context of spatial segregation. This theory states that migrants prefer to live in ethnically homogenous neighborhoods, based on their own ethnicity (Portes and Jensen, 1987). Our results show that the preference for living with the own ethnic group goes together with living in a neighborhood that is ethnically heterogeneous to some extent. This contrasts the findings of Schaake et al. (2014), who report that ethnic minorities have a tendency to move into neighborhoods with more Dutch, supporting the idea of increasingly multicultural neighborhoods. Our results are more in line with results found in the US, which show that ethnic groups have a tendency to cluster and that the underlying mechanism is not only driven by differences in socioeconomic status, but also by preferences for the own ethnic group (Krysan and Farley, 2002; Saiz and Wachter, 2011; Bayer et al., 2014). A generalization of our results would indicate that ethnic groups prefer some level of local dominance, even although not all ethnic groups can enjoy numerical superiority (Schelling, 1971). However, we do note that the link between individual preferences for an individual's ethnic group and the resulting ethnic composition of neighborhoods is not straightforward, especially in cities that have many ethnic groups, because diversity and ethnic clustering are not mutually exclusive.

With respect to the role of other neighborhood effects we find—in accordance with the gravity design—that the physical distance between neighborhoods impacts the size of mover flows negatively, while mover flows are greater into larger and growing neighborhoods. Furthermore, we find a negative correlation between income differences and mover flows, indicating that most mover flows are between neighborhoods that are alike in terms of average income. This result contrasts gravity analyses at the more aggregate spatial level, where trade or migration flows usually depend positively on income differences between the area of origin and destination. Furthermore, we find that life-cycle related characteristics of neighborhoods (captured by a friction measure of the share of children in a neighborhood) are strongly related to the size of mover flows, with higher mover flows into neighborhoods with a higher share of children. We find no statistically significant differences between ethnic groups in the correlation between income distance and mover flows and between high- or low-income ethnic groups and the role of the ethnic distance in influencing ethnic mover flows.

Our analysis contributes to the existing literature by incorporating ethnic friction measures and socioeconomic neighborhood characteristics in an empirical gravity model at a disaggregate spatial level. We think that this approach helps to assess place-based policies that predominantly aim to counter the tendency to segregation. These policies often deal with socioeconomic drivers of location choice, while our results suggest that preference heterogeneity may be at least equally important in driving intra-urban spatial segregation. For example, in The Netherlands locally designed urban renewal projects in low-income or deprived neighborhoods compete for higher-income natives with newly constructed suburbs and so called "new towns" that are planned at the national level (Bolt et al., 2008; Bolt and Van Kempen, 2010; Boschman et al., 2013). Our approach could be of use in understanding residential mobility dynamics that result from the complex interplay between, inter alia, household preferences, housing market and neighborhood characteristics, economic disparities and urban policies (Waldorf, 1993).

In future research much ground can still be gained by translating the analysis of mover flows into actual long-term neighborhood change and the role of socioeconomic distances across neighborhoods. This would require a dynamic approach of neighborhood change that incorporates the interplay of preferences of minority versus majority ethnic groups in a more structural way than we have done in the analysis in this paper. Such a model can show whether misalignments of the preferences of different ethnic groups for the ethnic compositions of neighborhoods exist. The equilibrium neighborhood composition would then also depend on the development of the economic position of ethnic minority groups (Ihlanfeldt and Scafidi, 2002; Krysan and Farley, 2002; Vigdor, 2003; Saiz and Wachter, 2011; Ioannides and Zabel, 2008). Bayer et al. (2014), for example, show for the US that the relationship between segregation and income inequality in cities is likely to be negative. A different, and promising, line of research is to develop a gravity approach for an even lower spatial level of aggregation than the neighborhood level, to correctly identify the (causal) processes of intra-urban residential mobility in the context of ethnic clustering. The clustering of ethnic groups at much lower spatial scales, such as at the block (group) level, will likely provide better insight into the spatial clustering patterns of ethnic groups and the persistence of these patterns (Vigdor, 2003). An analysis at such a low spatial scale should also put emphasis on spatial dependence of mobility choices. Especially when owner-occupied and social housing blocks are located in close proximity, this type of research may yield interesting insights for the design and evaluation of local housing policies.

References

Bayer, P., H. Fang, and R. McMillan (2014). Separate when Equal? Racial Inequality and Residential Segregation. *Journal of Urban Economics 82*(1), 32–48.

Bayer, P., R. McMillan, and K. S. Rueben (2004). What Drives Racial Segregation? New Evidence Using Census Microdata. *Journal of Urban Economics 56*(3), 514–535.

Becker, G. S. and K. M. Murphy (2000). *Social Economics: Market Behavior in a Social Environment*. Cambridge and London: The Belknap Press of Harvard University Press.

Blackburn, M. L. (2014). The Relative Performance of Poisson and Negative Binomial Regression Estimators. *Oxford Bulletin of Economics and Statistics 77*(4), 605–616.

Bolt, G. and R. Van Kempen (2010). Dispersal Patterns of Households who are Forced to Move: Desegregation by Demolition: A Case Study of Dutch Cities. *Housing Studies 25*(2), 159–180.

Bolt, G., R. Van Kempen, and M. Van Ham (2008). Minority Ethnic Groups in the Dutch Housing Market: Spatial Segregation, Relocation Dynamics and Housing Policy. *Urban Studies 45*(7), 1359–1384.

Borjas, G. J. (1995). Ethnicity, Neighborhoods, and Human Capital Externalities. *American Economic Review 85*(3), 365–390.

Boschman, S., G. Bolt, R. Van Kempen, and F. Van Dam (2013). Mixed Neighbourhoods: Effects of Urban Restructuring and New Housing Development. *Tijdschrift voor Economische en Sociale Geografie 104*(2), 233–242.

Brock, W. A. and S. N. Durlauf (2001). Interactions-Based Models. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics* (5 ed.)., pp. 3297–3380. Amsterdam: North-Holland.

Burger, M. J., F. G. Van Oort, and G.-J. M. Linders (2009). On the Specification of the Gravity Model of Trade: Zeros, excess Zeros and Zero-Inflated Estimation. *Spatial Economic Analysis 4*(2), 167–190.

Cadwallader, M. (1992). *Migration and Residential Mobility; Macro and Micro Approaches*. Madison: The University of Wisconsin Press.

Cameron, A. C. and P. K. Trivedi (2009). *Mircoeconometrics using Stata*. Texas: Stata Press.

Cameron, A. C. and P. K. Trivedi (2013). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Caragliu, A., C. Del Bo, H. L. F. De Groot, and G.-J. M. Linders (2013). Cultural Determinants of Migration. *Annals of Regional Science 51*(1), 7–32.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the Dynamics of Segregation. *Quaterly Journal of Economics 123*(1), 177–218.

Chambers, D. N. (1992). The Racial Housing Price Differential and Racially Transitional Neighborhoods. *Journal of Urban Economics 32*(2), 214–232.

Cutler, D. M. and E. L. Glaeser (1997). Are Ghettos Good or Bad? *Quarterly Journal of Economics 112*(3), 827–872.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999). The Rise and Decline of the American Ghetto. *Journal of Political Economy 107*(3), 455–506.

Ellen, I. G. and K. M. O'Regan (2011). How Low Income Neighborhoods Change: Entry, Exit, and Enhancement. *Regional Science and Urban Economics 41*(2), 89–97.

Falck, O., S. Heblich, A. Lameli, and J. Suedekum (2012). Dialects, Cultural Identity, and Economic Exchange. *Journal of Urban Economics 72*(2), 225–239.

Gabriel, S. A. and G. D. Painter (2012). Household Location and Race: a 20-Year Retrospective. *Journal of Regional Science 52*(5), 809–818.

Gabriel, S. A. and S. S. Rosenthal (1989). Household Location and Race: Estimates of a Multinomial Logit Model. *Review of Economics and Statistics 71*(2), 240–249.

Greenwood, M. J. (1975). Research on Internal Migration in the United States: a Survey. *Journal of Economic Literature 13*(2), 397–433.

Grosjean, P. (2011). The Weight of History on European Cultural Integration: A Gravity Approach. *American Economic Review 101*(3), 504–508.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural Biases in Economic Exchange? *Quarterly Journal of Economics 124*(3), 1095–1131.

Van Ham, M. and W. A. V. Clark (2009). Neighbourhood Mobility in Context: Household Moves and Changing Neighbourhoods in The Netherlands. *Environment and Planning A 41*(6), 1442–1459.

Ihlanfeldt, K. R. and B. Scafidi (2002). Black Self-Segregation as a Cause of Housing Segregation: Evidence from the Multi-City. *Journal of Urban Economics 51*(2), 366–390.

Ioannides, Y. M. and J. E. Zabel (2008). Interactions, Neighborhood Selection and Housing Demand. *Journal of Urban Economics 63*(1), 229–252.

Karemera, D., V. I. Oguledo, and B. Davis (2000). A Gravity Model Analysis of International Migration to North America. *Applied Economics 32*(13), 1745–1755.

Kiel, K. and J. E. Zabel (1996). House Price Differentials in U.S. Cities : Household and Neighborhood Racial Effects. *Journal of Housing Economics 5*(2), 143–165.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Krysan, M. and R. Farley (2002). The Residential Preferences of Blacks: Do they Explain Persistent Segregation? *Social Forces 80*(3), 937–980.

LeSage, J. P. and R. K. Pace (2008). Spatial Econometric Modeling of Origin-Destination Flows. *Journal of Regional Science 48*(5), 941–967.

Lewer, J. J. and H. Van den Berg (2008). A Gravity Model of Immigration. *Economics Letters 99*(1), 164–167.

Linders, G.-J. M. and H. L. F. De Groot (2006). Estimation of the Gravity Equation in the Presence of Zero Flows. Tinbergen Institute Discussion Paper 2006-072/3.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Long, J. S. and J. Freese (2006). *Regression Models for Categorical Dependent Variables Using Stata*. Texas: Stata Press.

Massey, D. S. and N. A. Denton (1988). Suburbanization and Segregation in US Metropolitan Areas. *American Journal of Sociology 94*(3), 592–626.

Mayda, A. M. (2010). International Migration: a Panel Data Analysis of the Determinants of Bilateral Flows. *Journal of Population Economics 23*(4), 1249–1274.

McFadden, D. (1978). Modelling the Choice of Residential Location. In A. Karlqvist, L. Lundqvist, and J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models*, pp. 75–96. Amsterdam: North Holland.

McKinnish, T., R. Walsh, and T. K. White (2010). Who Gentrifies Low-Income Neighborhoods? *Journal of Urban Economics 67*(2), 180–193.

Mueser, P. R. (1989). The Spatial Structure of Migration: An Analysis of Flows Between the States in the USA Over Three Decades . *Regional Studies 23*(3), 185–200.

Myers, C. K. (2004). Discrimination and Neighborhood Effects: understanding Racial Differentials in US Housing Prices. *Journal of Urban Economics 56*(2), 279–302.

Ortega, F. and G. Peri (2013). The Effect of Income and Immigration Policies on International Migration. *Migration Studies 1*(1), 47–74.

Papke, L. E. and J. M. Wooldridge (1996). Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics 11*(6), 619–632.

Peeters, L. (2012). Gravity and Spatial Structure: the Case of Interstate Migration in Mexico. *Journal of Regional Science 52*(5), 819–856.

Portes, A. and L. Jensen (1987). What's an ethnic enclave? The case for conceptual clarity. *American Sociological Review 52*(2), 768–771.

Rogers, A. and J. Raymer (1998). The Spatial Focus of U.S. Interstate Migration Flows. *International Journal of Population Geography 4*(1), 63–80.

Saiz, A. and S. Wachter (2011). Immigration and the Neighborhood. *American Economic Journal: Economic Policy 3*(2), 169–188.

Santos Silva, J. M. C. and S. Tenreyro (2006). The Log of Gravity. *Review of Economics and Statistics 88*(4), 641–658.

Santos Silva, J. M. C. and S. Tenreyro (2011). Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator. *Economics Letters 112*(2), 220–222.

Schaake, K., J. Burgers, and C. H. Mulder (2010). Ethnicity at the Individual and Neighborhood Level as an Explanation for Moving out of the Neighborhood. *Population Research and Policy Review 29*(4), 593–608.

Schaake, K., J. Burgers, and C. H. Mulder (2014). Ethnicity, Education and Income, and Residential Mobility Between Neighbourhoods. *Journal of Ethnic and Migration Studies 40*(4), 512–527.

Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology 1*(2), 143–186.

Vigdor, J. L. (2003). Residential Segregation and Preference Misalignment. *Journal of Urban Economics 54*(3), 587–609.

Waldorf, B. (1993). Segregation in Urban Space: A New Measurement Approach. *Urban Studies 30*(7), 1151–1164.

Winkelmann, R. (2008). *Econometric Analysis of Count Data* (5 ed.). Berlin, Heidelberg: Springer-Verlag.

Tables and Figures

Table 1: Population composition 2004.

	Amsterdam	The Hague	MA Amsterdam	MA The Hague
Total population	733,047	463,599	1,720,267	966,916
Dutch (percentage)	52.00	55.96	66.62	67.91
Caribbeans	10.99	7.68	7.19	7.68
Moroccans	8.62	5.06	4.56	3.00
Turks	5.13	6.71	3.47	3.67
Dissimilarity index[a]				
Dutch	0.27	0.34	0.30	0.32
Caribbeans	0.34	0.33	0.39	0.39
Moroccans	0.44	0.49	0.51	0.52
Turks	0.43	0.53	0.44	0.59
Yearly income per capita (€)[b]				
Overall	16,170	15,444	16,586	16,048
Dutch	19,283	17,532	18,091	17,263
Caribbeans	13,118	13,447	13,690	13,590
Moroccans	8,117	7,451	8,336	7,581
Turks	8,768	8,569	9,567	8,726

Source: GBA Statistics Netherlands.

[a] The calculation of the dissimilarity index is given in equation (12).

[b] Yearly income per capita is calculated in 2008 prices.

Table 2: Dependent variable: size of mover flows 2004–2008.

	Amsterdam				The Hague			
Percentage of non-zero flows								
Dutch	77.68				62.42			
Caribbeans	44.75				31.98			
Moroccans	32.51				15.49			
Turks	23.24				15.58			
Rest	65.69				46.97			
	mean	st.dev.	min.	max.	mean	st.dev.	min.	max.
Size flows excl. zero flows								
Dutch	17.37	66.70	1	3331	11.63	37.53	1	1744
Caribbeans	6.74	28.59	1	994	5.32	57.52	1	142
Moroccans	5.87	11.28	1	193	4.12	6.36	1	102
Turks	5.25	10.18	1	189	5.42	10.44	1	131
Rest	8.31	21.44	1	685	5.33	9.87	1	290
Size total flows								
Dutch	13.49	59.22	0	3331	7.26	30.18	0	1744
Caribbeans	3.02	19.42	0	994	1.70	4.96	0	142
Moroccans	1.91	7.00	0	193	0.64	2.91	0	102
Turks	1.22	5.38	0	189	0.84	4.56	0	131
Rest	5.46	17.82	0	685	2.51	7.27	0	290

Data source: GBA Statistics Netherlands.

Table 3: Independent variables: neighborhood characteristics 2004.[a]

	Amsterdam				The Hague			
	Friction measure		Stock value		Friction measure		Stock value	
	mean	st.dev.	min.	max.	mean	st.dev.	min.	max.
Δs								
Dutch	0	0.23	0.14	0.89	0	0.27	0.07	0.92
Caribbeans	0	0.10	0.01	0.46	0	0.11	0.00	0.30
Moroccans	0	0.11	0	0.35	0	0.07	0	0.28
Turks	0	0.07	0	0.24	0	0.10	0	0.35
Rest	0	0.09	0.09	0.36	0	0.11	0.05	0.39
ΔH								
Dutch	0	0.08	0.73	0.97	0	0.07	0.74	0.95
Caribbeans	0	0.22	0.19	0.90	0	0.24	0.12	0.84
Moroccans	0	0.21	0.20	0.89	0	0.24	0.14	0.95
Turks	0	0.21	0.20	0.89	0	0.24	0.15	0.85
Rest	0	0.31	0.03	0.72	0	0.34	0.01	0.77
Distance (km)[b]	6.79	4.99	0.28	42.49	5.20	2.97	0.38	19.79
No. dwellings (× 100)	0	225.05	0.40	942.64	0	108.52	0.32	465.10
Housing stock growth$_i$	0.20	1.08	−0.15	9.79	0.26	1.77	−0.24	14.65
Mean income (× €1000)[c]	0	7.49	8.90	35.76	0	7.96	3.94	29.74
Share social rent	0	0.30	0.05	0.88	0	0.35	0.00	0.96
Share owner occupied	0	0.25	0	0.81	0	0.33	0	0.91
Share children	0	0.08	0.05	0.30	0	0.10	0.04	0.36

Data source: GBA, SSB and Housingregister Statistics Netherlands.
[a] All variables are measured as a friction between the neighborhood of destination j and the neighborhood of origin i, unless explicitly stated otherwise.
[b] Distance is calculated as the Euclidean distance between centroids of neighborhoods.
[c] Calculated as 2008 per capita prices.

Table 4: Negative binomial regression results, base specification.[a]

	Amsterdam		The Hague	
	coef.	s.e.	coef.	s.e.
Dutch	−0.128***	0.039	0.126***	0.045
Caribbeans	−0.252***	0.042	0.200***	0.049
Moroccans	−0.374***	0.054	−0.059	0.060
Turks	0.273***	0.062	−0.016	0.060
Dutch Δs_{-}	3.559***	0.339	4.839***	0.299
Caribbeans	6.881***	0.807	2.506***	0.890
Moroccans	3.122***	0.967	5.032***	1.392
Turks	6.220***	1.815	3.904***	1.191
Dutch Δs_{+}	−1.611***	0.317	−1.318***	0.264
Caribbeans	5.172***	0.779	4.458***	0.936
Moroccans	9.005***	0.908	9.480***	1.460
Turks	9.784***	1.742	6.616***	1.289
Dutch Δs_{-}^2	2.008***	0.735	0.419	0.553
Caribbeans	10.037***	2.294	−21.822***	4.100
Moroccans	1.549	3.494	2.342	6.175
Turks	−1.422	10.604	2.392	4.369
Dutch Δs_{+}^2	0.880	0.629	1.610***	0.420
Caribbeans	1.377	2.124	−24.271***	4.495
Moroccans	−23.255***	3.652	−30.295***	6.747
Turks	−34.564***	11.290	−15.913***	5.125
Dutch ΔH	1.828***	0.300	1.001***	0.237
Caribbeans	−0.635***	0.113	−0.266**	0.125
Moroccans	0.242*	0.146	0.821***	0.164
Turks	0.210	0.169	1.007***	0.156
Distance (km)	−0.131***	0.003	−0.234***	0.004
No. dwellings (×100)	0.004***	0.000	0.003***	0.000
Housing stock growth$_j$	0.162***	0.007	0.095***	0.004
Mean income (×€1000)	−0.003	0.002	−0.035***	0.002
Share social rent	1.512***	0.054	−0.226***	0.073
Share owner occupied	0.609***	0.071	0.113	0.088
Share children	3.186***	0.193	2.777***	0.154
Constant	−2.093***	0.084	−5.153***	0.066
Observations	48,510		63,280	
Pseudo log-likelihood	–83605.698		–77175.635	
α[b]	1.140***	0.015	1.260***	0.018
Wald test χ^2 (df)[c]	23769.46 (66)***		23668.04 (64)***	

Data source: GBA, SSB and Housingregister Statistics Netherlands.

[a] Huber-White robust standard errors are reported. The statistical significance of coefficients is indicated by ***, **, and * for the 0.01, 0.05, and 0.1 significance levels, respectively. All variables are measured as a friction between the neighborhood of destination j and the neighborhood of origin i, unless explicitly stated otherwise. An ethnic rest-group, as well as neighborhood of origin and destination dummies are included in the estimation but the estimated coefficients are not reported here. The population at risk (exposure variable) is included in the estimation.

[b] The significance of α is based on a χ^2 likelihood-ratio test for overdispersion estimated on a model with non-robust standard errors with the null hypothesis being that the model is Poisson, corresponding to $H_0 : \alpha = 1$.

[c] The Wald test performs a full-model test of joint significance of all coefficients, with the null hypothesis being that all coefficients are simultaneously equal to zero.

Table 5: Marginal effects of the negative binomial regression results at percentiles.[a]

	Amsterdam				
	5%	25%	50%	75%	95%
Δs					
Dutch	−4.837	−0.715	2.450***	8.012***	32.518**
Caribbeans	0.746***	1.385***	2.026***	3.032***	6.504***
Moroccan	0.257***	0.633***	0.813***	0.955***	0.986
Turks	0.417***	0.812***	0.951***	1.075***	1.319
ΔH					
Dutch	2.412***	2.901***	3.130***	3.378***	4.063***
Caribbeans	−0.201***	−0.176***	−0.160***	−0.145***	−0.127***
Moroccan	0.038*	0.040*	0.041	0.043	0.045
Turks	0.022	0.023	0.024	0.025	0.026
Distance (km)	−3.234***	−2.502***	−1.903***	−1.263***	−0.417***
No. dwellings (×100)	0.007***	0.044***	0.050***	0.056***	0.359***
Housing stock growth$_j$	0.089***	0.089***	0.089***	0.090***	0.111***
Mean income (×€1000)	−0.017	−0.017	−0.017	−0.016	−0.016
Share social rent	0.438***	0.672***	0.936***	1.304***	2.000***
Share owner occupied	0.265***	0.316***	0.344***	0.373***	0.446***
Share Children	1.179***	1.508***	1.809***	2.171***	2.777***

	The Hague				
	5%	25%	50%	75%	95%
Δs					
Dutch	−3.369**	0.354	2.763***	7.329***	37.108**
Caribbeans	0.588***	0.965***	0.661***	−0.056	−0.956***
Moroccans	0.170***	0.332***	0.378***	0.422***	0.542
Turks	0.110***	0.327***	0.363***	0.400***	0.790
ΔH					
Dutch	0.719***	0.778***	0.814***	0.852	0.922***
Caribbeans	−0.046*	−0.043**	−0.041**	−0.040**	−0.037**
Moroccan	0.034***	0.040***	0.046***	0.053***	0.064***
Turks	0.049***	0.061***	0.072***	0.085***	0.106***
Distance (km)	−2.230***	−1.483***	−0.997***	−0.608***	−0.242***
No. dwellings (×100)	0.008***	0.010***	0.011***	0.011***	0.015***
Housing stock growth$_j$	0.025***	0.025***	0.025***	0.025***	0.025***
Mean income (×€1000)	−0.156***	−0.117***	−0.097***	−0.081***	−0.060***
Share social rent	−0.071***	−0.065***	−0.062***	−0.059***	−0.054***
Share owner occupied	0.029	0.030	0.031	0.032	0.033
Share Children	0.481***	0.638***	0.761***	0.907***	1.203***

Data source: GBA, SSB and Housing Register of Statistics Netherlands.

[a] The statistical significance of marginal effects is indicated with ***, **, and * for the 0.01, 0.05, and 0.1 significance levels, respectively, and calculated using the variance-matrix of Huber-White robust standard errors from the regression results presented in Table 4. Unit changes in x are 0.1 for the unscaled variables measured as shares, and 1 otherwise.

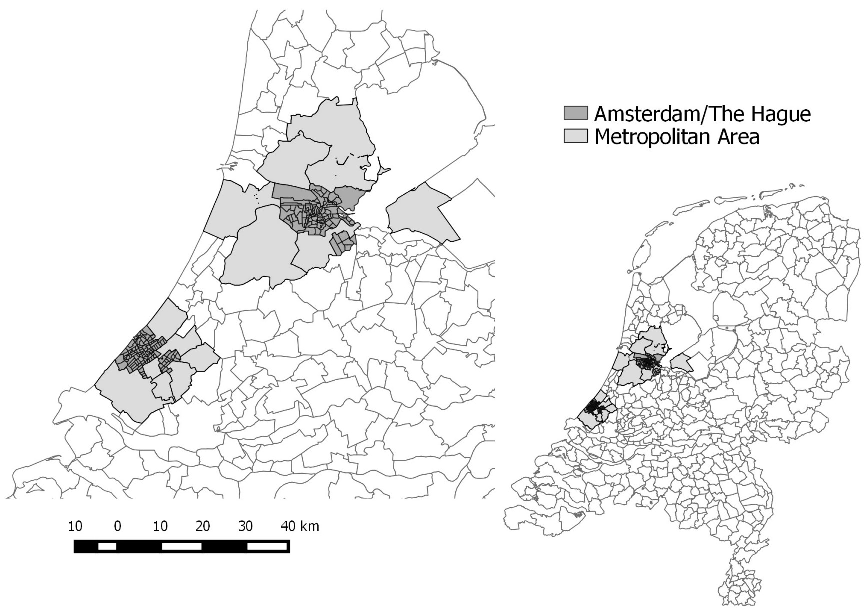
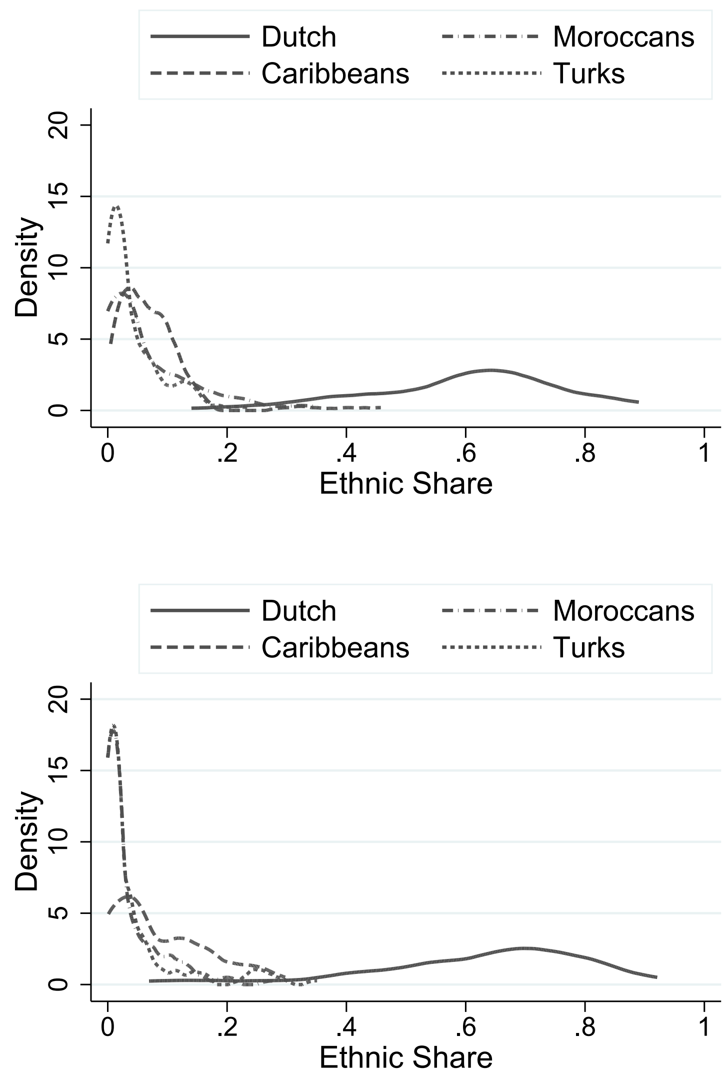


Figure 1: Neighborhoods and metropolitan areas of Amsterdam and The Hague.



Data source: GBA Statistics Netherlands.

Figure 2: Kernel density of ethnic neighborhood shares (s_i) for Amsterdam (top) and The Hague (bottom).

Data source: Housing Register of Statistics Netherlands.

Figure 3: Scatter plot of neighborhood shares of social rent and owner occupied housing in Amsterdam (top) and The Hague (bottom).

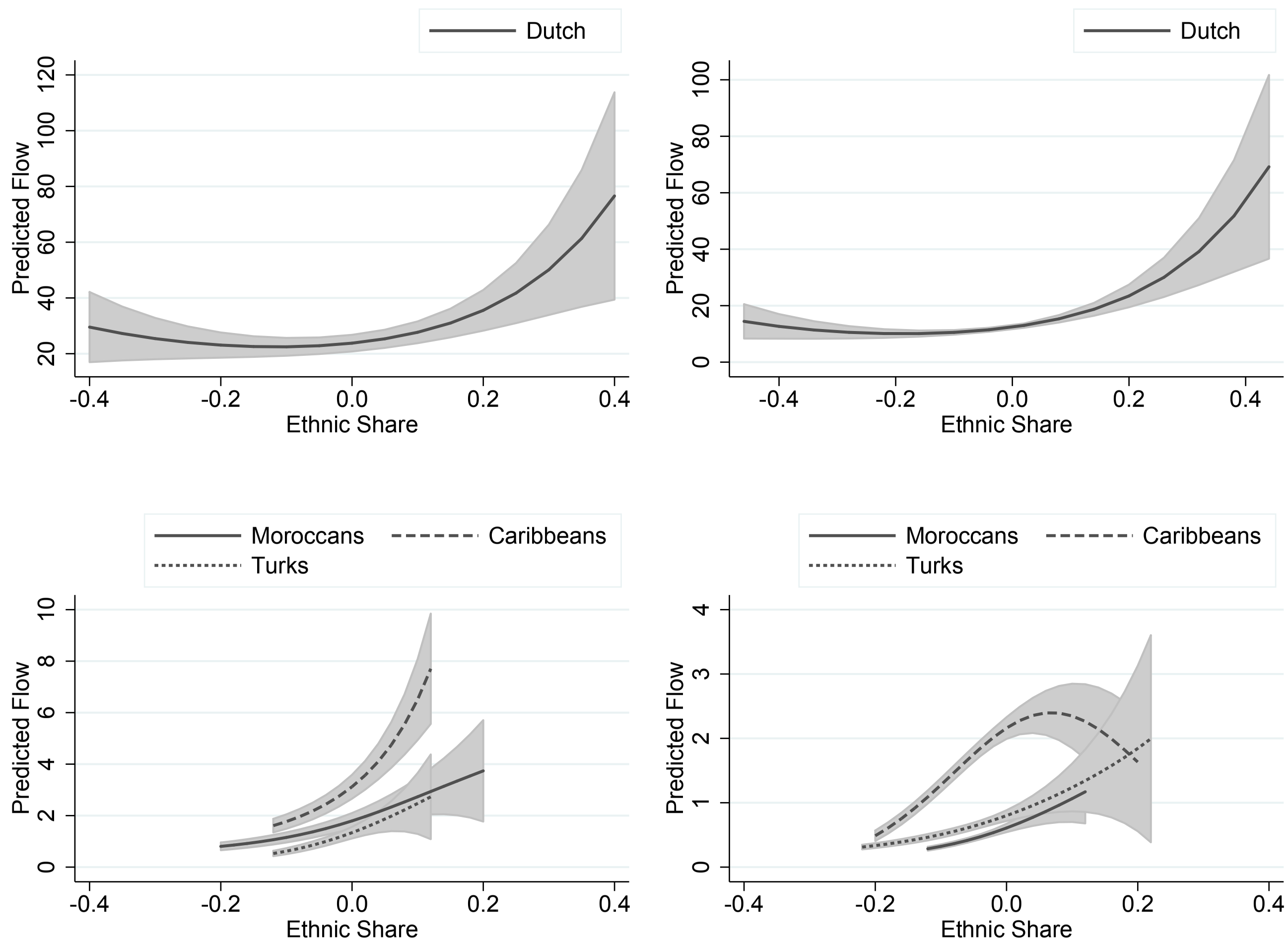


Figure 4: Mean predicted mover flows with 95% confidence interval for Amsterdam (left) and The Hague (right).

Appendix

This Appendix is not necessarily for publication. It can be available upon request or, alternatively, be published electronically. It contains details about the cities included in our analysis (Table A.1) as well as extensive tables with regression results for the robustness checks described in Section 4. Tables A.2 and A.3 contain the results of a zero-inflated negative binomial model of the base specification. Tables A.4 and A.5 include the diagonal flows within neighborhoods.

Table A.1: Aggregated neighborhoods in the metropolitan areas of Amsterdam and The Hague.[a]

Amsterdam	The Hague
Amsterdam (93)	The Hague (106)
Aalsmeer, Haarlemmermeer, Uithoorn (1)	Midden-Delfland, Westland (1)
Amstelveen, Diemen, Ouder-Amstel (1)	Delft (1)
Bennenbroek, Bloemendaal, Haarlem, Heemstede, Haarlemmerliede en Spaarnwoude, Zandvoort (1)	Leidschendam-Voorburg (1)
Zaanstad, Wormerland (1)	Pijnacker-Nootdorp (1)
Beemster, Edam-Volendam, Graft-De Rijp, Landsmeer, Oostzaan, Purmerend, Waterland, Zeevang (1)	Rijswijk (1)
Almere (1)	Wassenaar (1)
	Zoetermeer (1)

Data source: GBA Statistics Netherlands.

[a] Neighborhoods outside of the city of Amsterdam and The Hague are aggregated into 6 and 7 neighborhoods, respectively. The number of neighborhoods in the municipalities as is used in the analysis is shown in parentheses. The total number of neighborhoods in the metropolitan area of Amsterdam is 99, of which 93 are within the city of Amsterdam. The total number of neighborhoods in the metropolitan area of The Hague is 113, of which 106 are within the city of The Hague. All the neighborhood definitions are based on 2008 boundaries.

Table A.2: Zero-inflated negative binomial regression results for Amsterdam.[a]

	Negative Binomial estimation		Binary logit estimation	
	coef.	s.e.	coef.	s.e.
Dutch	−0.134***	0.039	0.692***	0.248
Caribbeans	−0.215***	0.042	0.430**	0.188
Moroccans	−0.225***	0.056	1.298***	0.196
Turks	−0.065	0.065	1.974***	0.209
Dutch Δs_- / Δs	3.143**	0.341	−2.932***	0.404
Caribbeans	5.812***	0.866		
Moroccans	3.316***	1.009		
Turks	6.257***	1.969		
Dutch Δs_+	−2.229**	0.323		
Caribbeans	6.028***	0.751		
Moroccans	8.424***	0.926		
Turks	8.828***	1.750		
Dutch Δs_-^2 / Δs^2	1.622**	0.749	3.461**	0.941
Caribbeans	7.606***	2.531		
Moroccans	4.609	3.639		
Turks	6.849	11.879		
Dutch Δs_+^2	2.755***	0.668		
Caribbeans	−0.856	2.009		
Moroccans	−20.448***	3.687		
Turks	−28.867***	11.247		
Dutch ΔH	0.938***	0.296	−0.507	0.483
Caribbeans	−0.579***	0.122		
Moroccans	0.225	0.166		
Turks	0.141	0.199		
S_i (× 100)			−0.037***	0.006
Distance (km)	−0.127***	0.003	0.054***	0.008
No. dwellings (×100)	0.004***	0.000	−0.006***	0.000
Housing stock increase$_j$	0.130***	0.006	−0.360***	0.028
Mean income (×€1000)	−0.010***	0.003	−0.024	0.016
Share social rent	0.655***	0.070	−9.030***	0.431
Share owner occupied	0.669***	0.082	0.511	0.392
Share children	3.235***	0.198	1.534*	0.912
Constant	−2.300***	0.084	−4.598***	0.276
Observations	48,510			
Nonzero observations	23,660			
Zero observations	24,850			
Pseudo log-likelihood	–82504.98			
α[b]	0.952 ***	0.014		
Vuong test z[c]	21.90 ***			
Wald test[d] χ^2_{66}	20358.98 ***			

[a] Huber-White robust standard errors are reported. The statistical significance of coefficients is indicated by ***, **, and * for the 0.01, 0.05, and 0.1 significance levels, respectively. All variables are measured as a friction between the neighborhood of destination j and the neighborhood of origin i, unless explicitly stated otherwise. An ethnic rest-group, as well as neighborhood of origin and destination dummies are included in the estimation but the estimated coefficients are not reported here. The population at risk (exposure variable) is included in the estimation.
[b] The significance of α is based on a χ^2 likelihood-ratio test for overdispersion estimated on a model with non-robust standard errors with the null hypothesis being that the model is Poisson, corresponding to $H_0 : \alpha = 1$.
[c] To test the zero-inflated negative binomial against the ordinary negative binomial model a Vuong test is performed on a model with non-robust standard errors with the null hypothesis that the expected probability of $y_i \mid x_i$ is the same for both models.
[d] The Wald-test performs a full-model test of joint significance of all coefficients, with the null hypothesis being that all coefficients are simultaneously equal to zero.

Table A.3: Zero-inflated negative binomial regression results for The Hague.[a]

	Negative Binomial estimation		Binary logit estimation	
	coef.	s.e.	coef.	s.e.
Dutch	0.060	0.043	0.450**	0.197
Caribbeans	0.283***	0.048	0.375***	0.141
Moroccans	0.069	0.061	0.797***	0.167
Turks	0.223***	0.063	1.245***	0.170
Dutch Δs_- / Δs	4.255***	0.289	−2.016***	0.413
Caribbeans	2.825***	0.892		
Moroccans	3.371**	1.417		
Turks	3.502***	1.204		
Dutch Δs_+ / Δs	−1.299***	0.257		
Caribbeans	3.664***	0.912		
Moroccans	9.908***	1.489		
Turks	4.926***	1.280		
Dutch Δs_-^2 / Δs^2	0.570	0.525	−1.340*	0.938
Caribbeans	−13.800***	4.249		
Moroccans	−1.318***	6.398		
Turks	2.064	4.532		
Dutch Δs_+^2	1.520***	0.416		
Caribbeans	−19.209***	4.308		
Moroccans	−29.369***	6.746		
Turks	−10.324**	4.990		
Dutch ΔH	1.080***	0.238	0.065	0.373
Caribbeans	0.176	0.133		
Moroccans	0.946***	0.171		
Turks	1.150***	0.167		
S_i ($\times$ 100)			−0.069***	0.006
Distance (km)	−0.226***	0.004	0.149***	0.015
No. dwellings (×100)	0.003***	0.000	−0.008***	0.000
Housing stock increase$_j$	0.078***	0.004	−11.219***	0.742
Mean income (×€1000)	−0.002	0.003	0.320***	0.011
Share social rent	−0.062	0.075	3.396***	0.318
Share owner occupied	0.454***	0.094	6.269***	0.458
Share children	2.785***	0.163	−0.350	0.658
Constant	−5.093***	0.066	−4.329***	0.182
Observations	63,280			
Nonzero observations	21,824			
Zero observations	41,456			
Pseudo log-likelihood	−75582.07			
α[b]	0.986 ***	0.015		
Vuong test z[c]	27.20 ***			
Wald test[d] χ^2_{64}	20037.18 ***			

[a] Huber-White robust standard errors are reported. The statistical significance of coefficients is indicated by ***, **, and * for the 0.01, 0.05, and 0.1 significance levels, respectively. All variables are measured as a friction between the neighborhood of destination j and the neighborhood of origin i, unless explicitly stated otherwise. An ethnic rest-group, as well as neighborhood of origin and destination dummies are included in the estimation but the estimated coefficients are not reported here. The population at risk (exposure variable) is included in the estimation.
[b] The significance of α is based on a χ^2 likelihood-ratio test for overdispersion estimated on a model with non-robust standard errors with the null hypothesis being that the model is Poisson, corresponding to $H_0 : \alpha = 1$.
[c] To test the zero-inflated negative binomial against the ordinary negative binomial model a Vuong test is performed on a model with non-robust standard errors with the null hypothesis that the expected probability of $y_i \mid x_i$ is the same for both models.
[d] The Wald-test performs a full-model test of joint significance of all coefficients, with the null hypothesis being that all coefficients are simultaneously equal to zero.

Table A.4: Negative binomial regression results Amsterdam including diagonal flows.[a]

	Inter-neighborhood flows		Intra-neighborhood flows	
	coef.	s.e.	coef.	s.e.
Dutch	-0.132^{***}	0.039	1.545	4.075
Caribbeans	-0.250^{***}	0.042	3.155^{***}	0.531
Moroccans	-0.372^{***}	0.054	2.998^{***}	0.598
Turks	-0.272^{***}	0.062	3.135^{***}	0.840
Dutch Δs_- / Δs_i[b]	3.538^{***}	0.338	5.319	4.274
Caribbeans	6.884^{***}	0.807	-6.376^{**}	2.963
Moroccans	3.101^{***}	0.967	1.962	2.384
Turks	6.144^{***}	1.816	-2.349	4.667
Dutch Δs_+	-1.592^{***}	0.316		
Caribbeans	5.173^{***}	0.778		
Moroccans	9.022^{***}	0.907		
Turks	9.830^{***}	1.740		
Dutch Δs_-^2 / $\Delta s_i^{2\mathrm{b}}$	1.985^{***}	0.732	-4.037	2.933
Caribbeans	10.040^{***}	2.294	16.139^{**}	6.648
Moroccans	1.523	3.495	1.013	7.624
Turks	-1.650	10.609	21.118	21.252
Dutch Δs_+^2	0.871	0.628		
Caribbeans	1.369	2.120		
Moroccans	-23.291^{***}	3.646		
Turks	-34.710^{***}	11.271		
Dutch ΔH	1.819^{***}	2.990	0.486	5.687
Caribbeans	-0.634^{**}	0.113	0.875	0.554
Moroccans	0.245^{*}	0.146	0.319	0.637
Turks	0.214	0.169	0.705	0.959
Distance (km)	-0.130^{***}	0.003		
No. dwellings ($\times 100$)	0.004^{***}	0.000	-0.002^{***}	0.000
Housing stock increase	0.161^{***}	0.007	0.140^{***}	0.020
Mean income ($\times$€1000)	-0.003	0.002	-0.028^{*}	0.015
Share social rent	1.510^{***}	0.054	-1.105^{***}	0.357
Share owner occupied	0.608^{***}	0.070	-0.918^{*}	0.489
Share children	3.184^{***}	0.193	-0.658	1.033
Constant	-2.111^{***}	0.083		
Observations	49005			
Pseudo log-likelihood	–85837.036			
α[c]	1.117^{***}	0.015		
Wald test[d] χ^2_{92}	41299.19^{***}			

[a] Huber-White robust standard errors are reported. The statistical significance of coefficients is indicated by ***, **, and * for the 0.01, 0.05, and 0.1 significance levels, respectively. All variables are measured as a friction between the neighborhood of destination j and the neighborhood of origin i, unless explicitly stated otherwise. An ethnic rest-group, as well as neighborhood of origin and destination dummies are included in the estimation but the estimated coefficients are not reported here. The population at risk (exposure variable) is included in the estimation.
[b] The variables for the intra-neighborhood flows are not estimated as differences between neighborhood i and j, but as stock variables of the neighborhood. For the intra-neighborhood flows, Δs_i and Δs_i^2 are estimated.
[c] The significance of α is based on a χ^2 likelihood-ratio test for overdispersion estimated on a model with non-robust standard errors with the null hypothesis being that the model is Poisson, corresponding to $\mathrm{H}_0 : \alpha = 1$.
[d] The Wald-test performs a full-model test of joint significance of all coefficients, with the null hypothesis being that all coefficients are simultaneously equal to zero.

Table A.5: Negative binomial regression results The Hague including diagonal flows.[a]

	Inter-neighborhood flows		Intra-neighborhood flows	
	coef.	s.e.	coef.	s.e.
Dutch	0.122***	0.045	−1.837	1.396
Caribbeans	0.202***	0.049	4.080***	0.741
Moroccans	−0.054	0.060	4.186***	0.793
Turks	−0.012	0.060	5.114***	0.830
Dutch Δs_-/ Δs_i[b]	4.820***	0.298	1.301	1.972
Caribbeans	2.475***	0.887	−12.932***	3.988
Moroccans	5.030***	1.390	−2.953	9.865
Turks	3.871***	1.191	0.169	5.786
Dutch Δs_+	−1.293***	0.263		
Caribbeans	4.480***	0.933		
Moroccans	9.437***	1.457		
Turks	6.634***	1.285		
Dutch Δs_-^2 / Δs_i^2[b]	0.401	0.552	0.155	2.010
Caribbeans	−21.952***	4.091	26.120**	12.429
Moroccans	2.378	6.170	3.029	30.491
Turks	2.302	4.369	−0.557	14.978
Dutch Δs_+^2	1.589***	0.419		
Caribbeans	−24.374***	4.483		
Moroccans	−30.086***	6.736		
Turks	−15.977***	5.110		
DutchΔH	1.005***	0.236	5.629***	1.748
Caribbeans	−0.270**	0.125	1.388	0.956
Moroccans	0.821***	0.163	−0.233	1.504
Turks	1.005***	0.155	−1.686	1.171
Distance (km)	−0.234***	0.004		
No. dwellings (×100)	0.003***	0.000	−0.003**	0.000
Housing stock increase$_j$	0.095***	0.004	0.117***	0.036
Mean income (×€1000)	−0.035***	0.002	−0.027**	0.013
Share social rent	−0.229***	0.072	−2.220***	0.441
Share owner occupied	0.110	0.087	−2.124**	0.595
Share children	2.781***	0.154	0.980	0.838
Constant	−5.169***	0.065		
Observations	63845			
Pseudo log-likelihood	–79239.037			
α[c]	1.234***	0.017		
Wald test[d] χ^2_{90}	32242.42***			

[a] Huber-White robust standard errors are reported. The statistical significance of coefficients is indicated by ***, **, and * for the 0.01, 0.05, and 0.1 significance levels, respectively. All variables are measured as a friction between the neighborhood of destination j and the neighborhood of origin i, unless explicitly stated otherwise. An ethnic rest-group, as well as neighborhood of origin and destination dummies are included in the estimation but the estimated coefficients are not reported here. The population at risk (exposure variable) is included in the estimation.

[b] The variables for the intra-neighborhood flows are not estimated as differences between neighborhood i and j, but as stock variables of the neighborhood. For the intra-neighborhood flows, Δs_i and Δs_i^2 are estimated.

[c] The significance of α is based on a χ^2 likelihood-ratio test for overdispersion estimated on a model with non-robust standard errors with the null hypothesis being that the model is Poisson, corresponding to $H_0 : \alpha = 1$.

[d] The Wald-test performs a full-model test of joint significance of all coefficients, with the null hypothesis being that all coefficients are simultaneously equal to zero.