Modeling Inter-Regional Migration in Western Germany

Herman J. Bierens
Department of Economics, Pennsylvania State University, and Tilburg University, the Netherlands

Thomas Kontuly
Department of Geography, University of Utah

February 2002

Abstract

In this paper we investigate the degree to which the internal migration flows in Western Germany can be related to employment changes. The data consist of in- and out-migration of employed workers and employment in 75 regions in Western Germany over the period 1982-1997. Given a multinomial logit model for the individual migration decisions, we derive a conditional Poisson gravity model for the migration flows from one region to another. Although these migration flows are not observable, it appears that the in- and out-migration per region are sufficient statistics for maximum likelihood estimation, for each year separately. The correctness of the model specification is tested by the integrated conditional moment test on the basis of nonlinear regressions implied by the Poisson gravity model.

JEL codes: C12, C13, C21, C51, C52, R12

Key words: Migration, Poisson-gravity model, maximum likelihood, nonlinear regression, model specification testing.

1 Introduction

The literature suggests numerous explanations for either a slowing urbanization pattern or the existence of a counter-urbanization tendency in European countries during the 1970s, 1980s, and 1990s. These explanations can be organized into five groups, that is, economic structural change, demographic restructuring, government policy, housing costs, and period effects. Economic structural change refers to the regional deconcentration of manufacturing and service employment, and can be thought of as the "people follow jobs" explanation. Demographic restructuring refers to a change in the preferences of the working-age population in favor of locations in rural, low density, environmentally pleasant areas; this is the "jobs follow people" explanation. Period effects include business cycle fluctuations, regional boom and bust experiences, and changing socio-demographic compositions. The government policy explanation includes planned deconcentration initiatives to redistribute jobs and people to rural / peripheral areas, and includes increases in public service employment such as, child care, education for the young, and other services. Housing costs reasons include both the high cost and the availability of housing (Kontuly 1998; Kontuly and Dearden 1998).

Several researchers believe that the first three explanations provide the basis for understanding counter-urbanization and slowing urbanization in the developed world (Fielding 1989; Frey 1987, 1988, 1989, 1993; Hugo and Smailes 1985; Kontuly and Vogelsang 1988; Morrill 1978; Moseley 1984; Plane and Rogerson 1991). Kontuly and Vogelsang (1988) contend that government policy must be added to these three major explanations.

Economic structural change and demographic restructuring explain long-term tendencies in the human settlement patterns of developed countries. Structural change explanations viewed the de-industrialization of the 1970s as a short-term phenomenon, and forecasted new urbanization tendencies for large metropolitan areas that were able to function as advanced private service centers, the locations of corporate headquarters, banks, and other financial institutions, and for metropolitan areas that became the location of knowledge-based industries. Demographic restructuring refers to a change in residential preferences or in the ability to act on such preferences - a basic change either in the preferences of the working-age population or in their ability to act on such preferences, in favor of residences in rural or small-town environments and against large cities.

Period explanations, which account for short-term change, attributed the 1970s migration turnaround to the unique economic and demographic circumstances of the 1970s, and implied that when these economic and demographic shocks subsided, more traditional urbanization tendencies would re-emerge. The changing socio-demographic compositions explanation included the influence of relative age cohort sizes such as the proportion of the population in the Baby Boom generation and the proportion of the population that is elderly. The emergence of large birth cohorts during the 1950s and early 1960s impacted labor markets and the spatial structure of young adult migration flows became more geographically focused. The aging of the population has meant a rapid increase in the numbers of economically inactive people and a growing pool of potential migrants. Retirement enables those with sufficient resources to realize long-held desires for more realized life-styles in attractive, uncongested surroundings.

Structural economic change is occurring as the proportion of tertiary and quaternary employment increases relative to secondary employment. Also, the decline in primary employment has almost run its course, so there is a reduction in the stock or potential out-migrants living in rural areas. Older industrial countries have been going through a process of de-industrialization that has had a strong negative impact on larger cities, especially on their central areas. Fielding (1989) expands on this explanation by emphasizing the evolution of a new spatial division of labor. He argues that the growth of companies in to multi-plant, multi-product, and multi-national enterprises was accompanied by disinvestment in high-wage areas in favor of low-wage reserves of labor within national territories. The deconcentration of jobs was followed by population migration.

Economic structural change is one of the major factors suggested in the literature as an explanation for a slowing urbanization pattern or the existence of a counter-urbanization tendency in European countries during the 1970s and 1980s. Economic structural change was cited in the literature as an important explanation in eleven out of fourteen countries surveyed. These countries were Belgium, Great Britain, Denmark, Finland, France, Western Germany, Italy, the Netherlands, Norway, Sweden, and Switzerland (Kontuly 1998).

Renewed interest in internal migration processes in developed countries and the reasons behind the changes in these processes was stimulated by the discovery of a "new migration turnaround" in the USA during the 1990s (Fuguitt and Beale 1996; Johnson 1999; Long and Nucci 1997). A similar "new" turnaround was evident in western Germany during the period 1992-1996 (Kemper 1999).

In this paper we evaluate the importance of the economic structural change hypothesis, by determining the degree to which the internal migration flows in Western Germany can be related to employment changes, on the basis of the Poisson gravity model proposed by Flowerdew and Aitken (1982).

2 The data

For $j = 1, ..., N = 75$ regions in West Germany and years $t = 1982, ..., 1997$ we observe:

$$\begin{aligned} y_{j,t}^{-} &= \text{out-migration of employed workers,} \\ y_{j,t}^{+} &= \text{in-migration of employed workers,} \\ X_{j,t} &= \text{total employment.} \end{aligned}$$

If the migration data involved would concern genuine **internal** migration, then for each year the net migration should add up to zero:

$$\sum_{j=1}^{N} y_{j,t}^{-} - \sum_{j=1}^{N} y_{j,t}^{+} = 0. \qquad (1)$$

However, in reality this is not the case. Nevertheless, for the time being we will treat the data as if (1) holds. In Appendix 1 we will set forth mild regularity conditions such that, asymptotically, the econometric results derived under the restriction (1) carry over to the case where (1) is false.

In this paper we propose a multinomial logit model for the individual migration decisions, on the basis of which we derive a conditional Poisson model for the migration, $y_{i|j}(t)$, from region j to region $i \neq j$:

$$\begin{aligned} y_{i|j}(t) &= \text{total number of employed workers (exclusive family members)} \\ &\quad \text{migrating from region } j \text{ to region } i \text{ in year } t, \end{aligned}$$

Then the out-migration of employed workers (exclusive family members) from region j is

$$y_{j,t}^{-} = \sum_{i=1}^{N} I(i \neq j) y_{i|j}(t)$$

where $I(\bullet)$ is the indicator function[1], and the in-migration to region i is

$$y_{i,t}^{+} = \sum_{j=1}^{N} I(i \neq j) y_{i|j}(t),$$

so that the restriction (1) holds. Although the internal migration flows $y_{i|j}(t)$ are not observed, it turns out that the $y_{j,t}^{-}$'s and $y_{i,t}^{+}$'s are sufficient statistics for maximum likelihood estimation.

3 A multinomial logit model for individual migration decisions

Since moving takes time, we assume that the decision to migrate takes effect in the next year. Thus, let $\mathcal{F}_{N,t-1}$ be the σ-algebra (or information set) generated by $y_{i|j}(t-m)$, and $X_{j,t-m}$ for $i, j = 1, ...N$ and all $m \geq 1$. The model we propose is a model for the conditional probability $p_{i|j}(t)$, given $\mathcal{F}_{N,t-1}$, that in period t a representative **employed** worker in region j will migrate to region $i \neq j$.

First, it is plausible that $p_{i|j}(t)$ is positively related to the difference in net in-migration in regions i and j in the previous year. In particular, if

$$\begin{aligned} & \left(y_{i_1,t-1}^{+} - y_{i_1,t-1}^{-}\right) - \left(y_{j,t-1}^{+} - y_{j,t-1}^{-}\right) \\ < \quad & \left(y_{i_2,t-1}^{+} - y_{i_2,t-1}^{-}\right) - \left(y_{j,t-1}^{+} - y_{j,t-2}^{-}\right), \end{aligned}$$

then region i_2 has apparently been more attractive to migrate to than region i_1, hence the larger the difference

$$\left(y_{i,t-1}^{+} - y_{i,t-1}^{-}\right) - \left(y_{j,t-1}^{+} - y_{j,t-1}^{-}\right),$$

the larger the conditional probability $p_{i|j}(t)$ will be. One may consider this as a "keeping up with the Jones" effect.

Moreover, it will be assumed that the migration decision is also based on the difference in employment growth rates, i.e., $p_{i|j}(t)$ depends on

$$\left(X_{i,t-1} - X_{i,t-2}\right)/X_{i,t-2} - \left(X_{j,t-1} - X_{j,t-2}\right)/X_{j,t-2}, \qquad (2)$$

[1] $I(true) = 1$, $I(False) = 0$.

or (almost) equivalently,[2]

$$(\ln X_{i,t-1} - \ln X_{i,t-2}) - (\ln X_{j,t-1} - \ln X_{j,t-2}).$$

Furthermore, workers living in large regions have more possibilities of changing jobs within the region, hence the probability of migrating from a large region to another region will be smaller than from smaller regions. Consequently, we assume that $p_{i|j}(t)$ depends negatively on $X_{j,t-1}$.

In view of these arguments we propose the following specification of the conditional probability $p_{i|j}(t)$ that in period $t \geq 2$ an employed worker in region j migrates to region $i \neq j$:

$$\begin{aligned} & p_{i|j}(t) \\ = & \frac{X_{j,t-1}^{-1} \exp\left[\alpha_0 + \beta_0' \left(x_{i,t-1}^{\Delta} - x_{j,t-1}^{\Delta}\right)\right]}{1 + X_{j,t-1}^{-1} \sum_{m=1}^{N} I(m \neq j) \exp\left[\alpha_0 + \beta_0' \left(x_{m,t-1}^{\Delta} - x_{j,t-1}^{\Delta}\right)\right]} \\ = & \frac{X_{j,t-1}^{-1} \exp\left[\alpha_0 + \beta_0' \left(x_{i,t-1}^{\Delta} - x_{j,t-1}^{\Delta}\right)\right]}{1 + \frac{N}{X_{j,t-1}} \exp\left[\alpha_0 - \beta_0' x_{j,t-1}^{\Delta}\right] \frac{1}{N} \sum_{m=1}^{N} I(m \neq j) \exp\left[\beta_0' x_{m,t-1}^{\Delta}\right]}, \end{aligned} \tag{3}$$

with[3]

$$x_{j,t-1}^{\Delta} = \begin{pmatrix} \left(y_{j,t-1}^{+} - y_{j,t-1}^{-}\right)/10000 \\ \ln X_{j,t-1} - \ln X_{j,t-2} \end{pmatrix}.$$

Both components of β_0 are expected to be non-negative:

$$\beta_0' = (\beta_{0,1}, \beta_{0,2}), \text{ with } \beta_{0,1} \geq 0, \beta_{0,2} \geq 0.$$

This specification is akin to a multinomial logit specification.

The reason for the factor $X_{j,t-1}^{-1}$ in (3) is two-fold. First, as motivated before, the larger $X_{j,t-1}$, the smaller $p_{i|j}(t)$ will be. Moreover the particular form of the factor $X_{j,t-1}^{-1}$ facilitates the Poisson approximation discussed below.

Since we assume that the decision to move or to stay in period t is made in period $t-1$, we have

$$y_{j|j}(t) = X_{j,t-1} - \sum_{i \neq j} y_{i|j}(t),$$

[2] Because $d\ln(x)/dx = 1/x$, hence $d\ln(x) = x^{-1}dx$.

[3] The reason for the rescaling of $y_{j,t-1}^{+} - y_{j,t-1}^{-}$ is to prevent overflow in the exp(.) function.

where $y_{j|j}(t)$ is the number of workers in region j that stay in region j. Then conditional on $\mathcal{F}_{N,t-1}$, the random variables

$$y_{1|j}(t), ..., y_{j-1|j}(t), y_{j+1|j}(t), ..., y_{N|j}(t) \tag{4}$$

have a multinomial distribution:

$$P\left[\wedge_{i\neq j}\left\{y_{i|j}(t)=y_i\right\}|\mathcal{F}_{t-1}\right]=\frac{X_{j,t-1}!}{y_1!....y_N!}\prod_{i=1}^{N}p_{i|j}(t)^{y_i},$$

where

$$\begin{aligned} y_j &= X_{j,t-1}-\sum_{i\neq j}y_i, \\ p_{j|j}(t) &= 1-\sum_{i\neq j}p_{i|j}(t). \end{aligned}$$

4 Poisson approximation of the multinomial distribution

The general form of the probability function of the multinomial distribution is

$$q\left(y_1,..,y_{k-1}\right)=\frac{n!}{y_1!....y_{k-1}!\left(n-\sum_{j=1}^{k-1}y_j\right)}\left(\prod_{j=1}^{k-1}p_j^{y_j}\right)\left(1-\sum_{j=1}^{k-1}p_j\right)^{n-\sum_{j=1}^{k-1}y_j}.$$

Define

$$\lambda_j=np_j$$

and suppose that $p_j\downarrow 0$ as $n\to\infty$ such that the λ_j's remain constant. Denote

$$z=\sum_{j=1}^{k-1}y_j.$$

Then similarly to the Poisson approximation of the Binomial distribution it follows that for fixed $y_1,..,y_{k-1}$ and $n\to\infty$,

$$q\left(y_1,..,y_{k-1}\right) = \frac{n!}{y_1!....y_{k-1}!\left(n-z\right)!}n^{-z}\left(\prod_{j=1}^{k-1}\lambda_j^{y_j}\right)\left(1-\sum_{j=1}^{k-1}\frac{\lambda_j}{n}\right)^{n-z}$$

$$
\begin{aligned}
&= \frac{1}{y_1!....y_{k-1}!}\prod_{k=1}^{z}\left(1-\frac{z-k}{n}\right)\prod_{j=1}^{k-1}\lambda_j^{y_j} \\
&\quad\times\left(1-\sum_{j=1}^{k-1}\frac{\lambda_j}{n}\right)^n\times\left(1-\sum_{j=1}^{k-1}\frac{\lambda_j}{n}\right)^{-z} \\
&\to \frac{1}{y_1!....y_{k-1}!}\left(\prod_{j=1}^{k-1}\lambda_j^{y_j}\right)\exp\left[-\sum_{j=1}^{k-1}\lambda_j\right] \\
&= \prod_{j=1}^{k-1}\left[\exp\left(-\lambda_j\right)\left(\lambda_j^{y_j}/y_j!\right)\right].
\end{aligned}
$$

Hence, for large n and small p_j's the multinomial distribution is approximately equal to the distribution of $k-1$ **independent** Poisson(λ_j) variates, where $\lambda_j = np_j$.

Thus, assuming that the probabilities $p_{i|j}(t)$ are small and $X_{j,t-1}$ is large, it follows that for fixed j the random variables (4) are approximately independent Poisson$\big(X_{j,t-1}p_{i|j}(t)\big)$ distributed, conditional on $\mathcal{F}_{t-1}$. Moreover, since for $j = 1,..,N$ the random variables (4) are due to the decisions of employed workers in different regions, we may without loss of generality assume that conditional on $\mathcal{F}_{N,t-1}$, the $y_{i|j}(t)$'s are independent Poisson$\big(X_{j,t-1}p_{i|j}(t)\big)$ distributed for all unequal i and j. Furthermore, it follows from (3) that

$$
\begin{aligned}
& X_{j,t-1}p_{i|j}(t) \\
&= \frac{\exp\left[\alpha_0+\beta_0'\left(x_{i,t-1}^{\Delta}-x_{j,t-1}^{\Delta}\right)\right]}{1+\frac{N}{.X_{j,t-1}}\exp\left[\alpha_0-\beta_0'x_{j,t-1}^{\Delta}\right]\frac{1}{N}\sum_{m=1}^{N}I(m\neq j)\exp\left[\beta_0'x_{m,t-1}^{\Delta}\right]} \\
&\approx \exp\left[\alpha_0+\beta_0'\left(x_{i,t-1}^{\Delta}-x_{j,t-1}^{\Delta}\right)\right]
\end{aligned}
$$

where the latter approximation holds if $X_{j,t-1}/N$ is large relative to $\exp(\alpha_0 - \beta_0'x_{j,t-1}^{\Delta})N^{-1}\sum_{m=1}^{N}\exp\left[\beta_0'x_{m,t-1}^{\Delta}\right] - \exp(\alpha_0)/N$. Therefore, adopting the specification (3) we have approximately:

Conjecture 1 *For $i,j = 1,...,N$, $i \neq j$ and each t the $y_{i|j}(t)$'s are independent Poisson$(\eta_{ijt}(\alpha_0,\beta_0))$ distributed, conditional on $\mathcal{F}_{N,t-1}$, where*

$$
\eta_{ijt}(\alpha,\beta) = \exp\left[\alpha+\beta'\left(x_{i,t-1}^{\Delta}-x_{j,t-1}^{\Delta}\right)\right]. \tag{5}
$$

Note that this is just a version of the Poisson gravity model proposed by Flowerdew and Aitken (1982). See also, e.g., Okoruwa, Terza and Nourse (1988) and Okoruwa, Nourse and Terza (1994) for applications.

Moreover, note that the $y_{i|j}(t)$'s are not independent over time, because they depend on past values via $\mathcal{F}_{N,t-1}$. Furthermore, it is unlikely that the parameters (α_0, β_0) are constant over time. But even if they are we cannot estimate them jointly for all t because the time dimension is too small to use laws of large numbers and central limit theorems for dependent random variables. Thus, the parameters α_0 and β_0 will be estimated **for each time period t separately.**

5 Conditional Maximum Likelihood (ML) estimation

Since sums of independent Poisson variates are Poisson distributed themselves, it follows from Conjecture 1 that for $j = 1,..,N$ the $y_{j,t}^{-}$'s are independently Poisson$[\mu_{j,t,N}(\alpha_0, \beta_0)]$ distributed, conditional on $\mathcal{F}_{N,t-1}$, where

$$\begin{aligned} \mu_{j,t,N}(\alpha,\beta) &= \sum_{i=1}^{N} I(i \neq j)\eta_{ijt}(\alpha,\beta) \\ &= \sum_{i=1}^{N} I(i \neq j)\exp\left[\alpha + \beta'\left(x_{i,t-1}^{\Delta} - x_{j,t-1}^{\Delta}\right)\right] \\ &= N\exp(\alpha)\exp(-\beta' x_{j,t-1}^{\Delta})\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta' x_{i,t-1}^{\Delta}\right) - \exp(\alpha), \end{aligned} \tag{6}$$

and similarly, the $y_{j,t}^{+}$'s are independently Poisson$[\upsilon_{j,t,N}(\alpha_0, \beta_0)]$ distributed, conditional on $\mathcal{F}_{N,t-1}$, where

$$\begin{aligned} \upsilon_{j,t,N}(\alpha,\beta) &= \sum_{i=1}^{N} I(i \neq j)\eta_{jit}(\alpha,\beta) \\ &= \sum_{i=1}^{N} I(i \neq j)\exp\left[\alpha - \beta'\left(x_{i,t-1}^{\Delta} - x_{j,t-1}^{\Delta}\right)\right] \\ &= N\exp(\alpha)\exp(\beta' x_{j,t-1}^{\Delta})\frac{1}{N}\sum_{i=1}^{N}\exp\left(-\beta' x_{i,t-1}^{\Delta}\right) - \exp(\alpha). \end{aligned} \tag{7}$$

Note that the pairs $(y^-_{j,t}, y^+_{j,t})$ are not independent, conditional on $\mathcal{F}_{N,t-1}$. Thus, it is possible to estimate the parameters by ML on the basis of either the $y^-_{j,t}$'s or the $y^+_{j,t}$'s., but in that case not all available information will be used. However, as is well-known, for each time period t the parameters α_0 and β_0 can be estimated by ML on the basis of the unobserved migration flows $y_{i|j}(t)$, because the likelihood function depends on the observed $y^-_{j,t}$'s, $y^+_{j,t}$'s and $x^\Delta_{j,t-1}$'s only. In other words, the $y^+_{i,t}$'s and $y^-_{i,t}$'s are sufficient statistics for the maximum likelihood estimation of α_0 and β_0:[4]

$$
\begin{aligned}
\widetilde{L}_t(\alpha,\beta) &= \sum_{i\neq j} y_{i|j}(t)\ln\left(\eta_{ijt}(\alpha,\beta)\right) - \sum_{j=1}^N \mu_{j,t,N}(\alpha,\beta) - \sum_{i\neq j}\ln\left(y_{i|j}(t)!\right) \quad (8)\\
&= \alpha\sum_{i\neq j} y_{i|j}(t) + \beta'\sum_{i\neq j} y_{i|j}(t)x^\Delta_{i,t-1} - \beta'\sum_{i\neq j} y_{i|j}(t)x^\Delta_{j,t-1}\\
&\quad -\sum_{j=1}^N \mu_{j,t,N}(\alpha,\beta) - \sum_{i\neq j}\ln\left(y_{i|j}(t)!\right)\\
&= \alpha\sum_{j=1}^N y^-_{j,t} + \beta'\sum_{j=1}^N\left(y^+_{j,t} - y^-_{j,t}\right)x^\Delta_{j,t-1}\\
&\quad -\sum_{i\neq j}\exp\left[\alpha + \beta'\left(x^\Delta_{i,t-1} - x^\Delta_{j,t-1}\right)\right]\\
&\quad -\sum_{i\neq j}\ln\left(y_{i|j}(t)!\right)
\end{aligned}
$$

The first-order conditions for a maximum of $\widetilde{L}_t(\alpha,\beta)$ are*:*

$$
\begin{aligned}
0 &= \frac{\partial\widetilde{L}_t(\widehat{\alpha},\widehat{\beta})}{\partial\widehat{\alpha}} = \sum_{j=1}^N y^-_{j,t} - \sum_{i\neq j}\exp\left[\widehat{\alpha} + \widehat{\beta}'\left(x^\Delta_{i,t-1} - x^\Delta_{j,t-1}\right)\right] \quad (9)\\
&= \sum_{j=1}^N\left(y^-_{j,t} - \mu_{j,t,N}(\widehat{\alpha},\widehat{\beta})\right),\\
0 &= \frac{\partial\widetilde{L}_t(\widehat{\alpha},\widehat{\beta})}{\partial\widehat{\beta}'} = \sum_{j=1}^N\left(y^+_{j,t} - y^-_{j,t}\right)x^\Delta_{j,t-1}
\end{aligned}
$$

[4] For notational convenience we will write the double sum $\sum_{j=1}^N\sum_{i=1}^N I(i\neq j)$ as $\sum_{i\neq j}$.

$$
\begin{aligned}
&-\sum_{i\neq j}\exp\left[\widehat{\alpha}+\widehat{\beta}'\left(x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\right)\right]\left(x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\right)\\
=\;&\sum_{j=1}^{N}\left(y^{+}_{j,t}-\nu_{j,t,N}(\widehat{\alpha},\widehat{\beta})\right)x^{\Delta}_{j,t-1}-\sum_{j=1}^{N}\left(y^{-}_{j,t}-\mu_{j,t,N}(\widehat{\alpha},\widehat{\beta})\right)x^{\Delta}_{j,t-1}.
\end{aligned}
$$

Moreover, the matrix of second derivatives of $\widetilde{L}_t(\alpha,\beta)$ is

$$
\frac{\partial^2\widetilde{L}_t(\alpha,\beta)}{\partial(\alpha,\beta')'\partial(\alpha,\beta')}=-N(N-1).\widehat{H}_t(\alpha,\beta) \tag{10}
$$

where

$$
\begin{aligned}
\widehat{H}_t(\alpha,\beta) &= \frac{\exp(\alpha)}{N(N-1)}\sum_{i\neq j}\exp\left(\beta' x^{\Delta}_{i,t-1}-\beta' x^{\Delta}_{j,t-1}\right) \\
&\quad\times\begin{pmatrix}1\\ x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\end{pmatrix}\begin{pmatrix}1\\ x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\end{pmatrix}'.
\end{aligned} \tag{11}
$$

Thus (10) is negative definite for all values of α and β, hence $\widetilde{L}_t(\alpha,\beta)$ is unimodal, and consequently the solutions $\widehat{\alpha}$ and $\widehat{\beta}$ of the first-order conditions (9) are unique.

It follows from standard textbook maximum likelihood theory[5] that under some regularity conditions ML estimators are consistent, asymptotically normally distributed, and asymptotically efficient. However, the present case is not exactly a standard textbook case. In particular, given the consistency of $\widehat{\alpha}$ and $\widehat{\beta}$, the standard approach for deriving asymptotic normality is to set forth conditions such that for $N\to\infty$,

$$
\frac{1}{\sqrt{N(N-1)}}\frac{\partial\widetilde{L}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')}\to N\left[0,\overline{H}_t(\alpha_0,\beta_0)\right] \tag{12}
$$

in distribution[6], and

$$
\frac{1}{N(N-1)}\frac{\partial^2\widetilde{L}_t(\alpha,\beta)}{\partial(\alpha,\beta')'\partial(\alpha,\beta')}=-\widehat{H}_t(\alpha,\beta)\to-\overline{H}_t(\alpha,\beta) \tag{13}
$$

[5] See for example Bierens (1994, Ch. 4).

[6] Note that in our case $N(N-1)$ is the effective sample size.

in probability, uniformly on a compact parameter space $\mathbb{A}\times\mathbb{B}$, say, containing $(\alpha_0, \beta_0')'$, where

$$\overline{H}_t(\alpha,\beta) = \lim_{N\to\infty} E\left[\widehat{H}_t(\alpha,\beta)\right]. \tag{14}$$

In the present case (14) does **not** hold, but under some regularity conditions (13) and (12) hold with

$$\overline{H}_t(\alpha,\beta) = p\lim_{N\to\infty} \widehat{H}_t(\alpha,\beta)$$

instead of (14). Hence:

Conjecture 2 *Under some regularity condition,*

$$\sqrt{N(N-1)}\begin{pmatrix} \widehat{\alpha}-\alpha_0 \\ \widehat{\beta}-\beta_0 \end{pmatrix} \to \mathcal{N}\left[0, \overline{H}_t(\alpha_0,\beta_0)^{-1}\right] \tag{15}$$

in distribution as $N\to\infty$. Moreover,

$$p\lim_{N\to\infty} \widehat{H}_t(\widehat{\alpha},\widehat{\beta})^{-1} = \overline{H}_t(\alpha_0,\beta_0)^{-1}. \tag{16}$$

It is not too hard to derive these regularity conditions on the basis of the unobserved migration flows $y_{i|j}(t)$ and the properties of the Poisson distribution. However, the derivations involved are pretty tedious, and therefore given in Appendix 1.

6 Nonlinear least squares estimation

It follows from the properties on the Poisson distribution, Conjecture 1, and (6) and (7) that

$$\begin{aligned} & E\left[y_{j,t}^{-}\middle|\mathcal{F}_{N,t-1}\right] \\ = & N\exp(\alpha_0)\widehat{a}_t(\beta_0)\exp(-\beta_0' x_{j,t-1}^{\Delta}) - \exp(\alpha_0), \\ & E\left[\left(y_{j,t}^{-} - E\left[y_{j,t}^{-}\middle|\mathcal{F}_{N,t-1}\right]\right)^2\middle|\mathcal{F}_{N,t-1}\right] \\ = & N\exp(\alpha_0)\widehat{a}_t(\beta_0)\exp(-\beta_0' x_{j,t-1}^{\Delta}) - \exp(\alpha_0), \end{aligned} \tag{17}$$

and

$$
\begin{aligned}
& E\left[y_{j,t}^{+} \middle| \mathcal{F}_{N,t-1} \right] & (18) \\
= \;& N \exp(\alpha_0) \widehat{a}_t(-\beta_0) \exp(\beta_0' x_{j,t-1}^{\Delta}) - \exp(\alpha_0), \\
& E\left[\left(y_{j,t}^{+} - E\left[y_{j,t}^{+} \middle| \mathcal{F}_{N,t-1} \right] \right)^2 \middle| \mathcal{F}_{N,t-1} \right] \\
= \;& N \exp(\alpha_0) \widehat{a}_t(-\beta_0) \exp(\beta_0' x_{j,t-1}^{\Delta}) - \exp(\alpha_0),
\end{aligned}
$$

where

$$
\widehat{a}_t(\beta) = \frac{1}{N} \sum_{i=1}^{N} \exp\left(\beta' x_{i,t-1}^{\Delta} \right). \tag{19}
$$

Hence in principle we can estimate the parameters by nonlinear least squares (NLLS), on the basis of either the $y_{j,t}^{-}$'s or the $y_{j,t}^{+}$'s only. Of course, NLLS is less efficient than ML, but the NLLS model can be consistently tested by the Integrated Conditional Moment Test (ICM) of Bierens and Ploberger (1997). This is the main reason for considering NLLS.

The asymptotic properties of the NLLS estimators involved are non-standard. To see this, recall that if Y_N is Poisson$(N.\theta)$ distributed then its characteristic function is

$$
E\left[\exp(\mathbf{i}\tau Y_N) \right] = \exp\left[N.\theta \left(\exp(\mathbf{i}\tau) - 1 \right) \right],
$$

where $\mathbf{i} = \sqrt{-1,}$ so that the characteristic function of $U_N = (Y_N - N.\theta)/\sqrt{N}$ is given by

$$
E\left[\exp(\mathbf{i}\tau U_N) \right] = \exp\left[N\theta \left(\exp(\mathbf{i}.\tau/\sqrt{N}) - \mathbf{i}.\tau/\sqrt{N} - 1 \right) \right].
$$

It is a standard exercise to verify that for $N \to \infty, E\left[\exp(\mathbf{i}\tau U_N) \right] \to \exp\left(-\theta/2 \right),$ and consequently, $U_N \to \mathcal{N}(0, \theta)$ in distibution. This result suggests that the nonlinear regression model for $y_{j,t}^{-}$ is approximately of the form

$$
y_{j,t}^{-}/\sqrt{N} = \sqrt{N} \exp(\gamma_{0,N}^{-} - \beta_0' x_{j,t-1}^{\Delta}) + u_{N,j,t}, \tag{20}
$$

where

$$
u_{N,j,t} | \mathcal{F}_{N,t-1} \sim \mathcal{N}(0, \exp(\gamma_{0,N}^{-} - \beta_0' x_{j,t-1}^{\Delta})).
$$

with

$$
\gamma_{0,N}^{-} = \alpha_0 + \ln \widehat{a}_t(\beta_0).
$$

The problem now is the factor $\sqrt{N}$ in the reponse function, which asymptotically blows it up to infinity. However, it is easy to show (see Appendix 2) that this is actually an advantage, as the nonlinear least squares estimator $\widehat{\theta}$ of $\theta_0 = (\gamma_{0,N}^{-} - \beta_0')'$ is asymptotically normally distributed with convergence rate N rather than the usual rate $\sqrt{N}$:

$$N\left(\widehat{\theta} - \theta_0\right) \to \mathcal{N}(0, B^{-1}AB^{-1})$$

in distribution, where $B^{-1}AB^{-1}$ is the probability limit of N times White's (1980) heteroskedasticity-consistent covariance matrix.

Moreover, the ICM test carries over to this model without the need for modifications. See Appendix 3.

7 The ICM test

The ICM test is based on the following theorem:

Theorem 1 *Let u be a random variable satisfying $E|u| < \infty$, and $P[E(u|x) = 0] < 1$, where $x \in \mathbb{R}^k$ is a bounded random vector.*

(a) *Let $w(u)$ be a complex or real valued function that is infinitely many times differentiable in $u = 0$ and satisfies the condition that the set*

$$\{s \in \mathbb{N} : (d/du)^s w(u) \mid_{u=0} = 0\}$$

is finite. Then for every $\varepsilon > 0$ there exists a $\xi \in \Re^k$ such that $E[u.w(\xi' x)] \neq 0$ and $\|\xi\| < \varepsilon$.

(b) *If in addition $w(u)$ is a power series in an open neighborhood of $u = 0$, i.e., for some $\delta > 0$, $w(u) = \sum_{s=0}^{\infty} (\gamma_s / s!)\, u^s$ for $|u| < \delta$, where $\gamma_s = (d/du)^s w(u) \mid_{u=0}$, then the set $\left\{\xi \in \mathbb{R}^k : E[u.w(\xi' x)] = 0\right\}$ has Lebesgue measure zero and is nowhere dense.*

Proof: See Bierens (1982) for part (a) with $w(u) = \exp(i.u)$, Bierens (1990) for the case $w(u) = \exp(u)$, and Bierens and Ploberger (1997) for the general case. Examples of suitable functions $w(u)$ in the general case are $w(u) = \cos(u) + \sin(u)$, and $w(u) = 1/[1 + \exp(c - u)]$ for $c \neq 0$. See also Stinchcombe and White (1998) for further elaborations on this theorem.

The condition that the random vector x is bounded can be get rid off by replacing x with $\Phi(x)$, where Φ is a Borel measurable bounded one-to-one mapping, because the σ-algebra generated by x is then the same as the

σ-algebra generated by $\Phi(x)$, hence conditioning on $\Phi(x)$ is equivalent to conditioning on x. See Bierens (1982, 1990).

Theorem 1 suggests that, given a random sample (y_t, x_t), $t = 1, .., n$, $x_t \in \mathbb{R}^k$, and a conditional expectation model $E(y_t|x_t) = g(x_t, \theta_0)$, the null hypothesis $P\left[E(y_t|x_t) = g(x_t, \theta_0)\right] = 1$ for some θ_0, can be consistently tested on the basis of the Integrated Conditional Moment (ICM) statistic

$$\int \left|\widehat{z}(\xi)\right|^2 d\mu(\xi),$$

where

$$\widehat{z}(\xi) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{u}_t w\left(\xi' \Phi(x_t)\right), \tag{21}$$

with

$$\widehat{u}_t = y_t - g(x_t, \widehat{\theta}),$$

where $\widehat{\theta}$ is the nonlinear least squares estimator of θ_0, Φ is a bounded one-to-one mapping, $w(.)$ is a weighting function satisfying the conditions of Theorem 1, and μ a probability measure on a compact set $\Xi \subset \mathbb{R}^k$ with positive Lebesgue measure, which is absolute continuous with respect to Lebesgue measure. The ICM test was proposed by Bierens (1982), for the case $w(u) = \exp(i.u)$, Ξ a hypercube in $\mathbb{R}^k$, μ the Lebesgue measure on Ξ, and i.i.d. observations (y_t, x_t).

It has been shown by Bierens (1990) and Bierens and Ploberger (1997) that under some mild regularity conditions (among which the assumption that the function $w(.)$ is real-valued), and the null hypothesis involved, $\widehat{z} \Rightarrow z$ on Ξ, where z is a zero-mean Gaussian process with covariance function $\Gamma(\xi_1, \xi_2) = E\left[z(\xi_1)z(\xi_2)\right]$, hence $\int \left|\widehat{z}(\xi)\right|^2 d\mu(\xi) \rightarrow \int \left|z(\xi)\right|^2 d\mu(\xi)$ in distribution, whereas under the general alternative that the null is false, $\widehat{z}(\xi)/\sqrt{n} \rightarrow \eta(\xi) = p\lim_{n\rightarrow\infty} \frac{1}{n} \sum_{t=1}^{n} \widehat{u}_t w\left(\xi'\Phi(x_t)\right)$ in probability, uniformly on Ξ, where $\eta(\xi) \neq 0$ except on a set with zero Lebesgue measure. Consequently, under the alternative $(1/n) \int \left|\widehat{z}(\xi)\right|^2 d\mu(\xi) \rightarrow \int \left|\eta(\xi)\right|^2 d\mu(\xi) > 0$, a.s.

The asymptotic null distribution of the ICM statistic is of the type

$$\int \left|z(\xi)\right|^2 d\mu(\xi) = \sum_{i=1}^{\infty} \lambda_i \varepsilon_i^2,$$

where the ε_i's are i.i.d. $N(0,1)$ and the λ_i's are the eigenvalues of the covariance function Γ. Moreover,

$$\frac{\int |z(\xi)|^2 \, d\mu(\xi)}{\int \Gamma(\xi,\xi) d\mu(\xi)} = \frac{\sum_{i=1}^{\infty} \lambda_i \varepsilon_i^2}{\sum_{i=1}^{\infty} \lambda_i} \leq \sup_{m \geq 1} \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i^2 = \overline{T},$$

say, so that asymptotic critical values can be derived from the latter distribution. The actual test statistic of the ICM test is therefore

$$\widehat{T}_{ICM} = \frac{\int |\widehat{z}(\xi)|^2 \, d\mu(\xi)}{\int \widehat{\Gamma}(\xi,\xi) d\mu(\xi)}, \tag{22}$$

where $\widehat{\Gamma}(\xi_1, \xi_2)$ is a consistent estimator of $\Gamma(\xi_2, \xi_2)$, uniformly on $\Xi \times \Xi$.

The asymptotic null distribution of $\widehat{T}_{ICM}$ is case-dependent, because the eigenvalues λ_i depend on the distribution of (y_t, x_t) and the conditional expectation model $g(x_t, \theta_0)$, but is dominated by the distribution of $\overline{T}$. Thus, denoting the $1-\alpha$ quantile of $\overline{T}$ by T_α, i.e., $P(\overline{T} \geq T_\alpha) = \alpha$, the null hypothesis is rejected at the $\alpha \times 100\%$ significance level if $\widehat{T}_{ICM} \geq T_\alpha$. The values of T_α for $\alpha = 0.10, 0.05, 0.01$ can be found in Bierens and Ploberger (1997).

8 Empirical results

8.1 Maximum likelihood results

Recall that our model for migration flows $y_{i|j}(t)$ from region j to region $i \neq j$ in year t is

$$y_{i|j}(t) \sim \text{Poisson}\left(\eta_{ijt}(\alpha_0, \beta_0)\right),$$

where

$$\begin{aligned} \ln\left(\eta_{ijt}(\alpha_0, \beta_0)\right) &= \alpha_0 + \beta_{0,1}\left[\left(y_{i,t-1}^{+} - y_{i,t-1}^{-}\right) - \left(y_{j,t-1}^{+} - y_{j,t-1}^{-}\right)\right]/10000 \\ &\quad + \beta_{0,2}\left[\left(\ln X_{i,t-1} - \ln X_{i,t-2}\right) - \left(\ln X_{j,t-1} - \ln X_{j,t-2}\right)\right], \end{aligned}$$

with $y_{j,t}^{+}$ = in-migration, $y_{j,t}^{-}$ = out-migration, and $X_{j,t}$ = employment in region j in year t.

The maximum likelihood estimation results[7], $\widehat{\beta}' = (\widehat{\beta}_1, \widehat{\beta}_2)$, for $\beta_0' = (\beta_{0,1}, \beta_{0,2})$ are presented in Table 1.

[7] The ML estimation, NLLS estimation, and the ICM tests in this section have been conducted by using the modules MIGRATION and NLLS in the software package *EasyReg International* developed by the first author. See Bierens (2001).

Table 1: *Maximum Likelihood Estimation results*

Year	$\widehat{\beta}_1$	*t-value*	$\widehat{\beta}_2$	*t-value*	*log-likelihood*[8]
1984	0.1394199	50.71	1.3788532	18.86	3757199.9
1985	0.5469267	150.02	0.2468324	5.12	3785956.0
1986	0.2151766	88.35	0.6200153	11.26	3282169.6
1987	0.2984167	72.78	1.6516470	17.61	3490124.1
1988	0.4984087	176.67	-0.0949351	-1.01	3654890.0
1989	0.1672895	81.62	2.9367199	26.73	3657387.8
1990	0.2698038	96.89	1.3021499	12.26	4052663.2
1991	0.4535728	149.64	26.6814053	291.35	4206964.5
1992	0.3303883	125.69	4.1647239	74.58	4273508.9
1993	0.2843011	95.99	2.4979466	31.03	3992985.4
1994	0.3412842	103.24	-0.4642191	-7.27	4000533.5
1995	0.3357902	111.09	0.0286840	0.45	3692279.9
1996	0.4892792	166.27	-0.4929617	-6.23	3415942.8
1997	0.4628604	187.22	-0.6146902	-5.74	3367311.0

As we suspected, the parameters are not constant over time. Moreover, the $\widehat{\beta}_1$'s have the expected sign, and are all strongly significant. However, the results for the $\widehat{\beta}_2$'s are puzzling, with respect to their signs as well as their variation over time.

As to the latter, the extremely large value of $\widehat{\beta}_2$ in 1991 likely corresponds to the German unification. It seems that the influx of workers from the former DDR in 1991 has spread over the regions with the highest employment growth rates.

Up to 1993 the $\widehat{\beta}_2$'s are positive, except for the insignificant value in 1988, but in 1994, 1996 and 1997 they become negative. The positive signs of the $\widehat{\beta}_2$'s up to 1993 seem to corroborate the "people follow jobs" explanation of internal migration discussed in the Introduction. A possible explanation for the negative effect after 1993 may be that, ceteris paribus, in regions with low employment growth rates quality housing is more affordable and available than in high employment growth regions, and that due to the economic boom of the nineties the cost of commuting relative to income has fallen. Cf.. Kontuly (1998) and Kontuly and Dearden (1998). Also, these negative signs

[8] Only the part that depends on parameters.

seem to corroborate the demographic restructuring explanation discussed in the Introduction. Admittedly, these explanations for the negative signs of $\widehat{\beta}_2$'s are merely speculations, but at least it is clear that the internal migration patterns in Western Germany have structurally changed over time.

However, before we draw definite conclusions, we have to check whether the model is correctly specified, by re-estimation it for the in- and out-migration seperately by nonlinear least squares, and testing the conditional expectiation specification by the ICM test.

8.2 Nonlinear least squares estimation and ICM test results

If the ML model is correctly specified, the NLLS estimation results should yield estimates of $\beta_0' = (\beta_{0,1}, \beta_{0,2})$ comparable with those in Table 1, for both in- and out-migration. Moreover, the NLLS models involved should be accepted by the ICM test.

The ICM test will be conducted with weight function $w(u) = \cos(u) + \sin(u)$, nuisance parameter space $\Xi = [-c, c] \times [-c, c]$ with $c = 5$, and $\mu(.)$ the probability measure of the uniform distribution on Ξ. Note that the ICM test is a right-sided test, with 10% critical value 3.23 and 5% critical value 4.26.

The NLLS estimation and ICM test results are presented in Tables 2 and 3.

Table 2: *NLLS estimation and test results for in-migration*

Year	$\widehat{\beta}_1$	*t-value*	$\widehat{\beta}_2$	*t-value*	R^2	ICM test
1984	0.70396	11.393	-12.78149	-1.632	0.1618	1.33
1985	0.76883	4.470	6.98364	1.491	0.0901	2.53
1986	0.39717	2.391	-22.96084	-3.827	0.1470	1.37
1987	0.98524	2.004	0.65807	0.059	0.0557	1.61
1988	1.34166	20.216	-2.35473	-0.270	0.5362	1.76
1989	1.22402	18.143	-7.29398	-0.835	0.4996	1.96
1990	1.85579	4.108	8.47995	0.630	0.2747	1.99
1991	1.17825	10.275	-17.26297	-1.496	0.2604	2.02
1992	1.83819	9.227	-28.32716	-3.242	0.5935	2.49
1993	1.46859	15.962	2.56380	0.304	0.3548	1.46
1994	1.08552	15.592	-2.16469	-0.310	0.2530	1.63
1995	1.38785	21.051	-14.38693	-2.226	0.3393	1.32
1996	1.48135	9.846	-23.42422	-2.744	0.5545	1.22
1997	1.65209	12.737	16.24337	1.271	0.6718	1.44

Table 3: *NLLS estimation and test results for out-migration*

Year	$\widehat{\beta}_1$	*t-value*	$\widehat{\beta}_2$	*t-value*	R^2	ICM test
1984	-0.65411	-7.346	17.54040	2.100	0.1276	1.44
1985						
1986						
1987						
1988						
1989						
1990						
1991						
1992						
1993						
1994						
1995						
1996						
1997						

To be completed

9 Appendix 1

In this appendix we will set forth the conditions for consistency and asymptotic normality of the maximum likelihood estimators of the parameters α_0 and β_0, for a given year t. Note that α_0 and β_0 may be different for different years t.

9.1 Consistency

9.1.1 General conditions

The conditional log-likelihood for year t can be written as

$$\widetilde{L}_t(\alpha,\beta) = N^2\widetilde{Q}_t(\alpha,\beta) - \sum_{i\neq j} \ln\left(y_{i|j}(t)!\right)$$

where

$$\begin{aligned} & \widetilde{Q}_t(\alpha,\beta) \qquad (23) \\ = \; & \alpha\frac{1}{N^2}\sum_{i\neq j} y_{i|j}(t) + \beta'\frac{1}{N^2}\sum_{i\neq j} y_{i|j}(t)x^{\Delta}_{i,t-1} - \beta'\frac{1}{N^2}\sum_{i\neq j} y_{i|j}(t)x^{\Delta}_{j,t-1} \\ & -\exp(\alpha)\left(\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta' x^{\Delta}_{i,t-1}\right)\right)\left(\frac{1}{N}\sum_{j=1}^{N}\exp\left(-\beta' x^{\Delta}_{j,t-1}\right)\right) \\ & +\exp(\alpha)/N. \end{aligned}$$

For proving the consistency of the maximum likelihood estimators of α_0 and β_0 only the part $N^2\widetilde{Q}_t(\alpha,\beta)$ of $\widetilde{L}_t(\alpha,\beta)$ is relevant.

Let $\mathbb{A}\times\mathbb{B}$ be the parameter space satisfying

Assumption 1
$\mathbb{A}$ is a given closed and bounded interval in $\mathbb{R}$ containing α_0,
$\mathbb{B}$ is a given closed and bounded hypercube in $\mathbb{R}^2$ containing β_0,
$\beta\in\mathbb{B}$ implies $-\beta\in\mathbb{B}$.

Note that the first two parts of Assumption 1 imply that $\mathbb{A}\times\mathbb{B}$ is compact and convex. The last part of Assumption 1 together with the convexity of $\mathbb{B}$ implies that $0 \in \mathbb{B}$, which is necessary for testing the significance of the components of β_0. The symmetry of $\mathbb{B}$ is not essential, though, but has notational advantages (see Assumption 2 below).

For proving the (weak) consistency of the maximum likelihood estimators $\widehat{\alpha}$ and $\widehat{\beta}$, i.e.,

$$p \lim_{N\to\infty} \widehat{\alpha} = \alpha_0,\ p \lim_{N\to\infty} \widehat{\beta} = \beta_0, \tag{24}$$

we will restrict $(\widehat{\alpha}, \widehat{\beta})$ to $\mathbb{A} \times \mathbb{B}$:

$$(\widehat{\alpha}, \widehat{\beta}) = \arg \max_{(\alpha,\beta)\in\mathbb{A}\times\mathbb{B}} \widetilde{Q}_t(\alpha, \beta).$$

As is well-known[9], for proving (24) it suffices to establish the following conditions:

$$p \lim_{N\to\infty} \sup_{(\alpha,\beta)\in\mathbb{A}\times\mathbb{B}} \left| \widetilde{Q}_t(\alpha, \beta) - \overline{Q}_t(\alpha, \beta) \right| = 0, \tag{25}$$

where

$$\overline{Q}_t(\alpha, \beta) \text{ is continuous on } \mathbb{A} \times \mathbb{B}; \tag{26}$$

$$(\alpha_0, \beta_0) = \arg \max_{(\alpha,\beta)\in\mathbb{A}\times\mathbb{B}} \overline{Q}_t(\alpha, \beta); \tag{27}$$

and

$$(\alpha_0, \beta_0) \text{ is unique.} \tag{28}$$

9.1.2 Further conditions for the uniform law of large numbers

In this subsection we will set forth further conditions for the validity of (25) and (26).

Recall that conditional on $\mathcal{F}_{N,t-1}$, and for $i \neq j$, the $y_{i|j}(t)$'s are independent Poisson distributed with expected value and variance equal to

$$\begin{aligned} E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right] &= E\left[\left(y_{i|j}(t) - E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right]\right)^2 \middle| \mathcal{F}_{N,t-1}\right] \\ &= \exp(\alpha_0)\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right)\exp\left(-\beta_0' x_{j,t-1}^{\Delta}\right) \end{aligned} \tag{29}$$

Consequently,

[9] See for example Bierens (1994, Chap. 4).

$$
\begin{aligned}
& E\left[\left.\left(\frac{1}{N^2}\sum_{i\neq j}\left(y_{i|j}(t)-E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right]\right)\right)^2\right|\mathcal{F}_{N,t-1}\right] \qquad (30)\\
= \;& \exp(\alpha_0)\frac{1}{N^4}\sum_{i\neq j}\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right)\exp\left(-\beta_0' x_{j,t-1}^{\Delta}\right)\\
= \;& \frac{\exp(\alpha_0)}{N^2}\left(\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right)\right)\left(\frac{1}{N}\sum_{j=1}^{N}\exp\left(-\beta_0' x_{j,t-1}^{\Delta}\right)\right)\\
& -\frac{\exp(\alpha_0)}{N^3}
\end{aligned}
$$

and similarly, for arbitrary $\xi \in \mathbb{R}^3$,

$$
\begin{aligned}
& E\left[\left.\left(\frac{1}{N^2}\sum_{i\neq j}\left(y_{i|j}(t)-E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right]\right)\xi' x_{i,t-1}^{\Delta}\right)^2\right|\mathcal{F}_{N,t-1}\right] \qquad (31)\\
= \;& \frac{\exp(\alpha_0)}{N^2}\xi'\left(\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right)x_{i,t-1}^{\Delta}x_{i,t-1}^{\Delta}{}'\right)\xi\left(\frac{1}{N}\sum_{j=1}^{N}\exp\left(-\beta_0' x_{j,t-1}^{\Delta}\right)\right)\\
& -\frac{\exp(\alpha_0)}{N^3}\xi'\left(\frac{1}{N}\sum_{i=1}^{N}x_{i,t-1}^{\Delta}x_{i,t-1}^{\Delta}{}'\right)\xi.
\end{aligned}
$$

Now assume

Assumption 2

$$
\widehat{a}_t(\beta)=\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta' x_{i,t-1}^{\Delta}\right)\to a_t(\beta)>0,
$$

$$
\widehat{b}_t(\beta)=\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta' x_{i,t-1}^{\Delta}\right)x_{i,t-1}^{\Delta}\to b_t(\beta),
$$

$$
\widehat{C}_t(\beta)=\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta' x_{i,t-1}^{\Delta}\right)x_{i,t-1}^{\Delta}x_{i,t-1}^{\Delta}{}'\to C_t(\beta),
$$

in probability, uniformly on $\mathbb{B}$, where the limits involved are finite. Moreover, $a_t(\beta)$, $b_t(\beta)$ and $C_t(\beta)$ are continuous on $\mathbb{B}$, and furthermore,

$$\frac{\partial a_t(\beta)}{\partial \beta'} = b_t(\beta),\ \frac{\partial b_t(\beta)}{\partial \beta} = C_t(\beta)\ for\ \beta \in \mathbb{B}. \tag{32}$$

Assumption 2 is more restrictive than necessary for proving consistency only, but its full extent will be needed later.

It follows from Assumption 2 that

$$\begin{aligned}
p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right) &= a_t(\beta_0),\\
p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\exp\left(-\beta_0' x_{i,t-1}^{\Delta}\right) &= a_t(-\beta_0),\\
p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right) x_{i,t-1}^{\Delta} x_{i,t-1}^{\Delta}{}' &= C_t(\beta_0),\\
p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\exp\left(-\beta_0' x_{i,t-1}^{\Delta}\right) x_{i,t-1}^{\Delta} x_{i,t-1}^{\Delta}{}' &= C_t(-\beta_0),\\
p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} x_{i,t-1}^{\Delta} x_{i,t-1}^{\Delta}{}' &= C_t(0),
\end{aligned}$$

hence (30), and (31) are $O_p(N^{-2})$. Using the conditional Chebishev inequality and the bounded convergence theorem it follows now that

$$p\lim_{N\to\infty}\frac{1}{N^2}\sum_{i\neq j}\left(y_{i|j}(t) - E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right]\right) = 0,$$

$$p\lim_{N\to\infty}\frac{1}{N^2}\sum_{i\neq j}\left(y_{i|j}(t) - E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right]\right) x_{i,t-1}^{\Delta} = 0,$$

$$p\lim_{N\to\infty}\frac{1}{N^2}\sum_{i\neq j}\left(y_{i|j}(t) - E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right]\right) x_{j,t-1}^{\Delta} = 0.$$

Similarly, it follows from Assumptions 1 and 2 that

$$\frac{1}{N^2}\sum_{i\neq j} E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right] \to \exp(\alpha_0)a_t(\beta_0)a_t(-\beta_0),$$

$$\frac{1}{N^2}\sum_{i\neq j} E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right] x^{\Delta}_{i,t-1} \to \exp(\alpha_0)b_t(\beta_0)a_t(-\beta_0),$$

$$\frac{1}{N^2}\sum_{i\neq j} E\left[y_{i|j}(t)|\mathcal{F}_{N,t-1}\right] x^{\Delta}_{j,t-1} \to \exp(\alpha_0)b_t(-\beta_0)a_t(\beta_0),$$

in probability. Therefore, (25) and (26) hold under Assumptions 1 and 2, with

$$\begin{aligned}\overline{Q}_t(\alpha,\beta) &= \alpha\exp(\alpha_0)a_t(\beta_0)a_t(-\beta_0) \\ &\quad +\beta'\left(\exp(\alpha_0)b_t(\beta_0)a_t(-\beta_0) - \exp(\alpha_0)b_t(-\beta_0)a_t(\beta_0)\right) \\ &\quad -\exp(\alpha)a_t(\beta)a_t(-\beta).\end{aligned} \tag{33}$$

9.1.3 Conditions for uniqueness of the parameters

It follows from part (32) of Assumption 2 and (33) that

$$\begin{aligned}\frac{\partial\overline{Q}_t(\alpha,\beta)}{\partial\alpha} &= \exp(\alpha_0)a_t(\beta_0)a_t(-\beta_0) - \exp(\alpha)a_t(\beta)a_t(-\beta), \qquad (34)\\ \frac{\partial\overline{Q}_t(\alpha,\beta)}{\partial\beta'} &= \exp(\alpha_0)a_t(-\beta_0)b_t(\beta_0) - \exp(\alpha_0)a_t(\beta_0)b_t(-\beta_0) \\ &\quad -\exp(\alpha)a_t(-\beta)b_t(\beta) + \exp(\alpha)a_t(\beta)b_t(-\beta).\end{aligned}$$

Thus the first-order conditions

$$\frac{\partial\overline{Q}_t(\alpha_0,\beta_0)}{\partial\alpha_0} = 0,\ \frac{\partial\overline{Q}_t(\alpha_0,\beta_0)}{\partial\beta_0'} = 0 \tag{35}$$

hold.

In order to check the second order conditions, denote

$$\overline{H}_t(\alpha,\beta) = -\frac{\partial^2\overline{Q}_t(\alpha,\beta)}{\partial(\alpha,\beta')'\partial(\alpha,\beta')}. \tag{36}$$

If (36) is positive definite for all $\alpha \in \mathbb{A}$ and $\beta \in \mathbb{B}$ then $\overline{Q}_t(\alpha,\beta)$ is unimodal, hence the solution of the first order condition (35) is then unique.

It follows straightforwardly from (34) that

$$\begin{aligned}
&\overline{H}_t(\alpha,\beta) = \exp(\alpha)\times \\
&\begin{pmatrix} a_t(\beta)a_t(-\beta) & a_t(-\beta)b_t(\beta)' - a_t(\beta)b_t(-\beta)' \\ a_t(-\beta)b_t(\beta) - a_t(\beta)b_t(-\beta) & a_t(-\beta)C_t(\beta) - b_t(\beta)b_t(-\beta)' - b_t(-\beta)b_t(\beta)' + a_t(\beta)C_t(-\beta) \end{pmatrix} \\
&= \exp(\alpha)a(\beta)a(-\beta)\times \\
&\begin{pmatrix} 1 & b_*(\beta)' - b_*(-\beta)' \\ b_*(\beta) - b_*(-\beta) & C_*(\beta) - b_*(\beta)b_*(-\beta)' + C_*(-\beta) - b_*(-\beta)b_*(\beta)' \end{pmatrix},
\end{aligned} \tag{37}$$

where

$$b_*(\beta) = a(\beta)^{-1}b(\beta),\ C_*(\beta) = a(\beta)^{-1}C(\beta).$$

Thus, $\overline{H}_t(\alpha,\beta)$ is positive definite for all $\alpha \in \mathbb{A}$ and $\beta \in \mathbb{B}$ if and only if the matrix (37) is positive definite for all $\beta \in \mathbb{B}$. To establish the latter, denote the matrix (37) by $D(\beta)$, and let $\varsigma \in \mathbb{R}$ and $\xi \in \mathbb{R}^2$ be arbitrary. Then

$$\begin{aligned}
& (\varsigma,\xi')\,D(\beta)\,(\varsigma,\xi')' \\
= \ & (\varsigma + b_*(\beta)'\xi - b_*(-\beta)'\xi)^2 \\
& + \xi'\left(C_*(\beta) - b_*(\beta)b_*(\beta)\right)\xi + \xi'\left(C_*(-\beta) - b_*(-\beta)b_*(-\beta)\right)\xi \\
\geq \ & \xi'\left(C_*(\beta) - b_*(\beta)b_*(\beta)\right)\xi + \xi'\left(C_*(-\beta) - b_*(-\beta)b_*(-\beta)\right)\xi \\
= \ & \frac{\xi'\left(a(\beta)C(\beta) - b(\beta)b(\beta)'\right)\xi}{a(\beta)^2} + \frac{\xi'\left(a(-\beta)C(-\beta) - b(-\beta)b(-\beta)'\right)\xi}{a(-\beta)^2} \\
> \ & 0 \text{ for all } \xi \neq 0 \text{ and } \beta \in \mathbb{B}
\end{aligned}$$

if

Assumption 3 *$a(\beta)C(\beta) - b(\beta)b(\beta)'$ is positive definite for all $\beta \in \mathbb{B}$.*

This assumption is not too farfetched, because by Assumption 2,

$$\begin{aligned}
& a(\beta)C(\beta) - b(\beta)b(\beta)' \\
= \ & p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta' x_{i,t-1}^{\Delta}\right)\left(a_t(\beta)^{1/2}x_{i,t-1}^{\Delta} - a_t(\beta)^{-1/2}b_t(\beta)\right) \\
& \times\left(a_t(\beta)^{1/2}x_{i,t-1}^{\Delta} - a_t(\beta)^{-1/2}b_t(\beta)\right)'.
\end{aligned}$$

Assumptions 1-3 now imply the weak consistency of the maximum likelihood estimators $\widehat{\alpha}$ and $\widehat{\beta}$.

9.2 Asymptotic normality

9.2.1 General conditions

As said before, the standard approach in deriving the asymptotic normality result (15) is to set forth conditions such that, for given t and $N \to \infty$, (12) holds, and (13) holds uniformly on $\mathbb{A} \times \mathbb{B}$. However, for notational convenience we will set forth additional conditions such that,

$$\frac{1}{N}\frac{\partial \widetilde{L}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')} = N\frac{\partial \widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')} \to \mathcal{N}\left[0, \overline{H}_t(\alpha_0,\beta_0)\right] \tag{38}$$

in distribution, and

$$\frac{1}{N^2}\frac{\partial^2 \widetilde{L}_t(\alpha,\beta)}{\partial(\alpha,\beta')'\partial(\alpha,\beta')} = \frac{\partial^2 \widetilde{Q}_t(\alpha,\beta)}{\partial(\alpha,\beta')'\partial(\alpha,\beta')} \to -\overline{H}_t(\alpha,\beta) \tag{39}$$

in probability, uniformly on $\mathbb{A} \times \mathbb{B}$, where $\overline{H}_t(\alpha_0,\beta_0)$ is non-singular. Note that (38) is asymptotically equivalent to (12), and (39) is asymptotically equivalent to (13).

Condition (39) follows straightforwardly from Assumption 2, with $\overline{H}_t(\alpha,\beta)$ defined by (36), because

$$\begin{aligned}\frac{\partial^2 \widetilde{Q}_t(\alpha,\beta)}{\partial(\alpha,\beta')'\partial(\alpha,\beta')} &= -\frac{\partial^2\left(\exp(\alpha)\widehat{a}_t(\beta)\widehat{a}_t(-\beta) - \exp(\alpha)/N\right)}{\partial(\alpha,\beta')'\partial(\alpha,\beta')} \\ &= -\exp(\alpha)\times\end{aligned}$$

$$\begin{pmatrix} \widehat{a}_t(\beta)\widehat{a}_t(-\beta) & \widehat{a}_t(-\beta)\widehat{b}_t(\beta)' - \widehat{a}_t(\beta)\widehat{b}_t(-\beta)' \\ \widehat{a}_t(-\beta)\widehat{b}_t(\beta) - \widehat{a}_t(\beta)\widehat{b}_t(-\beta) & a_t(-\beta)C_t(\beta) - b_t(\beta)b_t(-\beta)' - b_t(-\beta)b_t(\beta)' + a_t(\beta)C_t(-\beta) \end{pmatrix}$$

$$+\frac{1}{N}\times\begin{pmatrix} \exp(\alpha) & 0' \\ 0 & O \end{pmatrix}.$$

Moreover, the non-singularity of $\overline{H}_t(\alpha_0,\beta_0)$ follows from Assumption 3.

9.2.2 Further conditions for asymptotic normality

It is not hard to verify that

$$\begin{aligned}
& N\frac{\partial \widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')} \\
= & \frac{1}{N}\sum_{i\neq j}\left(y_{i|j}(t)-\exp\left[\alpha_0+\beta_0'\left(x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\right)\right]\right) \\
& \times\begin{pmatrix} 1 \\ x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\end{pmatrix}.
\end{aligned}$$

Thus, for proving (38) it suffices to show that for arbitrary $\varsigma\in\mathbb{R}$ and $\xi\in\mathbb{R}^2$

$$\begin{aligned}
& N(\varsigma,\xi')\frac{\partial \widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')} \\
= & \frac{1}{N}\sum_{i\neq j}\left(y_{i|j}(t)-\exp\left[\alpha_0+\beta_0'\left(x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\right)\right]\right) \\
& \times\left[\varsigma+\xi'\left(x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\right)\right]
\end{aligned}$$

converges in distribution to the normal distribution with zero mean and variance

$$p\lim_{N\to\infty}E\left[\left.\left(N(\varsigma,\xi')\frac{\partial \widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')}\right)^2\right|\mathcal{F}_{t-1}\right]=(\varsigma,\xi')\overline{H}_t(\alpha_0,\beta_0)\begin{pmatrix}\varsigma\\ \xi\end{pmatrix},$$

say.

As to the latter, it follows from the conditional independence of the $y_{i|j}(t)$'s, and the properties of the Poisson distribution [cf. (29)] that

$$\begin{aligned}
& E\left[\left.\left(N(\varsigma,\xi')\frac{\partial \widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')}\right)^2\right|\mathcal{F}_{N,t-1}\right] \\
= & \exp(\alpha_0)\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\exp\left(\beta_0'x^{\Delta}_{i,t-1}\right)\exp\left(-\beta_0'x^{\Delta}_{j,t-1}\right) \\
& \times(\varsigma,\xi')\begin{pmatrix} 1 \\ x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\end{pmatrix}\begin{pmatrix} 1 \\ x^{\Delta}_{i,t-1}-x^{\Delta}_{j,t-1}\end{pmatrix}'\begin{pmatrix}\varsigma\\ \xi\end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{N}\varsigma^2\exp(\alpha_0) \\
= \;& \exp(\alpha_0)\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta' x^{\Delta}_{i,t-1}\right)\frac{1}{N}\sum_{j=1}^{N}\exp\left(-\beta' x^{\Delta}_{j,t-1}\right)(\varsigma,\xi')\times \\
& \begin{pmatrix} 1 & x^{\Delta\prime}_{i,t-1} - x^{\Delta\prime}_{j,t-1} \\ x^{\Delta}_{i,t-1} - x^{\Delta}_{j,t-1} & x^{\Delta}_{i,t-1}x^{\Delta\prime}_{i,t-1} - x^{\Delta}_{j,t-1}x^{\Delta\prime}_{i,t-1} - x^{\Delta}_{i,t-1}x^{\Delta\prime}_{j,t-1} + x^{\Delta}_{j,t-1}x^{\Delta\prime}_{j,t-1} \end{pmatrix}\begin{pmatrix}\varsigma \\ \xi\end{pmatrix} \\
& -\frac{1}{N}\varsigma^2\exp(\alpha_0) \\
= \;& (\varsigma,\xi')\widehat{H}^*_t(\alpha_0,\beta_0)\begin{pmatrix}\varsigma \\ \xi\end{pmatrix} - \frac{1}{N}\varsigma^2\exp(\alpha_0) \to (\varsigma,\xi')\overline{H}_t(\alpha_0,\beta_0)\begin{pmatrix}\varsigma \\ \xi\end{pmatrix}
\end{aligned}
$$

in probability, where

$$
\widehat{H}^*_t(\alpha,\beta) = \exp(\alpha)\times
$$
$$
\begin{pmatrix} \widehat{a}_t(\beta)\widehat{a}_t(-\beta) & \widehat{a}_t(-\beta)\widehat{b}_t(\beta)' - \widehat{a}_t(\beta)\widehat{b}_t(-\beta)' \\ \widehat{a}_t(-\beta)\widehat{b}_t(\beta) - \widehat{a}_t(\beta)\widehat{b}_t(-\beta) & a_t(-\beta)C_t(\beta) - b_t(\beta)b_t(-\beta)' - b_t(-\beta)b_t(\beta)' + a_t(\beta)C_t(-\beta) \end{pmatrix}.
$$

As is well-known, if Y is Poisson(θ) distributed then its characteristic function is

$$
E\left[\exp(\mathbf{i}\tau Y)\right] = \exp\left[\theta\left(\exp(\mathbf{i}\tau) - 1\right)\right],
$$

where $\mathbf{i} = \sqrt{-1},$ so that the characteristic function of $U = (Y-\theta)\,c$, with c a constant, is given by

$$
E\left[\exp(\mathbf{i}\tau U)\right] = \exp\left[\theta\left(\exp(c.\mathbf{i}.\tau) - c.\mathbf{i}.\tau - 1\right)\right]
$$

Substituting $\theta = \exp\left[\alpha_0 + \beta_0'\left(x^{\Delta}_{i,t-1} - x^{\Delta}_{j,t-1}\right)\right]$, $c = \varsigma + \xi'\left(x^{\Delta}_{i,t-1} - x^{\Delta}_{j,t-1}\right)$, it follows that

$$
\begin{aligned}
& \log\left(E\left[\exp\left(\left(\mathbf{i}.\tau N\right)(\varsigma,\xi')\frac{\partial\widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')}\right)\middle|\mathcal{F}_{N,t-1}\right]\right) \\
= \;& \sum_{i\neq j}\exp\left[\alpha_0 + \beta_0'\left(x^{\Delta}_{i,t-1} - x^{\Delta}_{j,t-1}\right)\right] \\
& \times\left(\exp\left[\left(\varsigma + \xi' x^{\Delta}_{i,t-1} - \xi' x^{\Delta}_{j,t-1}\right).\mathbf{i}\tau/N\right]\right. \\
& \left. - \left(\varsigma + \xi' x^{\Delta}_{i,t-1} - \xi' x^{\Delta}_{j,t-1}\right).\mathbf{i}\tau/N - 1\right).
\end{aligned}
$$

Moreover, it follows from Taylor's theorem that there exists a random variable $\lambda_{N,i,j,t-1}$ such that

$$
\exp\left[\left(\varsigma + \xi' x^{\Delta}_{i,t-1} - \xi' x^{\Delta}_{j,t-1}\right).\mathbf{i}\tau/N\right] - 1
$$

$$
\begin{aligned}
& -\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right).\mathbf{i}\tau/N \\
= \; & \frac{1}{2}\left(\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right).\mathbf{i}\tau/N\right)^{2} \\
& +\frac{1}{3!}\left(\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right).\mathbf{i}\tau/N\right)^{3} \\
& \times\exp\left[\lambda_{N,i,j,t-1}\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right).\mathbf{i}\tau/N\right] \\
= \; & -\frac{1}{2}\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right)^{2}\tau^{2}/N^{2} \\
& -\frac{\mathbf{i}}{6N^{3}}\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right)^{3}\tau^{3} \\
& \times\exp\left[\lambda_{N,i,j,t-1}\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right).\mathbf{i}\tau/N\right],
\end{aligned}
$$

hence

$$
\begin{aligned}
& \log\left(E\left[\exp\left(\left(\mathbf{i}.\tau N\right)\left(\varsigma,\xi'\right)\frac{\partial\widetilde{Q}_{t}(\alpha_{0},\beta_{0})}{\partial(\alpha_{0},\beta_{0}')}\right)\middle|\mathcal{F}_{N,t-1}\right]\right) \\
= \; & -\frac{1}{2}\times\frac{1}{N^{2}}\sum_{i\neq j}\exp\left[\alpha_{0}+\beta_{0}'\left(x_{i,t-1}^{\Delta}-x_{j,t-1}^{\Delta}\right)\right] \\
& \times\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right)^{2}\tau^{2}+R_{N}(\alpha,\beta,\tau,\varsigma,\xi) \\
= \; & -\frac{1}{2}\tau^{2}(\varsigma,\xi')\widehat{H}_{t}(\alpha_{0},\beta_{0})\begin{pmatrix}\varsigma\\ \xi\end{pmatrix}+R_{N}(\tau,\varsigma,\xi),
\end{aligned}
$$

where

$$
\begin{aligned}
& \left|R_{N}(\tau,\varsigma,\xi)\right| \\
= \; & \frac{|\tau|^{3}\exp(\alpha_{0})}{6N}\left|\frac{1}{N^{2}}\sum_{j=1}^{N}\sum_{i=1}^{N}I(i\neq j)\exp\left[\beta_{0}'\left(x_{i,t-1}^{\Delta}-x_{j,t-1}^{\Delta}\right)\right]\right. \\
& \left.\times\left(\varsigma+\xi' x_{i,t-1}^{\Delta}-\xi' x_{j,t-1}^{\Delta}\right)^{3}\right| \\
\leq \; & \frac{|\tau|^{3}\exp(\alpha_{0})}{6N}\times\frac{1}{N^{2}}\sum_{j=1}^{N}\sum_{i=1}^{N}\exp\left(\beta_{0}' x_{i,t-1}^{\Delta}\right)\exp\left(-\beta_{0}' x_{j,t-1}^{\Delta}\right) \\
& \times 27\left(\frac{|\varsigma|+\|\xi\|.\left\|x_{i,t-1}^{\Delta}\right\|+\|\xi\|.\left\|x_{j,t-1}^{\Delta}\right\|}{3}\right)^{3} \\
& +\frac{|\tau.\varsigma|^{3}\exp(\alpha_{0})}{6N^{2}}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{|\tau|^3\exp(\alpha_0)}{6N}\times\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right)\exp\left(-\beta_0' x_{j,t-1}^{\Delta}\right)\\
&\quad\times\frac{27}{3}\left(|\varsigma|^3+\|\xi\|^3\cdot\left\|x_{i,t-1}^{\Delta}\right\|^3+\|\xi\|^3\cdot\left\|x_{j,t-1}^{\Delta}\right\|^3\right)\\
&\quad+\frac{|\tau.\varsigma|^3\exp(\alpha_0)}{6N^2}\\
&= \frac{3|\tau|^3\exp(\alpha)|\varsigma|^3}{2\sqrt{N}}\widehat{a}_t(\beta_0)\widehat{a}_t(-\beta_0)\\
&\quad+\frac{3|\tau|^3\exp(\alpha)\,\|\xi\|^3}{2N}\widehat{a}_t(-\beta_0)\left(\frac{1}{N}\sum_{i=1}^{N}\exp\left(\beta_0' x_{i,t-1}^{\Delta}\right)\left\|x_{i,t-1}^{\Delta}\right\|^3\right)\\
&\quad+\frac{3|\tau|^3\exp(\alpha)\,\|\xi\|^3}{2N}\widehat{a}_t(\beta_0)\left(\frac{1}{N}\sum_{i=1}^{N}\exp\left(-\beta_0' x_{i,t-1}^{\Delta}\right)\left\|x_{j,t-1}^{\Delta}\right\|^3\right)\\
&\quad+\frac{|\tau.\varsigma|^3\exp(\alpha_0)}{6N^2}.
\end{aligned}
$$

Thus, if in addition to Assumptions 1-3,

Assumption 4 $\frac{1}{N}\sum_{i=1}^{N}\left\|x_{i,t-1}^{\Delta}\right\|^3\exp\left(\left|\beta_0' x_{i,t-1}^{\Delta}\right|\right)=O_p(1)$,

then $R_N(\tau,\varsigma,\xi)=O_p\left(1/N\right)$, hence

$$
\begin{aligned}
&p\lim_{N\to\infty}E\left[\exp\left(\mathbf{i}.\tau N(\varsigma,\xi')\frac{\partial\widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')}\right)\middle|\mathcal{F}_{N,t-1}\right] \\
&= \exp\left[-\frac{1}{2}\tau^2(\varsigma,\xi')\overline{H}_t(\alpha_0,\beta_0)\begin{pmatrix}\varsigma\\ \xi\end{pmatrix}\right].
\end{aligned}
\tag{40}
$$

Substituting $\tau=1$ in (40), and applying the bounded convergence theorem, it follows that

$$
\begin{aligned}
&\lim_{N\to\infty}E\left[\exp\left(\mathbf{i}.N(\varsigma,\xi')\frac{\partial\widetilde{Q}_t(\alpha_0,\beta_0)}{\partial(\alpha_0,\beta_0')}\right)\right]\\
&= \exp\left[-\frac{1}{2}(\varsigma,\xi')\overline{H}_t(\alpha_0,\beta_0)\begin{pmatrix}\varsigma\\ \xi\end{pmatrix}\right].
\end{aligned}
$$

This result implies (38). Thus, (15) holds under Assumptions 1-4. Finally, note that (16) follows from Assumptions 1-4 as well.

9.3 The add-up restriction reconsidered

The asymptotic properties of the maximum likelihood estimators have been derived under the implicit assumption that the add-up restriction (1) holds. However, in reality it does not hold. So the question arises whether this matters for the asymptotic results.

In order to address this question, let us introduce region 0, the world outside the $N = 75$ regions under review, and let us assume that to Conjecture 1 holds for regions $i, j = 0, 2, ..., N$. Then the actual log-likelihood becomes

$$\begin{aligned}
\widehat{L}_t(\alpha, \beta) &= \alpha \sum_{j=0}^{N} y_{j,t}^{-} + \beta' \sum_{j=0}^{N} \left(y_{j,t}^{+} - y_{j,t}^{-}\right) x_{j,t-1}^{\Delta} \\
&\quad - \exp(\alpha) \sum_{j=0}^{N} \exp\left(-\beta' x_{j,t-1}^{\Delta}\right) \sum_{i=0}^{N} \exp\left(\beta' x_{i,t-1}^{\Delta}\right) + (N+1)\exp(\alpha) \\
&\quad + (N+1)\exp(\alpha) + r_N \\
&= \alpha \sum_{j=1}^{N} \left(y_{j,t}^{-} - y_{0|j}(t)\right) + \beta' \sum_{j=1}^{N} \left(\left(y_{j,t}^{+} - y_{j|0}(t)\right) - \left(y_{j,t}^{-} - y_{0|j}(t)\right)\right) x_{j,t-1}^{\Delta} \\
&\quad - \exp(\alpha) \sum_{j=1}^{N} \exp\left(-\beta' x_{j,t-1}^{\Delta}\right) \sum_{i=1}^{N} \exp\left(\beta' x_{i,t-1}^{\Delta}\right) + N\exp(\alpha) \\
&\quad - \exp(\alpha) - \left(y_{0,t}^{+} - y_{0,t}^{-}\right)\left(\beta' x_{0,t-1}^{\Delta} - \alpha\right) - \beta' \sum_{j=1}^{N} \left(y_{j|0}(t) - y_{0|j}(t)\right) x_{j,t-1}^{\Delta} \\
&\quad - \exp(\alpha)\exp\left(-\beta' x_{0,t-1}^{\Delta}\right) \sum_{i=1}^{N} \exp\left(\beta' x_{i,t-1}^{\Delta}\right) \\
&\quad - \exp(\alpha)\exp\left(\beta' x_{0,t-1}^{\Delta}\right) \sum_{j=1}^{N} \exp\left(-\beta' x_{j,t-1}^{\Delta}\right) + r_N \\
&= N^2 \widetilde{Q}_t(\alpha, \beta) - N\widehat{q}_t(\alpha, \beta) + r_N
\end{aligned}$$

where

$$r_N = -\sum_{i=0}^{N} \sum_{j=0}^{N} I(i \neq j) \ln\left(y_{i|j}(t)!\right),$$

$\widetilde{Q}_t(\alpha,\beta)$ is defined by (23), and

$$\begin{aligned}\widehat{q}_t(\alpha,\beta) &= \frac{\exp(\alpha)}{N}+\left(\frac{y^+_{0,t}-y^-_{0,t}}{N}\right)\left(\beta' x^\Delta_{0,t-1}-\alpha\right)\\ &+\beta'\left[\frac{1}{N}\sum_{j=1}^N\left(y_{j|0}(t)-y_{0|j}(t)\right)x^\Delta_{j,t-1}\right]\\ &+\exp(\alpha-\beta' x^\Delta_{0,t-1})\frac{1}{N}\sum_{i=1}^N\exp\left(\beta' x^\Delta_{i,t-1}\right)\\ &+\exp(\alpha+\beta' x^\Delta_{0,t-1})\frac{1}{N}\sum_{j=1}^N\exp\left(-\beta' x^\Delta_{j,t-1}\right)+r_N\end{aligned}$$

It follows now easily that, if

Assumption 5 $(y^+_{0,t}-y^-_{0,t})/N=O_p(1)$ *and* $\frac{1}{N}\sum_{j=1}^N\left(y_{j|0}(t)-y_{0|j}(t)\right)x^\Delta_{j,t-1}=O_p(1)$

then asymptotically the term $N\widehat{q}_t(\alpha,\beta)$ in the log-likelihood $\widehat{L}_t(\alpha,\beta)$, and its first and second order partial derivatives, becomes negligeable, hence the asymptotic results do not hinge on the add-up restriction (1).

10 Appendix 2

In this appendix we consider a nonlinear regression model of the form

$$y_j=\sqrt{N}\exp(\theta_0' x_j)+u_j,\ u_j|X\sim N\left[0,\exp(\theta_0' x_j)\right],\ j=1,...,N, \tag{41}$$

where $X=(x_1,....,x_N)'$, which correponds to (20), with $y_j=y^-_{j,t}/\sqrt{N}$ and $x_j'=(1,{}^{\Delta}_{j,t-1}{}')$.

Denote

$$\begin{aligned}M_N(\theta) &= \frac{1}{N^2}\sum_{j=1}^N\left(y_j-\sqrt{N}\exp(\theta' x_j)\right)^2\\ &= \frac{1}{N^2}\sum_{j=1}^N\left(u_j+\sqrt{N}\exp(\theta_0' x_j)-\sqrt{N}\exp(\theta' x_j)\right)^2\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{j=1}^{N} u_j^2 + 2\frac{1}{N\sqrt{N}}\sum_{j=1}^{N} u_j\left(\exp(\theta_0' x_j) - \exp(\theta' x_j)\right) \\
&\quad + \frac{1}{N}\sum_{j=1}^{N}\left(\exp(\theta_0' x_j) - \exp(\theta' x_j)\right)^2
\end{aligned}
$$

Then

$$
\begin{aligned}
\frac{\partial M_N(\theta)}{\partial \theta'} &= -2\frac{1}{N\sqrt{N}}\sum_{j=1}^{N} u_j \exp(\theta' x_j) x_j \\
&\quad -2\frac{1}{N}\sum_{j=1}^{N} \exp(\theta_0' x_j)\exp(\theta' x_j) x_j \\
&\quad +2\frac{1}{N}\sum_{j=1}^{N} \exp(2\theta' x_j) x_j
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 M_N(\theta)}{\partial \theta \partial \theta'} &= -2\frac{1}{N\sqrt{N}}\sum_{j=1}^{N} u_j \exp(\theta' x_j) x_j x_j' \\
&\quad -2\frac{1}{N}\sum_{j=1}^{N} \exp(\theta_0' x_j)\exp(\theta' x_j) x_j x_j' \\
&\quad +4\frac{1}{N}\sum_{j=1}^{N} \exp(2\theta' x_j) x_j x_j'
\end{aligned}
$$

Under some regularity conditions we have

$$
N\frac{\partial M_N(\theta_0)}{\partial \theta_0'} = -2\frac{1}{\sqrt{N}}\sum_{j=1}^{N} u_j \exp(\theta_0' x_j) x_j \to N\left[0, 4A\right]
$$

where

$$
\begin{aligned}
A &= p\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N} u_j^2 \exp(2\theta_0' x_j) x_j x_j' \\
&= p\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N} \exp(3\theta_0' x_j) x_j x_j'
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 M_N(\theta_0)}{\partial\theta_0\partial\theta_0'} &= -2\frac{1}{N\sqrt{N}}\sum_{j=1}^{N} u_j \exp(\theta_0' x_j) x_j x_j' \\
&\quad +2\frac{1}{N}\sum_{j=1}^{N} \exp(2\theta_0' x_j) x_j x_j' \\
&\to 2B
\end{aligned}
$$

in prob, where

$$B = p \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \exp(2\theta_0' x_j) x_j x_j'.$$

Thus

$$N\left(\widehat{\theta} - \theta_0\right) \to \mathcal{N}(0, B^{-1}AB^{-1}).$$

in distribution, where $\widehat{\theta}$ is the NLLS estimator. Therefore, the nonlinear regression model involved can be estimated by standard nonlinear least squares with heteroskedasticity consistent variance matrix, provided that the latter is multiplied by N. However, the t-values and p-values remain valid.

11 Appendix 3

Recall that the main reason for considering NLLS is that we then can test the validity of the model by the ICM test. The question now arises whether for this case the ICM test carries over. In our case (21) becomes

$$
\begin{aligned}
\widehat{z}(\xi) &= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(y_j - \sqrt{N}\exp(\widehat{\theta}' x_j)\right) w\left(\xi'\Phi(x_j)\right) \qquad (42) \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} u_j w\left(\xi'\Phi(x_j)\right) \\
&\quad -\sum_{j=1}^{N}\left(\exp(\widehat{\theta}' x_j) - \exp(\theta_0' x_j)\right) w\left(\xi'\Phi(x_j)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} u_j w\left(\xi'\Phi(x_j)\right) \\
&\quad -\left(\frac{1}{N}\sum_{j=1}^{N}\exp(\theta_0' x_j)x_j w\left(\xi'\Phi(x_j)\right)\right)' N\left(\widehat{\theta}-\theta_0\right) \\
&\quad +O_p(N^{-1}),
\end{aligned}
$$

where the last equality follows from Taylor's theorem. Note that the remainder term $O_p(N^{-1})$ is uniform in $\xi \in \Xi$. Moreover, it follows easily from the derivations in Appendix 2 that

$$
N\left(\widehat{\theta}-\theta_0\right) = B^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N} u_j \exp(\theta_0' x_j)x_j\right) + o_p(1). \tag{43}
$$

Substituting (43) in (42) now yields

$$
\begin{aligned}
\widehat{z}(\xi) &= \frac{1}{\sqrt{N}}\sum_{j=1}^{N} u_j w\left(\xi'\Phi(x_j)\right) \\
&\quad -\left(\frac{1}{N}\sum_{i=1}^{N}\exp(\theta_0' x_i)x_j w\left(\xi'\Phi(x_i)\right)\right)' B^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N} u_j \exp(\theta_0' x_j)x_j\right) \\
&\quad +o_p(1),
\end{aligned}
$$

which is the same expression as in the case $E(y_j|x_j) = g(x_j, \theta_0)$ with $g(x_j, \theta_0) = \exp(\theta_0' x_j)$. Consequently, the ICM test is directly applicable to model (41), without modification.

12 References

Bierens, H. J. (1982): "Consistent model specification tests", *Journal of Econometrics* 20, 105-134.

Bierens, H. J. (1990): "A consistent conditional moment test of functional form", *Econometrica* 58, 1443-1458.

Bierens, H. J. (1994): *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models.* Cambridge University Press.

Bierens, H. J. (2001): *EasyReg International 2001*, Department of Economics, Pennsylvania State University (Free econometric software downloadable from URL http://econ.la.psu.edu/~hbierens/EASYREG.HTM).

Bierens, H. J. and W. Ploberger (1997): "Asymptotic theory of integrated conditional moment tests", *Econometrica* 65, 1129-1151.

Fielding, A. J. (1989): "Migration and urbanization in western Europe since 1950", *The Geographical Journal*, 155: 60-69.

Flowerdew, R. and M. Aitken (1982): "A method of fitting the gravity model based on the Poisson distribution", *Journal of Regional Science* 22, 191-202.

Frey, W. H. (1987), "Migration and depopulation of the metropolis: regional restructuring or rural renaissance", *American Sociological Review*, 52: 240-257.

Frey, W. H. (1988): "Migration and metropolis decline in developed countries: a comparative national study", *Population Development Review*, 14: 595-628.

Frey, W. H. (1989): "United States: counterurbanization and metropolis depopulation". In: A. G. Champion (ed.), *Counterurbanization: The Changing Pace and Nature of Population Deconcentration*, London: Edward Arnold, 34-61.

Frey, W. H. (1993): "The new urban revival in the United States", *Urban Studies*, 30: 741-774.

Fuguitt, G. V. and C. L. Beale (1996): "Recent trends in nonmetropolitan migration: toward a new turnaround?", *Growth and Change*, 27: 156-174.

Hugo, G. J. and P. J. Smailes (1985): "Urban-rural migration in Australia: a process view of the turnaround", *Journal of Rural Studies*, 1: 11-30.

Johnson, K. M. (1999): "The Rural Rebound", *Population Reference Bureau. Reports on America*, 1: 1-22.

Kemper, F. J. (1999): "Binnenwanderungen und Dekonzentration der Bevölkerung - Jüngere Entwicklungen in Deutschland", *Berliner Geographische Arbeiten*, 90: 105-122.

Kontuly, T. (1998): "Contrasts of the counterurbanization experience in European nations". In: P.J. Boyle and K. Halfacree (eds.), *Migration into Rural Areas Theories and Issues*, London: John Wiley, 61-78.

Kontuly, T. and B. Dearden (1998): "Regionale Umverteilungsprozesse der Bevölkering in Europa seit 1970", *Informationen zur Raumentwicklung. Heft* 11/12: 713-722.

Kontuly, T. and R. Vogelsang (1988): "Explanations for the intensification of counterurbanization in the Federal Republic of Germany". *Professional Geographer*, 40: 42-54.

Long, L. and A. Nucci (1997): "The 'clean break' revisited: is U.S. population again deconcentrating?", *Environment and Planning A*, 29: 1355-1366.

Morrill, R. (1978): "Population redistribution, 1965-75", *Growth and Change*, 9: 35-43.

Moseley, M. J. (1984): "The revival of rural areas in advanced countries. A review of some causes and consequences", *Geoforum*, 15: 447-456.

Okoruwa, A.A., J. V. Terza, and H. O. Nourse (1988): "Estimating patronization shares for urban retail centers: An extension of the Poisson gravity model", *Journal of Urban Economics* 24, 241-259.

Okoruwa, A.A., H. O. Nourse, and J. V. Terza (1994): "Estimating sales for retail centers: An application of the Poisson gravity model", *Journal of Real Estate Research* 9, 85-97.

Plane, D. A. and P. A. Rogerson (1991): "Tracking the baby boom, the baby bust, and the echo generations: How age composition regulates U.S. migration", *Professional Geographer*, 43: 416-430.

Stinchcombe, M. B., and H. White (1998), "Consistent specification testing with nuisance parameters present only under the alternative", *Econometric Theory* 14, 295-325.

White, H.(1980): "A heteroskedasticity-consistent covariance matrix estimator, and a direct test for heteroskedasticity", *Econometrica* 48, 817-838.