Rouwendal, Jan; Levkovich, Or; Mulalic, Ismir

Working Paper

Discrete Choice Models for Commuting Interactions

Tinbergen Institute Discussion Paper, No. 17-067/VIII

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Rouwendal, Jan; Levkovich, Or; Mulalic, Ismir (2017) : Discrete Choice Models for Commuting Interactions, Tinbergen Institute Discussion Paper, No. 17-067/VIII, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/177635

TI 2017-067/VIII
Tinbergen Institute Discussion Paper

Discrete Choice Models for Commuting Interactions

Jan (J.) Rouwendal[1]
Or Levkovich[2]
Ismir Mulalic[3]

1: Vrije Universiteit Amsterdam; Tinbergen Institute, The Netherlands
2: Vrije Universiteit Amsterdam
3: DTU, KRAKS

Discrete choice models for commuting interactions

Jan Rouwendal[1,2]**, Or Levkovich**[1] **and Ismir Mulalic**[3]

This version: July 27, 2017

Key words: quantitative spatial economics, multinomial logit, mixed logit, independence of irrelevant alternatives
JEL codes: R1, R2, R4

Abstract
An emerging quantitative spatial economics literature models commuting interactions by a gravity equation that is mathematically equivalent to a multinomial logit model. This model is widely viewed as restrictive because of the independence of irrelevant alternatives (IIA) property that links substitution behavior in response to changes in the attractiveness of choice alternatives to choice probabilities in a mechanistic way. This is relevant for counterfactual analysis. In this paper we examine the appropriateness of the commuting model from a theoretical as well as an empirical point of view. We show that conventional specification tests of the multinomial logit model are of limited use when alternative specific constants are used, as is common in the recent literature, and offer no information with respect to the validity of IIA. In particular, we show that maximum likelihood estimation of relevant nested logit model is impossible because the crucial parameters are not identified. We discuss cross-nested and mixed logit as alternatives. We argue that a comparison between predicted and actual changes in commuting flows in response to a change in the attractiveness of choice alternatives provides a more informative test for the validity of the multinomial logit model for commuting interaction and report the results of such a test – as well as others – for data referring to Copenhagen.

1 Department of Spatial Economics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands.
[2] Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, Netherlands.
[3] DTU, Anker Engelunds Vej 1, 2800 Kgs. Lyngby, Denmark, and Kraks Fond – Institute for Urban Economic Research, Frederiksholms Kanal 30, DK-1220 Copenhagen, Denmark. Copenhagen, Denmark.

1 Introduction

Commuting is an important source of traffic in every urban region and a main cause of congestion. Understanding this phenomenon is therefore crucial for urban transportation analyses. Commutes connect residential and work locations and are therefore crucial for the functioning of the urban economy. The monocentric model implies an extremely simple commuting pattern. The predictions of this model with respect to home-work trips were shown to be problematic by Hamilton (1982), see also Small & Song (1992). The introduction of market imperfections of the type commonly assumed in labor economics can explain apparently 'wasteful' commuting (Rouwendal, 1998). Alternatively the introduction of idiosyncratic preferences for residential and work locations, as is standard in discrete choice models based on additive random utility, can deal with crisscrossing commute patterns.

The most popular type of discrete choice model is the multinomial logit. Indeed, this model has been employed for modelling commuting interactions in a number of recent studies that analyze important issues that arise in heterogeneous urban areas with decentralized employment: see Ahlfeldt, Redding, Sturm, & Wolf (2016) for Berlin before, during and after the wall, Monte, Redding, & Rossi-Hansberg (2015) for trade and commuting between U.S. counties, Allen, Arkolakis, & Li (2015) for the optimal division the stock if real estate into residences and workplaces in Chicago.[1] Although these papers do not motivate their commuting models in the same way as McFadden (1974) did, using extreme value type I distributed random utilities, but refer to Eaton & Kortum (2002) who derive their interaction equation from Fréchet distributed utilities, they are formally equivalent to McFadden's multinomial logit model.[2]

In this emerging quantitative spatial economics literature (reviewed by Redding & Rossi-Hansberg, 2017) combinations of residential and work locations are considered as the choice alternatives while the population of commuters is considered as homogeneous: their preferences are assumed to identical except for the values of the iid idiosyncratic parts. The utility of these combinations is usually specified (possibly after a logarithmic transformation) as the sum of four terms corresponding to (i) the utility of living in a particular location, (ii) the utility of working in a particular location, (iii) the friction between the two locations caused by the distance that has to be bridged by commuting and (iv) a random term that reflects the idiosyncratic preferences of the worker for the choice alternative. The latter covers considerations that are difficult or impossible to measure such as the desire to live with one's relatives or where one has grown up. The description suggests immediately that such idiosyncratic preferences may be common to different choice alternatives. It seems reasonable that individual-specific preferences for a particular residential location affect *all* choice alternatives with this specific residential location (recall that the alternatives are *pairs* of residential and work locations). And there is a similar argument for work locations. However, McFadden's (1974) conditional multinomial logit model is built on the assumption that the idiosyncratic utilities of the choice alternatives are *independently* distributed. Indeed, much of the later development of discrete choice models, that turned to less restrictive structures like nested, cross-nested and mixed logit, has been driven by the desire to relax this restrictive assumption.

A significant innovation in the use of the multinomial logit model, introduced by Berry, Levinsohn, & Pakes (1995) (BLP), is the use of an alternative-specific error term, commonly

[1] See Redding & Rossi-Hansberg (2017) for a review of the broader literature.

[2] We discuss this in some detail in the Appendix.

denoted as ξ, that reflects the impact of unobserved characteristics on utility. In the first stage of the estimation procedure this term is absorbed in an alternative-specific constant, while in a second stage it acts as the error term in a linear decomposition of that constant. In the absence of alternative-specific constants it is possible that predicted choice probabilities differ from the observed choice frequencies that would be expected on the basis of sampling error, the only source of the discrepancy if the model is correctly specified.

Rouwendal (2017) has recently shown that the use of alternative-specific constants for all alternatives has important implications for the application of conventional specification tests for the multinomial logit model as proposed by Hausman & McFadden (1984) and others with a single type of commuters. Estimation of the model on subsets of the alternatives always leads to identical estimates of the alternative-specific constants which means that this is useless as a test for IIA. Moreover, the additional parameters that embody the difference between multinomial logit and a nested logit model are not identified. The first-order conditions for maximum likelihood estimation of these parameters turn out to be automatically satisfied when the first-order conditions for the alternative-specific constant hold. This result holds in fact for *any* GEV model one may attempt to use to relax the restrictive characteristics of the multinomial logit, for instance a cross nested logit. Also, if one formulates a mixed logit model by attaching a random error term to all choice alternatives, a similar result holds if that additional random term is an additive function of standardized random variables that is linear in the parameters. All this suggests strongly that without explicit information about heterogeneity of the actors, for instance related to their characteristics, introduction of alternative-specific constants implies that the MNL is essentially the only relevant model.

These results do not directly apply to the MNL model for commuting that we consider here. The reason is that this model does not have alternative-specific constants for all alternatives. That is, there are alternative-specific constants for all residential and employment locations, but not for all combinations of such locations. We will, nevertheless, show that the additional parameters introduced by the obvious generalizations of the multinomial logit model for commuting to nested logit are not identified. However, estimation of the model on subsets gives useful information about the appropriateness of the chosen specification compared to one in which there are alternative-specific constants for all alternatives. GEV models that differ from nested logit – for instance, cross-nested logit models - can be estimated using the same set of alternative-specific constants as in the MNL version of the commuting model.

2 Specification of discrete choice models for commuting

2.1 Introduction

In this section we discuss the specification of discrete choice models for commuting interactions in the framework of additive random utility models. The discussion refers throughout to a single type of commuters,[3] that is to a population of actors with the same deterministic parts of the utilities attached to choice alternatives and idiosyncratic utilities that are random draws from a given distribution.

We consider workers who choose a combination of a residential location $i, i = 1 \dots I$ and a work location $j, j = 1 \dots J$. The utility they attach to such combinations is the sum of a deterministic part v_{ij} and a random part ε_{ij}:

[3] This is the setting that has thus far been used in quantitative spatial economics models, see Redding & Rossi-Hansberg (2017)

$$u_{ij} = v_{ij} + \varepsilon_{ij}. \quad (1)$$

We specify the deterministic part as:

$$v_{ij} = vr_i + ve_j + t_{ij} \quad (2)$$

The first term vr_i represents the utility of living in location i, which depends on the housing price and amenities. The second term ve_j is the utility of working at location j, which depends on the wage and amenities. The third term, t_{ij}, refers to the match of both and depends on commuting costs. The first term is common to all alternatives that have i as the residential location, the second term to all alternatives that have j as the residential location. This is the specification that is common in the emerging quantitative spatial economics literature referred to in the introduction. Sometimes parts of the terms vr_i or ve_j are specified by reference to other studies, and then the remaining parts of these terms are estimated as alternative specific constants through equalizing the observed frequencies of living or working in a particular location to those implied by the model. For instance, Ahlfeldt et al. (2016) use this procedure to identify the impact of the Berlin wall, and through this, accessibility to jobs and people on the attractiveness of zones for residential and employment purposes.

It seems reasonable to specify the random parts of the utilities similar to the deterministic parts, that is, as:

$$\varepsilon_{ij} = \rho\theta_i + \sigma\varphi_j + \tau\omega_{ij} \quad (3)$$

In this equation θ_i, φ_j and ω_{ij} are random variables and ρ, σ and τ are parameters. The random variables refer to idiosyncratic preferences with respect to, respectively, the residential location, the work location and the combinations of both. We assume throughout this paper that the ω_{ij}'s are standard IID extreme value type I distributed. This implies that the multinomial logit model (MNL) will result if $\rho = \sigma = 0$. Choice probabilities then are:

$$\pi_{ij} = \frac{e^{v_{ij}/\tau}}{\sum_{i'}\sum_{j'}\left(e^{v_{i'j'}}\right)^{1/\tau}} \quad (4)$$

The term θ_i refers to idiosyncratic preferences for residential location i. Such preferences may, for instance, be related to one's place of birth, to locations where relatives live or that are known from experience. It seems a priori very likely that such preferences exist. The important point to note is that they are incompatible with the MNL (4). Similarly, it can easily be argued that there may exist idiosyncratic preferences that are related to specific employment locations and cause correlation between ε_{ij}'s that share such a location. Again, the implication would be that MNL is inappropriate. This discussion strongly suggests that ρ and σ can be positive. In the next subsections we will consider if and to what extent existing discrete choice models are able to reflect the proposed specification (3).

2.2 Nested logit models

The nested logit model was developed as an alternative to MNL that enables a researcher to avoid the IIA property. Since it has the MNL as a special case, its estimation has also been proposed as a way to test the appropriateness of the MNL.

Two nested logit models are especially relevant: (i) the one that one nests the work locations in the residential locations, and (ii) the one that nests the residential locations in the work locations. A researcher has to choose between these two, which means that with this discrete choice model either ρ or σ in (3) has to be set equal to 0 *a priori*. This is a clear drawback of the nested logit It means that it can address at most half of the restrictiveness of the MNL.

If the first variant is chosen, residential locations are at the top of the utility tree, and σ is set equal to 0. We can then rewrite the random part of the utilities as: $\varepsilon_{ij} = \rho\left(\theta_i + \frac{\tau}{\rho}\omega_{ij}\right)$. One may think about this term as the combination of a random draw from the distribution of θ_i for each residential location and a random draw of ω_{ij} for each combination of residential and work locations. For a given actor, the value of θ_i is this common to all choice alternatives with the same residential location i.

The nested logit model requires that $\theta_i + \frac{\tau}{\rho}\omega_{ij}$ is also standard extreme value type I distributed and that $0 < \tau/\rho \leq 1$. This implies that he random variable θ_i must have the particular distribution for which this is true. This suggests immediately that the distribution of θ_i depends on the value of the parameter $\frac{\tau}{\rho}$, which is somewhat peculiar. There does not appear to be any reason why someone's idiosyncratic preferences for a particular residential location are related to idiosyncratic preferences for commutes starting from that location in this specific way. Cardell (1997) studied the distribution of θ_i and presents an analytical formulation for its density function $f_{\frac{\tau}{\rho}}$:

$$f_{\tau/\rho}(x) = \frac{1}{\tau/\rho}\sum_{n=0}^{\infty}(-1)^n \frac{e^{-nx}}{n!\Gamma\left(-\frac{\tau}{\rho}n\right)} \qquad (5)$$

Since the Γ-function is not defined for non-negative integers, this implies that the density of θ_i is not defined for any rational $\frac{\tau}{\rho}$. This complicates the use of this formula. Moreover, Cardell (1997) shows that the distribution function of θ_i never has a closed-form representation.

We conclude that the nested logit model fits in the framework of (3), but in a somewhat unintuitive way, even if one accepts that the random parts of the utilities are only correlated when the choice alternatives share the residential location.

The implied choice probabilities are:

$$\pi_{ij} = \frac{e^{v_{ij}/\tau}}{\sum_{j'} e^{v_{ij'}/\tau}} \frac{\left\{\sum_{j'} e^{v_{ij'}/\tau}\right\}^{\tau/\rho}}{\sum_{i'}\left\{\sum_{j'} e^{v_{ij'}/\tau}\right\}^{\tau/\rho}} . \qquad (6)$$

In this equation π_{ij} denotes the probability that the combination of residential location i and work location j will be chosen. The multinomial logit is the special case in which $\rho = \tau$. Estimation of the model should reveal the value of $\frac{\tau}{\rho}$. If it differs significantly from 1, the nested logit is rejected by the data. However, Appendix B shows that when the deterministic parts of the utilities are specified as in (2), the likelihood function becomes flat with respect to $\frac{\tau}{\rho}$ if the first-order conditions for the alternative-specific constants are satisfied. Hence this nested logit model cannot tell us anything about the possible restrictiveness of the MNL.

Similar conclusions hold for the alternative nested logit structure.

2.3 Cross-nested logit

The cross-nested logit model is a generalization of the nested logit models discussed in the previous section that has both of them as special cases. It allows simultaneously for correlations between the idiosyncratic utilities of alternatives with the same residential location as well as between the idiosyncratic utilities of alternatives with the same work location.

The cross-nested logit is a GEV model with generator function:

$$G(y) = \alpha \sum_i \left\{\sum_j \left(y_{ij}\right)^{\delta/\tau_1}\right\}^{\tau_1/\delta} + (1-\alpha)\sum_j \left\{\sum_i \left(y_{ij}\right)^{\delta/\tau_2}\right\}^{\tau_2/\delta} . \qquad (7)$$

which is a weighted average of generator functions that lead to the two types of nested logit models discussed in the previous subsection. The choice probabilities are:

$$\pi_{ij} = \frac{\alpha e^{v_{ij}/\delta}\left\{\sum_{j\prime}\left(\alpha e^{v_{ij\prime}}\right)^{1/\tau_1}\right\}^{\tau_1/\delta-1} + (1-\alpha)e^{v_{ij}/\delta}\left\{\sum_{i\prime}\left((1-\alpha)e^{v_{i\prime j}}\right)^{1/\tau_2}\right\}^{\tau_2/\delta-1}}{\sum_{i\prime}\left\{\sum_{j\prime}\left(\alpha e^{v_{i\prime j\prime}}\right)^{1/\tau_1}\right\}^{\tau_1/\delta} + \sum_{j\prime}\left\{\sum_{i\prime}\left((1-\alpha)e^{v_{i\prime j\prime}}\right)^{1/\tau_2}\right\}^{\tau_2/\delta}} \quad (8)$$

If α equals 1 the nested logit model with the residential locations at the top of the decision tree results, if it equals 0 the nested logit model with the work locations at the top of the decision tree results.[4]

Abbe et al. (2007) have shown that the model (10) results from maximization of an additive random utility function if the random parts of the utilities are defined as:

$$\varepsilon_{ij} = \delta \max\left\{\ln\alpha + \theta_i + \frac{\tau_1}{\delta}\omega_{ij}, \ln(1-\alpha) + \varphi_j + \frac{\tau_2}{\delta}\omega'_{ij}\right\} \quad (9)$$

In this equation α is a deterministic weight, $0 \leq \alpha \leq 1$. The random parts of the utilities are thus either $\delta\left(\theta_i + \frac{\tau_1}{\delta}\omega_{ij}\right)$ or $\delta\left(\varphi_j + \frac{\tau_2}{\delta}\omega'_{ij}\right)$. The random variables ω_{ij} and ω'_{ij} are independent standard extreme value type I distributed random variables. As in the (non-cross-)nested logit models, the random variables defined by the expressions in parentheses, $\left(\theta_i + \frac{\tau_1}{\delta}\omega_{ij}\right)$ and $\left(\varphi_j + \frac{\tau_2}{\delta}\omega'_{ij}\right)$, must also be standard extreme value type I distributed. Moreover, $\frac{\tau_1}{\delta}$ and $\frac{\tau_2}{\delta}$ should be positive and at most equal to 1.

The specification of the random parts of the utilities in the cross-nested logit model thus differs from that in (3). The parameters for θ_i and φ_j are now restricted to be equal, whereas the parameter for ω_{ij} is now allowed to take on two different values, that in fact refer to two different random variables that are identically and independently extreme value type I distributed.

The generalization of the two types of nested logit models that the cross-nested logit realizes thus appears to be the result of a specification of the random parts of the utilities that differs substantially from (3). To allow the 'logsum coefficients' of the two types of nested logit models to be different in the cross-nested logit model the parameters associated with ω_{ij} and ω'_{ij} must be allowed to differ. Although this gives the impression that the cross nested logit generalizes the two types of nested logits, the underlying model structure is different and the generalization is thus more apparent than real. Moreover, the peculiar properties of the random parts of the utilities that we found in the two types of nested logit models are also present here.

We conclude that the cross nested logit model is an attractive alternative to the two types of nested logit models because it allows for correlation of the random parts of the utilities along the two lines we desired, but that it does so in a way that is not covered by (3).

2.4 Mixed logit

The last specification that we consider is the mixed logit model. With this model we neither 'integrate out' the θ_i and φ_j analytically, nor assume them away. Instead, we keep them in the model and integrate out only the ω_{ij}'s analytically, starting from (3), which leads to a multinomial logit model with choice probabilities:

$$\pi_{ij} = \frac{e^{(v_{ij}+\rho_i\theta_i+\sigma_j\varphi_j)/\tau}}{\sum_{i\prime}\sum_{j\prime}e^{(v_{i\prime j\prime}+\rho_{i\prime}\theta_{i\prime}+\sigma_{j\prime}\varphi_{j\prime})/\tau}} \quad (10)$$

[4] It is not difficult to verify that the choice probability in (11) can be rewritten as a weighted average of the choice probabilities associated with the two types of nested logit models discussed above.

In these choice probabilities the random variables θ_i and φ_j are still present. We can integrate them out numerically if we choose a parametric distribution for these variables and estimate the model by simulation.

One possibility would be to choose the distribution of θ_i and φ_j as those shown by Cardell (1997) which would make the mixed logit nesting the two types of nested logit models discussed in 2.2 in a way that differs from the cross-nested logit and fits better in the structure suggested by (3). However, this is unattractive. As we have seen, these distributions depend on the parameters to be estimated (that is, on the values of $\frac{\tau}{\rho}$ and $\frac{\tau}{\sigma}$) and that there is no closed form representation of their distribution functions. Apart from the practical difficulties associated with these aspects, there is the more fundamental point that this specification seems implausible.

With a specification of the distributions of θ_i and φ_j that is independent of the values of the coefficients associated with them, the mixed logit model represents the structure defined by (3) exactly and without imposing additional restrictions. In fact it is the only model discussed here that does so.

The result of Rouwendal (2017) for mixed logit models does not hold in the current model, because we have only introduced alternative-specific constants for origins and destinations, and not for all commuting flows. See the discussion in Appendix C. Hence there is no reason to expect that the parameters of a mixed logit model as specified in (10) cannot be estimated.

2.5 Conclusion

The results of this section are not entirely positive. The nested and cross-nested logit models are based on specifications of the idiosyncratic utilities that lack plausibility. Moreover, the parameters of the nested logit model that embody the deviation from MNL are not identified when alternative-specific constants for residential and employment locations are present. The mixed logit model appears to be the best candidate for a useful generalization of the MNL.

However, it should be noted that the results of Rouwendal (2017) imply that introduction of a full set of alternative-specific constants would make it impossible to estimate any of the models just mentioned. If we find that estimation of the model on subsets of alternatives leads to different parameter values, or that estimation of a mixed logit model of the type discussed above improves the likelihood of the model, this cannot be interpreted as evidence against the IIA property. It only shows that the specification of the deterministic part of the utility function in (2) is unable to approximate the alternative specific constants that would be estimated if we would have specified the deterministic part of the utility function instead as $v_{ij} = asc_{ij}, i = 1 \dots I, j = 1 \dots J$.

3 Testing the MNL model for commuting

In this section we consider specification tests for the MNL model for commuting. The first is estamation of the model on a subset of alternatives as proposed in Hausman and McFadden (1984). The second test exploits variation over time in the attractiveness of alternatives to compare the actual changes in commuting flows with those predicted by the model. This test is at the heart of the potentially restrictive IIA property: the substitution behavior that occurs in response to changes in the deterministic utilities.

3.1 Estimating the model on subsets

Hausman & McFadden (1984) have considered a comparison of the coefficients estimated on (i) all observations and (ii) a subset of the observations in which choices referring to a limited

number of alternatives are left out. For instance, one could leave out some residential locations or some employment locations (or both). One would then compare the estimated coefficients that are common to the unrestricted and restricted model, following the recipe of these authors.

Since the terms t_{ij} in (2) are restricted by a functional form assumption, for instance by linking them to an indicator of the distance between residential location i and employment location j, the commuting model is a restricted version of a model with a full set of alternative-specific constants. The specification test can thus be considered as referring to the appropriateness of the chosen specification of t_{ij}.

To formalize this a bit, we introduce the notation asc_{ij} for the unbiased estimate of the deterministic parts of the utilities when the full set of alternative-specific constants is introduced. After estimating the asc_{ij}'s we can specify them further as suggested in (2):

$$asc_{ij} = vr_i + ve_j + t(d_{ij}) + \xi_{ij}, \quad (11)$$

where $t(d_{ij})$ is a parametric specification of the 'friction' associated with the geographical separation of residential and work locations which is assumed to be a function of a distance measure, and ξ_{ij} is an error term. To make things concrete, in the empirical work that we report below we assume a linear specification in travel time by car, d_{ij}^{car}, and by public transport, d_{ij}^{pt}:

$$t(d_{ij}) = \alpha d_{ij}^{car} + \beta d_{ij}^{pt}. \quad (12)$$

The deterministic utilities of the residential and employment locations vr_i and ve_j can, of course, be functions of characteristics of these locations – some of them potentially unobserved – but we will not delve into that issue. We only observe that, once (2) is accepted, one can estimate them non-parametrically as constants referring to residential and employment locations, respectively. The only issue that remains is the specification of $t(d_{ij})$. If that term is specified correctly, one the ξ_{ij}'s would be identically equal to 0 for all combinations of residential and employment locations and there would be no difference between the model based on a full set of alternative-specific constants and one that specifies the deterministic utilities as

$$v_{ij} = vr_i + ve_j + t(d_{ij}) \quad (13)$$

However, if $t(d_{ij})$ is not correctly specified the ξ_{ij}'s will not all be equal to zero. Moreover, the estimated values of vr_i and ve_j may become biased.

Re-estimating the model on a subset of alternatives may then result in different values of some vr_i's, ve_j's or of the parameters of $t(d_{ij})$.

The discussion just given is based on two-step estimation of the model, but when maximum likelihood (ML) estimation of a logit model based on specification (13) is used, the issues are similar. The ML estimates of the deterministic utilities of the residential and employment locations are such that the predicted probability that a particular residential or employment location is chosen, is identically equal to the observed relative frequency that it is chosen. Misspecification of $t(d_{ij})$ thus implies the possibility that re-estimation of the model on a subset of choice alternatives results in different values of some or all of these deterministic utilities or of the parameters of $t(d_{ij})$.

Using the Hausman-McFadden specification test based on re-estimation of the model on subsets of choice alternatives thus provides useful information about the appropriateness of the choice specification for distance friction, although it is uninformative about the validity of IIA.

3.4 Estimating the model for two periods and compare the predicted results

The MNL for commuting flows plays a crucial role in the computation of counterfactuals in quantitative economic geography models. The reactions of workers to changes in the attractiveness of particular residential or work locations determine the new equilibrium. It is well known that in the logit model these reactions are determined in a mechanistic way by the choice probabilities. In the commuting model considered here, with deterministic utilities as in (2) and choice probabilities determined as in (4) we have:

$$\frac{\partial \pi_{ij}}{\partial vr_k} = \pi_{ij} I(i = k) - \pi_{ij}\pi_{k*} \quad (14)$$

$$\frac{\partial \pi_{ij}}{\partial ve_l} = \pi_{ij} I(j = l) - \pi_{ij}\pi_{*l} \quad (15)$$

where $\pi_{k*} = \sum_j \pi_{kj}$ and $\pi_{*l} = \sum_i \pi_{il}$ and the coefficient τ has been suppressed because it is in practice absorbed in the estimates for vr_i and ve_j. The important point is that the substitution behavior is linked to the choice frequencies in (14) and (15) in a mechanistic way that leaves no role for the estimated parameters.

With data on commuting flows in two different periods, say t and τ, we can test the validity of the predictions of the MNL in the following way:

1 Estimate the MNL for a given metropolitan area in two time periods. We thus estimate all coefficients of the model separately for the two periods.

2 We use the estimated models to derive predictions of the *change* in the commuting flows between each origin and destination. There will, of course, only be predicted changes if there are changes in the estimated deterministic utilities of some residential or work locations, or in the coefficients of the distance function $t(.)$. The predicted changes are compared with the observed changes in the commuting flows.

If the predictions and realizations are close, we cannot reject the model. If they are different, we should.[5] This test will be especially meaningful if the estimated models for both periods fit the data closely, and therefore it is important to check this first, using for instance the Hausman-McFadden methodology discussed in 3.1.

4 Data and results

4.1 Data

The data we use are flows of commuters in the Greater Copenhagen Area which consist of 20 municipalities. We know the residential and work locations of these commuters and treat them as a single type. That is, we do not distinguish them on the basis of age education or any other characteristic. This corresponds to current practice in quantitative economic geography models.

The table on the next page gives an example of the matrix of commuting flows for the year 2008. We have them also for the years 2009-2013. Moreover, we have travel times by car and public transport for commutes between the 20 municipalities.

4.2 Estimating the model on subsets

[5] We do not yet have a fully developed methodology for the test, but the idea is clear and the comparison relatively easy to make.

Table 4.1 Commuting flows in the Greater Copenhagen Area in 2008

	Work																				
Res	101	147	151	153	155	157	159	161	163	165	167	169	173	175	183	185	187	190	230	240	Total
101	172822	23845	4145	3713	2473	13121	8030	2099	2707	2531	7657	4374	6742	5024	2119	8215	1642	4098	6588	3510	285455
147	14651	10820	445	466	154	967	815	279	317	301	1020	459	559	886	180	532	184	389	544	357	34325
151	6345	1302	8468	552	108	885	1367	531	1343	565	708	1007	820	754	334	362	261	1368	778	2926	30784
153	4175	778	493	4230	110	353	473	792	271	660	1063	948	281	689	504	321	489	242	299	431	17602
155	710	0	0	0	1556	0	0	0	0	0	0	0	0	0	0	357	0	0	0	0	2623
157	9868	1575	533	268	103	10259	2318	184	409	189	437	281	1837	416	127	300	119	644	1380	434	31681
159	6637	1178	1061	326	0	1386	8930	312	1122	288	557	447	1300	577	199	252	155	1053	806	857	27443
161	3241	693	588	1102	0	435	528	2844	341	1226	605	1031	348	612	394	252	404	291	287	402	15624
163	3131	535	1327	261	0	453	1210	327	3772	259	316	365	382	493	138	117	119	489	358	788	14840
165	3026	560	536	733	0	303	423	742	258	4148	598	1209	258	526	351	250	493	203	207	450	15274
167	5719	1014	332	960	139	380	432	402	197	375	7688	669	260	878	458	432	318	205	224	371	21453
169	4000	772	585	614	0	383	456	447	247	881	669	8742	332	477	696	316	334	321	282	560	21114
173	6542	1200	618	235	0	1914	2575	218	499	177	381	300	8316	374	139	208	124	772	2145	482	27219
175	3428	600	404	632	0	252	410	437	286	365	856	414	198	4767	161	196	158	166	180	294	14204
183	1056	166	110	283	0	0	116	121	0	175	311	519	0	113	3030	0	195	0	0	0	6195
185	8223	844	196	298	1337	463	296	166	124	147	540	292	226	257	177	7555	121	153	209	202	21826
187	535	0	0	205	0	0	0	0	0	215	115	163	0	0	199	0	1092	0	0	0	2524
190	1684	289	703	0	0	219	527	0	243	0	108	123	267	131	0	0	0	5507	460	716	10977
230	4006	799	350	139	0	1047	827	136	216	101	240	193	1682	204	0	120	0	1011	8392	464	19927
240	936	163	709	0	0	0	171	0	116	0	0	130	0	103	0	0	0	237	0	5843	8408
Total	260735	47133	21603	15017	5980	32820	29904	10037	12468	12603	23869	21666	23808	17281	9206	19785	6208	17149	23139	19087	629498

Hausman & McFadden (1984) suggested to test the specification of MNL by a comparison of the coefficients estimated on (i) all observations and (ii) a subset of the observations in which choices referring to a limited number of alternatives are left out. As discussed above, this test gives potentially useful information about the appropriateness of the specification of the function $t(d_{ij})$ that represents the impact of the spatial separation of the residential and employment locations. One would then compare the estimated coefficients that are common to the unrestricted and restricted model, following the recipe of these authors.

In the first column of Table 4.1 we present the estimates of the unrestricted MNL model. Both travel time coefficients have the expected negative sign. Moreover, they are estimated quite precisely. Model [2] shows the estimates when we re-estimate the model after deletion of the central municipality, i.e. City of Copenhagen. Thus we re-estimate the model on the commuting flows for the subset of households that have residence or job location (or both) outside the City of Copenhagen. If the MNL is specified correctly, the estimated coefficients should remain approximately the same assuming that the unrestricted model is true. Specification error could be related to the fact that home-work trips that start or end in the central municipality take place under more congested traffic conditions and are therefore experienced as more stressful, at given travel times. The table shows that coefficients for travel times increase in absolute values by 11% and 24% for travel time with car and travel time with public transport, respectively, which seems to be at odds with the alternative hypothesis just discussed. The Hausman and McFadden (1984) test value is $T = 172$. Under the null hypothesis T is distributed as χ^2 with 2 degrees of freedom. We reject the MNL specification at beyond the 99 percent critical level.

In the last two columns of Table 4.1 we report two additional Hausman and McFadden tests where we have, in model [3] eliminated one suburban municipality (Egedal), and in model [4] three neighboring suburban municipalities (Egedal, Furesø and Rudersdal). We find now negative test values for both model [3] and model [4]. Hausman and McFadden (1984) report that they often found negative T statistics due to the lack of positive semidefiniteness in finite sample applications. Moreover, they show that replacement by the alternative covariance matrix always leads to a small positive number but they found this alternative statistics to be far away from any reasonable critical values for a χ^2 test (McFadden, 1984, pp. 1226).[6] Therefore, we cannot use the Hausman and McFadden test to accept or reject the MNL specification. However, since the

Table 4.1. MNL estimates on subsets of choice alternatives

	[1] Unrestricted estimate	[2] Restricted estimate 1	[3] Restricted estimate 2	[4] Restricted estimate 3
Travel time with car (minutes)	-0.1277	-0.1416	-0.1358	-0.1428
	(0.0004)	(0.0008)	(0.0005)	(0.0006)
Travel time with public transport (minutes)	-0.0225	-0.0279	-0.0196	-0.0209
	(0.0002)	(0.0003)	(0.0003)	(0.0004)
Log likelihood	-2,582,066	-1,222,917	-2,436,585	-2,090,784
H&M (1984) test statistic		172.070	<0	<0
Number of obs.	619,484	241,117	598,898	547,968

Notes: Standard errors are in parentheses. Restricted estimate 1 is without core municipality (Copenhagen), restricted estimate 2 excludes Egedal municipality, and restricted estimate 3 excludes three municipalities: Furesø, Rudersdal and Egendal.

[6] Small & Hsiao (1985) also report the computation difficulties when the covariance matrix is almost singular.

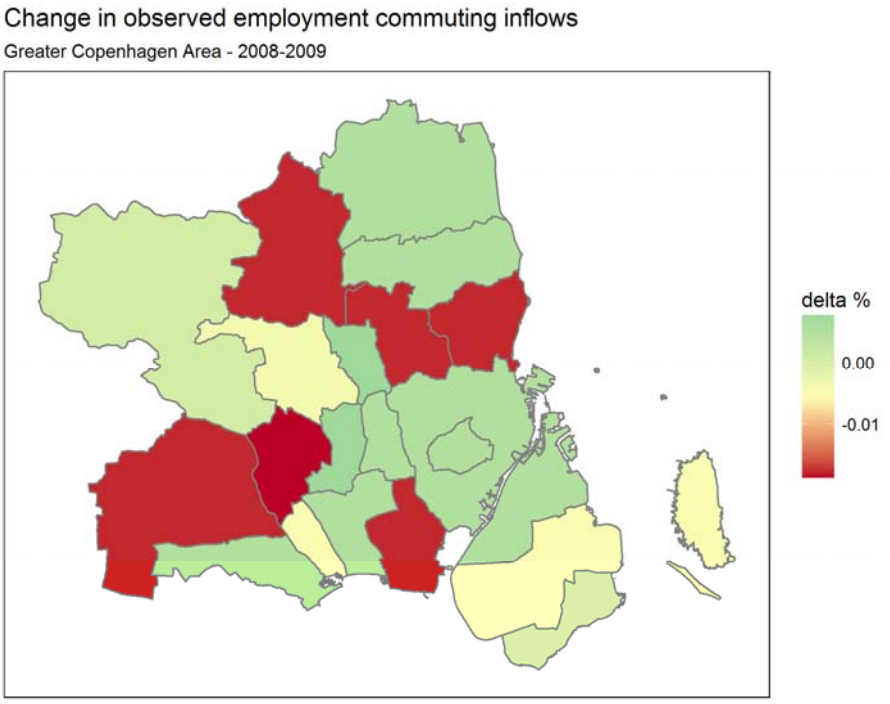
first specification tests lead to rejection of the MNL specification, we may conclude that the MNL specification of the impact of spatial separation on the commuting flows is likely incorrect.

4.3 Predicted and actual changes in commuting flows

As argued in the previous section, we regard the comparison of actual and predicted changes in commuting flows in which the attractiveness of one or more residential or employment locations change as the best test for the adequateness of the MNL as a model for commuting behavior. The reason is that the potentially unrealistic implications of the model for substitution behavior have immediate implications for the computation of counterfactuals in quantitative economic geography models that use gravity equations for commuting interactions.

We estimated the MNL (4) with deterministic utilities as in (2) and the friction of distance as specified in (13) and estimated it for the years 2008-2013. At the moment of submitting this paper for UEA 2017 we have not yet finished the analysis. We can only show some preliminary results. The two maps on the next page show the changes in employment commuting inflows and residential commuting outflows for the 20 Copenhagen municipalities between 2008 and 2009. We also have them for 2009-2010…2012-2013. The formulation of the MNL, with a separate constant representing the attractiveness of every residential and employment location in each year, ensures that the models will exactly predict the total numbers of commutes originating from any residential location or ending in any employment location. The changes in the predicted total flows thus corresponds to changes in the numbers of jobs and residences. The maps suggest that between the years 2008 and 2009 most of the action was in the employment locations, which is possibly related to the volatile economic developments at the time. The pictures for the other years are different.

The fact that the estimated models predict the total number of commutes ending in a particular employment zone exactly, implies that *on average* the predicted change in commuting flows ending up in a particular employment zone will be exactly equal to the observed change in the commuting flows ending in that zone. And that a similar statement holds for flow originating from a particular residential location. This leaves open the possibility of potentially substantial and systematic discrepancies for specific commuting flows. For instance, if there is correlation between the idiosyncratic utilities of commutes that end up in the same employment location (as in a nested logit model with the employment locations first in the utility branch), then a change in the attractiveness of a particular residential location will result in fewer job changes than would be predicted by the MNL and we would see a systematic pattern in the differences between the predicted and actual changes in commuting flows. This is just one possibility, there exist many other and it is the purpose of this exercise to investigate these discrepancies and evaluate their significance. We are currently working on this issue.



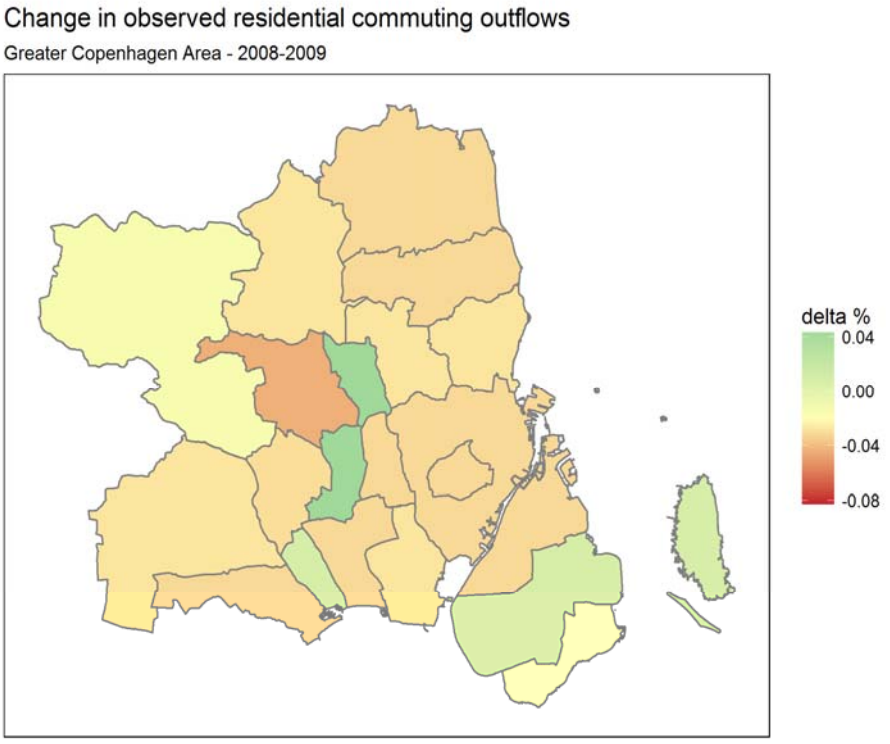


Figure 1 Changes in total commuting flows into and from municipalities

5 A revealed preference approach to the commuting model

In this section we briefly describe another way of approaching specification (2) of the deterministic part of the utility function. We can adopt the two-stage approach familiar from BLP. With a single type of commuters, the first stage would imply estimation of a full set of alternative-specific constants $asc_{ij}, i = 1 \dots I, j = 1 \dots J$. In the second stage they are the dependent variable in (11). Suppose that we have information about the distance or travel time between each pair (i,j), denoted as d_{ij} and that we know that t_{ij} is an increasing function, t, of d_{ij}:

$$d_{ij} > d_{kl} \;\Leftrightarrow\; t_{ij} = t(d_{ij}) > t(d_{kl}) = t_{kl} \qquad (16)$$

The this implies a number of inequalities for the asc_{ij}'s that can be used to investigate the function $t(d)$. In particular, it may be possible to choose the alternative-specific constants vr_i and ve_j in such a way that the (weighted) number of violations of (16) is minimized. The values of the alternative-specific constants then imply those of $t(d_{ij})$ for each commute.

6 Conclusion

In this paper we have considered the MNL as a model for commuting interactions in metropolitan regions. We used a specification in which the deterministic part of the utility attached combinations of residential and work locations is specified as the sum of two constants, representing the utilities of the residential and the work locations, respectively, and a term that reflects the friction of distance. The potential restrictiveness of the MNL results from the fact that it seems plausible that the random part of the utility is specified similarly, which suggests correlation between the random parts of utilities of alternatives sharing the residential or employment location. The MNL rules out such correlation because of its IIA property.

We considered several alternatives and found that the nested logit model, which is the obvious candidate for a more general model, is indistinguishable from the MNL with the chosen specification of the deterministic utilities. The cross nested logit can be estimated, but does not have a particularly attractive specification for the purpose of modelling commuting choices. The best alternative appears to be the mixed logit model.

A fundamental limitation of all the models considered here is that they can never do better than a multinomial logit model with a full set of alternative specific constants.

Appendix

A The equivalence of the Eaton-Kurtum discrete choice model and MNL.

The key observation is that if X is Frechet (EV type II) distributed, $\ln X$ is Gumbel (EV type I) distributed

B The indeterminateness of the logsum parameter in the nested logit models.

We show that the logsum coefficient is not identified in the nested logit version of the commuting model discussed in the main text when using maximum likelihood estimation. Recall that the choice probabilities are given by (7), which we rewrite here as:

$$\pi_{ij} = e^{v_{ij}/\tau}\left\{\sum_{j'} e^{v_{ij'}/\tau}\right\}^{\frac{\tau}{\rho}-1} \Big/ \sum_{i'}\left\{\sum_{j'} e^{v_{ij'}/\tau}\right\}^{\frac{\tau}{\rho}}$$

while the deterministic part of the utility is given by (12). The loglikelihood L is the sum of the logarithm of the appropriate choice probabilities over all observations:

$$L = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \ln \pi_{ij}$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \left(\frac{v_{ij}}{\tau} + \left(\frac{\tau}{\rho} - 1\right) \ln \sum_{j'} e^{\frac{v_{ij'}}{\tau}} - \ln \sum_{i'} \left\{ \sum_{j'} e^{\frac{v_{ij'}}{\tau}} \right\}^{\frac{\tau}{\rho}} \right)$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \frac{v_{ij}}{\tau} + \left(\frac{\tau}{\rho} - 1\right) \sum_{i=1}^{I} n_{i*} \ln \sum_{j'} e^{\frac{v_{ij'}}{\tau}} - \mathrm{n} \ln \sum_{i'} \left\{ \sum_{j'} e^{\frac{v_{ij'}}{\tau}} \right\}^{\frac{\tau}{\rho}},$$

where n_{ij} is the number of workers commuting from residential location i to work location j, n_{i*} is the number of workers residing in i ($n_{i*} = \sum_i n_{ij}$) and n is the total number of workers.

We cannot identify ρ and τ separately, but only the 'logsum' parameter $\frac{\tau}{\rho}$. We simplify the notation a bit by defining $\lambda = \frac{\tau}{\rho}$. All alternative-specific constants and the travel time parameter are estimated relative to ρ and to avoid notational clutter, we proceed as if $\rho = 1$.[7] Using these conventions, we rewrite the loglikelihood as:

$$L = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \frac{v_{ij}}{\lambda} + (\lambda - 1) \sum_{i=1}^{I} n_{i*} \ln \sum_{j'} e^{v_{ij'}/\lambda} - \mathrm{n} \ln \sum_{i'} \left\{ \sum_{j'} e^{v_{i'j'}/\lambda} \right\}^{\lambda}.$$

Note also that we have to choose reference residential and work locations for which we set the alternative-specific constants equal to zero. Choosing location 1 as such, we estimate: $I - 1$ alternative-specific constants for residential location $vr_i, i = 2 \ldots I$, $J - 1$ alternative-specific constants for work locations $ve_j, j = 2 \ldots J$, the travel time parameter γ and the logsum parameter λ.

We now consider the first derivatives of the logs of the choice probabilities, which we rewrite as: $\ln \pi_{ij} = \frac{v_{ij}}{\lambda} + (\lambda - 1) \ln \sum_{j'} e^{v_{ij'}/\lambda} - \ln \sum_{i'} \left\{ \sum_{j'} e^{v_{i'j'}/\lambda} \right\}^{\lambda}$:

$$\frac{\partial \ln \pi_{ij}}{\partial vr_k} = \delta_{i=k} \left(\frac{1}{\lambda} + \frac{\lambda - 1}{\lambda} \sum_{j'} \pi_{j'|i} \right) - \pi_{k*}$$
$$= \delta_{i=k} - \pi_{i*}. \qquad \text{(B.1)}$$

Here $\delta_{i=k} = 1$ if $i = k$ and 0 otherwise, $\pi_{j'|i}$ is the probability of choosing work location j conditional on the choice of residential location i, and π_{i*} is the probability that residential location i will be chosen ($\pi_{i*} = \sum_j \pi_{ij}$). The simplification in the second line follows because $\sum_{j'} \pi_{j'|i} = 1$.

$$\frac{\partial \ln \pi_{ij}}{\partial ve_l} = \frac{\delta_{j=l}}{\lambda} + \frac{\lambda - 1}{\lambda} \pi_{l|i} - \sum_{i'} \pi_{i'*} \pi_{l|i'}$$
$$= \frac{1}{\lambda} \left(\delta_{j=l} + (\lambda - 1) \pi_{l|i} - \lambda \sum_{i'} \pi_{i'*} \pi_{l|i'} \right) \qquad \text{(B.2)}$$

$$\frac{\partial \ln \pi_{ij}}{\partial \gamma} = \frac{tt_{ij}}{\lambda} + (\lambda - 1) \sum_{j'} \pi_{j'|i} \frac{tt_{ij'}}{\lambda} - \sum_{i'} \pi_{i'*} \sum_{j'} \pi_{j'|i'} tt_{i'j'}$$

[7] One may keep in mind that the alternative-specific constants and the travel time parameter are estimated after dividing them by the (unknown) value of ρ, which may differ from 1.

$$= \frac{1}{\lambda}\left(tt_{ij} + (\lambda - 1)\sum_{j'} \pi_{j'|i}\, tt_{ij'} - \lambda \sum_{i'} \pi_{i'*} \sum_{j'} \pi_{j'|i'} tt_{i'j'}\right) \quad \text{(B.3)}$$

$$\frac{\partial \ln \pi_{ij}}{\partial \lambda} = -\frac{v_{ij}}{\lambda^2} + +(\lambda - 1)\sum_{j'} \pi_{j'|i}\left(-\frac{v_{ij}}{\lambda^2}\right) - \sum_{i'} \pi_{i'*}\left[\ln \sum_{j'} e^{\frac{v_{ij'}}{\lambda}} + \lambda \sum_{j'} \pi_{j'|i'}\left(-\frac{v_{ij}}{\lambda^2}\right)\right]$$
$$= \left(-\frac{1}{\lambda^2}\right)\left(v_{ij} + \lambda^2 ls_i + (\lambda - 1)\sum_{j'} \pi_{j'|i}\, v_{ij'} + \lambda^2 \sum_{i'} \pi_{i'*} ls_{i'} - \lambda \sum_{i'} \pi_{i'*} \sum_{j'} \pi_{j'|i'} v_{i'j'}\right) \quad \text{(B.4)}$$

where $ls_i = \ln \sum_{j'} e^{\frac{v_{ij'}}{\lambda}}$, the logsum associated with residential location i.
Using these results, we find:

$$\left(\frac{\partial L}{\partial vr_k} = 0 \Longrightarrow\right) n_{k*} - n\,\pi_{k*} = 0. \quad \text{(B.5)}$$

$$\left(\frac{\partial L}{\partial ve_l} = 0 \Longrightarrow\right) n_{*l} + (\lambda - 1)\sum_i n_{i*}\,\pi_{l|i} - n\lambda \sum_i \pi_{i*}\pi_{l|i} = 0. \quad \text{(B.6)}$$

$$\left(\frac{\partial L}{\partial \gamma} = 0 \Longrightarrow\right) \frac{1}{\lambda}\left(\sum_i \sum_j n_{ij} tt_{ij} + (\lambda - 1)\sum_i n_i \sum_{j'} \pi_{j'|i}\, tt_{ij'} - n\lambda \sum_{i'} \pi_{i'*} \sum_{j'} \pi_{j'|i'} tt_{i'j'}\right) = 0 \quad \text{(B.7)}$$

$$\left(\frac{\partial L}{\partial \lambda} 0 \Longrightarrow\right) \left(-\frac{1}{\lambda^2}\right)\left\{\sum_i \sum_j n_{ij} v_{ij} + \lambda^2 \sum_i n_{i*} ls_i + (\lambda - 1)\sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, v_{ij'} + \lambda^2 n \sum_{i'} \pi_{i'*} ls_{i'} - \lambda n \sum_{i'} \pi_{i'*} \sum_{j'} \pi_{j'|i'} v_{i'j'}\right\} = 0 \quad \text{(B.8)}$$

We will show that the expression in curly brackets in (B.8) is identically 0 if (B.5), (B.6) and (B.7) hold. We do so as follows:

i) (B5) implies $\pi_{k*} = n_{k*}/n$. This result will be used repeatedly in what follows. Hence $\sum_i n_{i*} ls_i = n \sum_{i'} \pi_{i'*} ls_{i'}$. Hence the two terms with the logsums cancel out.

ii) For the same reason: $\sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, v_{ij'} = n \sum_i \pi_{i*} \sum_{j'} \pi_{j'|i}\, v_{ij'}$. Hence the terms with a λ in front of them cancel out.

These simplifications allow us to rewrite the expression in curly brackets in (B.8) as:
$\sum_i \sum_j n_{ij} v_{ij} - \sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, v_{ij'}$
After substitution of the deterministic part of the utility function, this can be rewritten as:
$$\left(\sum_i \sum_j n_{ij} vr_i - \sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, vr_i\right) + \left(\sum_i \sum_j n_{ij} ve_j - \sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, ve_{j'}\right) + \gamma\left(\sum_i \sum_j n_{ij} tt_{ij} - \sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, tt_{ij'}\right) \quad \text{(B.9)}$$

We now show that each term in parentheses equals 0:

iii) $\sum_i \sum_j n_{ij} vr_i = \sum_i n_{i*}\, vr_i$ and $\sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, vr_i = \sum_i n_{i*}\, vr_i$. Hence the first term equals 0.

iv) $\sum_i n_{i*}\,\pi_{l|i} = n \sum_i \pi_{i*}\,\pi_{l|i}$. It follows thus from (B.6) that $n_{*l} = \sum_i n_{i*}\,\pi_{l|i}$. After some rearrangement of the second term of the second expression in curly brackets, we can substitute this result in it. We then find that this second term equals $\sum_j n_{*j}\, ve_j$. This is identical to the first term. Hence the second term in parentheses must also be equal to 0.

v) $\sum_i n_{i*} \sum_{j'} \pi_{j'|i}\, tt_{ij'} = n \sum_i \pi_{i*} \sum_{j'} \pi_{j'|i}\, tt_{ij'}$. Hence we can simplify (B.7) to the third term, which must thus be equal to 0.

Since all three terms in (B.9) are zero, the expression in curly brackets in (B.8) must be equal to 0 as well.
The derivations for the other type of nested logit model are completely analogous.

C The determinateness of the parameters for the random term the mixed logit model.

We start from (11) and write the loglikelihood that the combination (i,j) is chosen as:

$$ll_{ij} = \ln\left(\int_\theta \int_\varphi \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right) \quad \text{(C.1)}$$

with $\pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi}) = \frac{e^{(v_{ij}+\rho_i\theta_i+\sigma_j\varphi_j)/\tau}}{\sum_{i'}\sum_{j'} e^{(v_{i'j'}+\rho_{i'}\theta_{i'}+\sigma_{j'}\varphi_{j'})/\tau}}$ as in (11). The first derivatives with respect to vr_k and ρ_k are:

$$\frac{\partial ll_{ij}}{\partial vr_k} = \frac{1}{Pr_{ij}}\left\{\int_\theta \int_\varphi \left[I(i=k)\pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi}) - \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})\right]g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\}$$
$$= I(i=k) - \frac{1}{Pr_{ij}}\left\{\int_\theta \int_\varphi \left[\pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})\right]g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\} \quad \text{(C.2)}$$

where $Pr_{ij} = \int_\theta \int_\varphi \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}$ and $\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi}) = \sum_j \pi_{kj}(\boldsymbol{\theta},\boldsymbol{\varphi})$, and:

$$\frac{\partial ll_{ij}}{\partial \rho_k} = \frac{1}{Pr_{ij}}\left\{\int_\theta \int_\varphi \left[I(i=k)\pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi}) - \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})\right]\theta_k g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\}$$
$$= \frac{1}{Pr_{ij}}\left\{\int_\theta \int_\varphi I(i=k)\pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\theta_k g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\} -$$
$$\frac{1}{Pr_{ij}}\left\{\int_\theta \int_\varphi \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})\theta_k g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\} \quad \text{(C.3)}$$

For the likelihood of the whole sample, $LL = \sum_i \sum_j ll_{ij}$ this gives:

$$\frac{\partial LL}{\partial vr_k} = f_{i*} - \sum_i \sum_j \frac{f_{ij}}{Pr_{ij}}\left\{\int_\theta \int_\varphi \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\} \quad \text{(C.4)}$$

$$\frac{\partial LL_{ij}}{\partial \rho_k} = \sum_j \frac{f_{kj}}{Pr_{kj}}\left\{\int_\theta \int_\varphi \pi_{kj}(\boldsymbol{\theta},\boldsymbol{\varphi})\theta_k g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\} -$$
$$\sum_i \sum_j \frac{f_{ij}}{Pr_{ij}}\left\{\int_\theta \int_\varphi \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})\theta_k g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\} \quad \text{(C.5)}$$

where f_{ij} is the number of commuters from i to j.Rouwendal (2017) dhows that if we could set the alternative-specific constants[8] so that: $Pr_{ij} = f_{ij}/F$, where $F = \sum_i \sum_j f_{ij}$ is the total number of commuters – which is in general not the case – it can be shown that (C.4) and (C.5) are identically equal to zero. This works as follows After substitution of the condition in (C.4), we can rewrite this equation as:

$$\frac{\partial LL}{\partial vr_k} = f_{i*} - F\left\{\int_\theta \int_\varphi \left(\sum_i \sum_j \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\right)\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta}\right\} \quad \text{(C.6)}$$

Which is equal to 0, since $\sum_i \sum_j \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi}) = 1$.

Substitution in (C.5) gives:

$$\frac{\partial LL_{ij}}{\partial \rho_k} = F\int_\theta \int_\varphi \pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})\theta_k g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta} -$$
$$F\int_\theta \int_\varphi \left(\sum_i \sum_j \pi_{ij}(\boldsymbol{\theta},\boldsymbol{\varphi})\right)\pi_{k*}(\boldsymbol{\theta},\boldsymbol{\varphi})\theta_k g(\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi}d\boldsymbol{\theta} \quad \text{(C.7)}$$

which is identically 0.

[8] Not only the vr_i's, but also the ve_j's could be used for that purpose. However, there are $I + J$ such constants, whereas we have IJ chouice alternatives.

If the condition is violated, the equations do not simplify and the result does not hold.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2016). The economics of density: Evidence from the Berlin wall. *Econometrica*, *83*(6), 1217–2189.

Allen, T., Arkolakis, C., & Li, X. (2015). Optimal city structure. *Working Paper*.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779. http://doi.org/10.1111/1468-0262.00352

Hamilton, B. W. (1982). Wasteful commuting. *Journal of Political Economy*, *90*(5), 1035–1053. http://doi.org/10.1086/261665

Hausman, J. A., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*. http://doi.org/10.2307/1910997

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (pp. 105–142).

Monte, F., Redding, S., & Rossi-Hansberg, E. (2015). Commuting, migration, and local employment elasticities. *NBER Working Paper*, *21706*, 1–57.

Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, *9*, 1–47.

Rouwendal, J. (1998). Search theory, spatial labor markets, and commuting. *Journal of Urban Economics*, *43*(1), 1–22.

Rouwendal, J. (2017). *Specification tests for the multinomial logit model revisited*.

Small, K. A., & Hsiao, C. (1985). Multinomial logit specification tests. *International Economic Reviewc Review*, *26*(3), 619–627. http://doi.org/10.2307/2526707

Small, K. A., & Song, S. (1992). " Wasteful " Commuting : A Resolution. *Journal of Political Economy*, *100*(4), 888–898. http://doi.org/10.1086/261844