I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11780

Interdependent Hazards, Local Interactions, and the Return Decision of Recent Migrants

Govert E. Bijwaard
Christian Schluter

AUGUST 2018

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11780

Interdependent Hazards, Local Interactions, and the Return Decision of Recent Migrants

Govert E. Bijwaard
Netherlands Interdisciplinary Demographic Institute and IZA

Christian Schluter
Aix-Marseille Université, CNRS & EHESS and University of Southampton

AUGUST 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Interdependent Hazards, Local Interactions, and the Return Decision of Recent Migrants*

Consider the duration of stay of migrants in a host country. We propose a statistical model of locally interdependent return hazards in order to examine whether interactions at the level of the neighbourhood are present and lead to social multipliers. To estimate this model we develop and study two complementary estimation strategies, demonstrate their good performance while standard non-spatial estimators are shown to be heavily biased. Using a unique large administrative panel dataset for the population of recent labour immigrants to the Netherlands, we quantify the local social multipliers in several factual and counterfactual experiments, and demonstrate that these are substantial.

JEL Classification: C41, C10, C31, J61

Keywords: interdependent hazards, local interaction, social multipliers, return migration

Corresponding author:
Govert E. Bijwaard
Netherlands Interdisciplinary Demographic Institute (NIDI)
University of Groningen
PO Box 11650
2502 AR The Hague
The Netherlands
E-mail: bijwaard@nidi.nl

* We thank Yann Bramoullé and Aureo de Paula for their comments, as well as the many conference and seminar participants. Financial support from Norface under the Choices project and ANR-DFG (grant ANR-15-FRAL-0007-01) is gratefully acknowledged.

1 Introduction

Much economic and social activity arises from and gives rise to local interactions at the level of the neighbourhood. Locally, social and geographic space coincides, and such interaction can lead to individual outcomes influencing and being influenced by outcomes of one's neighbours or peers, endogenous "neighbourhood effects" for short. This interaction gives rise in turn to local social multipliers which can amplify the social effect of idiosyncratic events. Here we consider the case of a duration outcome, specifically, the duration of recent labour immigration spells in the Netherlands. Such staying durations of immigrants in a host country is a key concern in the migration literature (see e.g. the survey by Dustmann and Görlach, 2016). Spatially, these immigrants not only cluster but also segregate along ethnic lines. In this setting, using Dutch administrative data of the entire population of recent labour immigrants, we seek to quantify the role of local interactions for the migrant's return (or outmigration) decision.

To this end, we propose a new statistical model of locally interdependent return hazards, and develop new methods for estimation and statistical inference. Specifically, an individual migrant's return hazard[1] is assumed to depend directly on the return hazard of others. This direct dependence captures the idea that an individual's return hazard is impacted on by the *propensity* of her peers or co-ethnics in the neighbourhood to return. Our model of interdependent hazards is thus proposed as a complement to the recent literature of social interactions in duration analysis which models directly durations as interdependent. This complementarity reflects our different empirical setting. In particular, de Paula (2009) and Honoré and de Paula (2010) obtain such dependence structurally by considering complete information 2-agent synchronisation games of optimal switching. We believe their modelling assumptions too demanding for our specific empirical setting: here, the numbers of peers is very large in some neighbourhoods, while information is necessarily incomplete, and ties are weak. For this reason, we have opted to allow directly the propensities of peers or co-ethnics to return to impact the individual hazard. For the same reason we do not model the individual hazard to depend on the peers' survival function, as in Sirakaya (2006) in a proportional hazard (PH) setting.

Our statistical model of hazards as interdependent has two antecedents. First, Lillard (1993) considers a simultaneous equation model for hazards in an accelerated failure time (AFT) setting. Specifically, he studies a statistical model in which the hazard of fertility depends directly on the hazard of marriage dissolution. The second antecedent is the so-called timing-of-events method (Abbring and van den Berg (2003)) in mixed proportional hazard (MPH) models. There, the hazard of one process depends directly on the duration and thus the hazard of another, and under the no-anticipation assumption the latter is interpreted as a dynamic causal treatment effect.[2] Rosholm and Svarer (2006) combine these two approaches in a study of a

[1]As is standard in duration analysis, we model directly the hazard function, "the focal point of econometric duration models" (van den Berg (2001)).

[2]For instance, Osikominu (2013) studies the effects of short-term job search-oriented training programmes on the time to job entry. In Bijwaard et al. (2014) we study the dynamic effects of unemployment on out-migration hazards. Drepper and Effraimidis (2015) examine the effect of first-time drug use on the hazard of drug taking by siblings.

model where the hazard rate out of unemployment depends on the hazard of entering an active labour market programme. In our case, the interdependence of the hazards arises because of local social interactions. As in the timing-of-events method we also rule out anticipatory effects, in that the expected hazards at some future duration have no effect on the hazard at the current duration. For these reasons, our approach also differs from models in which correlations between hazards arise because of correlated frailties.[3]

Our modelling approach has also close parallels to the recent literature of binary choice models with social interactions. Building on the pioneering insights of Brock and Durlauf (2001), Lee et al. (2014) examine this model under the assumption of heterogeneous expectations. The resulting statistical model is one of interdependent choice probabilities, and the equilibrium is the fixed point of this (continuous) mapping. The authors derive conditions under which this mapping defines a contraction, yielding a unique equlibrium. We follow this programme for our model of interdependent hazards, and obtain the conditions for a contraction mapping. In their empirical application, Lee et al. (2014) consider the incidence of smoking among adolescents under peer influence, and seek to disentangle empirically, in the language of Manski (1993) such endogenous effects -the influence of peer outcomes- from contextual effects - the influence of exogenous own and peer characteristics. In our case, the "outcome" is a hazard, and the perspective is dynamic in contrast to the static binary choice model. Interpreting the interaction as spatial at the level of the neighbourhood also enables us to contribute methodologically and empirically to a growing literature that confirms the importance of neighbourhood effects, while disentangling endogenous from contextual effects.[4]

The estimation of and inference for spatially interdependent hazard models requires new methods. In particular, although the statistical model has the structure of a mixed proportional hazard (MPH) model, the reduced form does not. We therefore develop and study two complementary estimation strategies. First, the reduced form suggests naturally the application of maximum likelihood techniques, which yields our spatial mixed proportional hazard (sMPH) estimator. Second, we propose a new spatial linear rank estimator (sLRE) that offers an interesting trade-off for applied work: while this estimator does not require the estimation of the distribution of individual-level unobvserved heterogeneity, it requires that the local social interaction parameter ρ be sufficiently small. We examine the performance of these two estimation approaches in several Monte Carlos. We first show that ignoring local interaction when they are present biases the standard MPH estimators. The induced spatial biases

[3]See e.g. Duffie et al. (2009) for a recent model. Such correlated frailties are often parametrised using copulas, see e.g. Goethals et al. (2008) for a discussion.

[4]For instance, Topa (2001) considers spatial dependence of unemployment rates in a setting in which spatial interaction arise from information spillovers. The spatial units are 863 census tracts in Chicago, and residents in one tract are assumed to exchange information locally with residents of the adjacent tracts. Instead of unemployment incidences, Gobillon et al. (2010) consider spatial differences in the duration of unemployment using administrative data for 1300 municipalities in the Paris region. In their paper the neighbourhood affects the outcome directly, thus defining an exogenous neighbourhood effect. Bayer et al. (2008) study the propensity of neighbours to work together by examining whether individuals residing in the same city block are more likely to work together than those in nearby blocks.

increases in ρ. The new sMPH estimates ρ very well throughout all experiments, while the sLRE is complementary for $\rho \leq .4$.

Turning to our empirical contributions, durations of stay in the host country are a key concern in the migration literature (Dustmann and Görlach, 2016). While there is a presumption in the literature that local social interactions could be important, standard methods do not permit empirical quantifications. Here, we bring together a new statistical model, new estimation strategies, and exploit exceptionally rich Dutch administrative data of the universe of recent immigrants.

The spatial clustering of immigrants is often a manifestation of local social networks at work (e.g. Munshi (2003), McKenzie and Rapoport (2010)), and local interactions are expected to be important for recent labour immigrants since they are newcomers to the host country and the local labour market.[5] Despite this presumption of their importance, severe data limitations - relating foremost to the size of the spatial units, to sample size and data reliability - have often prevented their rigorous empirical investigation. The usual data situation in migration analysis is one of small samples, possibly subject to selectivity and attrition issues, extracted from surveys of respondents who provide recall data; these problems are particularly acute in studies of migration durations since survey attrition usually confounds outmigration. If spatial units are reported in survey data at all, these are typically either municipalities or regions. Such spatial units are excessively large for analyses of *local* interactions.

We overcome these empirical data challenges using a unique administrative panel for the entire *population* of *recent labour immigrants* to the Netherlands covering the years 1999-2007, which is extensively described in Data Appendix B. The data characteristics -large size, repeated and accurate measurement- are fairly unique in migration analysis, as is the spatial unit, the neighbourhood. This Dutch immigrant register is based on the legal requirement for immigrants to register with the authorities upon arrival. Moreover, natives as well as immigrants are required to register with their municipality. Several other official registers are linked by Statistics Netherlands to this immigrant register, such as the social benefit and the income register (used by the tax authorities). Sojourn times in the Netherlands, in a specific neighbourhood, and in labour market states are thus recorded accurately. Consequently, no data based on individual recall has to be used, and the administrative population has no attrition. Moreover, the usual concerns about measurement error are less acute. Another attractive feature of our data is the administrative report in the immigrant register (consistent with the visa status at entry) of the immigration motive. This enables us to focus explicitly and exclusively on *labour* immigrants. The immigration motive is usually latent in standard datasets, and our previous work (Bijwaard (2010)) has confirmed that the systematically different behavioural patterns of labour and non-labour migrants confound the empirical analysis. The size of our population data of recent labour migrants permits us to consider specific groups. As in Adda et

[5]See e.g. Bartel (1989), and Logan et al. (2002) for an examination of ethnic immigrant enclaves in the US, and Clark and Drinkwater (2002) for the UK. Zorlu and Mulder (2008) observe for the Dutch case, the subject of our empirical investigation, that "in some neighbourhoods in The Hague, Amsterdam and Rotterdam, the share of non-Western foreigners has reached levels above 70 per cent and even 80 per cent" (p.1902). Such spatial concentration strongly suggests the presence of local interactions.

al. (2015) in the context of Germany, we consider here the largest ethnic group of recent labour immigrants, namely Turkish labour immigrants (about 8000 individuals), and refer to return and out-migration as "return" for short.

Our dataset identifies the neighbourhood the immigrant lives in, defined by Statistics Netherlands as areas that include approximately 2,000 households on average. There are about 14,000 neighbourhoods. The extent of spatial clustering and segregation among the four principal ethnic immigrant groups (Turks and three others for comparison) is extensively documented in Data Appendix B.1 and B.2. For instance, the Lorenz curve analysis shows that about 80% (70%) of this immigrant population lives in about the 20% (10%) most concentrated neighbourhoods, and the mapping of the 100 most concentrated neighbourhoods for each group reveals little spatial overlap. This descriptive evidence suggests that there is scope for local interactions resulting in interdependent return hazards.

The econometric analysis confirms the importance of such local social interactions. The estimated social interaction parameter is fairly high ($\hat{\rho} = .75$), which then induces a large social multiplier whose role increases as durations increase: local social interactions accelerate migrants' return probabilities, and this acceleration increases as the duration of stay increases. This multiplier effect (and its interaction with covariates) is further explored in several counterfactual experiments that manipulate individual migrant profiles and global push and pull factors. By contrast, the usual MPH estimator is shown to fail to account for these effets, which results in a substantial under-estimation of return probabilities. For instance, in our illustration, at month 60 the sMPH-based predicted return probability is about 3 times larger than the spatially biased MPH estimate. The new methods developed in this paper are thus also empirically important.

The outline of this paper is as follows. In the next section, the empirical setting is presented in greater detail, as it informs our modelling choices. In Section 3, we present and discuss the properties of the statistical model of locally interdependent hazards. Our two complementary approaches to estimation are set out in Section 4. We then study their performance in several Monte Carlos. Section 5 is devoted to the empirical analysis. We also quantify the social multipliers for return probabilities in several factual and counterfactual experiments, where we vary pull and push factors as well as immigrant characteristics. All proofs are collected in the Appendix A. Appendix B contains a detailed description of the data, as well as evidence about the spatial clustering and segregation of the different immigrant groups.

2 The empirical setting: The population of recent Turkish labour migrants

We proceed to present in some detail our empirical setting, focussing on spatial aspects, since the features of our statistical model of locally interdependent hazards will be informed by this. We consider the population of recent labour immigrants to The Netherlands, who have entered the host country during our observation window 1999-2007. The administrative data, covering this entire population, is extensively described in Data Appendix B.1. Specifically, we focus on the largest ethnic group, namely Turkish immigrants (as in Adda et al. (2015) in the context of Germany). In (Data) Appendix B.2 we have extensively documented that immigrants not only cluster but also spatially segregate along ethnic lines. This applies particularly to Turkish immigrants, and suggests that there is scope for local interactions as captured by the econometric model.

Table 1: Summary statistics by neighbourhood concentration: Recent Turkish labour migrants

	neighbourhoods				
	all	top 50	top 100	top 200	not top 200
N	7617	1109	1687	2582	5034
% Female	21	18	20	20	21
% Married	66	62	68	72	63
% with children	21	22	22	23	20
Average age at entry	28	28	28	27	28
Income at entry					
0 < income p.m. < €1000 [%]	77	65	73	78	77
Neighbourhood average :					
% Turks	10	42	33	35	2
% unemployed	3.5	3.2	3.3	3.4	3.5
average income [$\times 10^3$]	11.4	8.4	8.9	9.4	12.3
global :					
quaterly unempl. rate at entry	2.9	3.1	3.0	3.0	2.9
Length of stay at return [%]					
< 6 months	4.1	4.6	4.3	4.3	4.0
6-12 months	19.3	42.5	37.8	31.3	10.5
12-18 months	14.2	15.9	15.1	14.2	14.0
18-24 months	12.9	18.9	16.7	16.1	10.6
24-60 months	35.8	11.7	18.0	23.7	44.6
> 5 years	13.8	6.4	8.0	10.3	16.3
censoring[a][%]	80.2	60.6	70.2	75.9	82.5

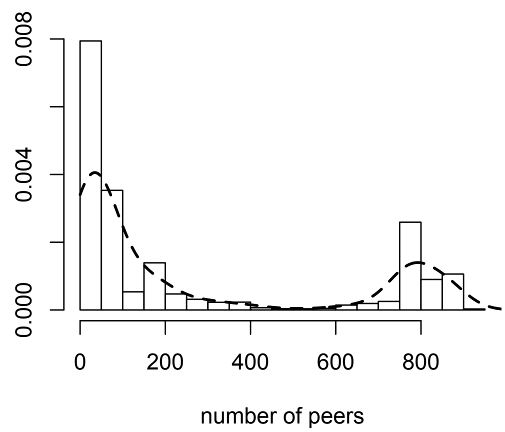
Notes. Summary statistics for all recent Turkish labour immigrants for the subpopulations residing in the 50, 100 or 200 most concentrated neighbourhoods in terms of the Turkish population.

[a] Migrants who remain in the country until the end of the observation period.

Table 1 provides selective summary statistics for our data. In order to explore spatial difference, we also contrast these summary statistics for all recent Turkish labour immigrants for the subpopulations residing in the 50, 100 or 200 most concentrated neighbourhoods (in terms of the Turkish population). 15% of our individuals reside in the top 50 neighbourhoods, and 22% in the top 100 neighbourhoods. We complete the description of the spatial concentration with Figure 1, where we plot the histogram and the kernel density estimate of the number of other (recently arrived) Turkish labour migrants ("peers" for short). It is evident that the density is bimodal, with a substantial number of individuals having many connections. The first and third quartile are 27 and 657 connections, while the median and mean number of connections are 75 and 263.

The majority of these recent Turkish labour immigrants are men, albeit married, and fairly young at arrival, the mean age being 28 years. These labour immigrants are typically poor, as the vast majority earn less that €1000 p.m. in their first job after entry. Turks living in more concentrated neighbourhoods are more often male and less often on very low incomes in the first job. The most concentrated neighbourhoods exhibit slightly lower unemployment rates (e.g. 3.2% compared to an average 3.5%), but also average lower incomes (e.g. €8.4K compared to €11.4K).

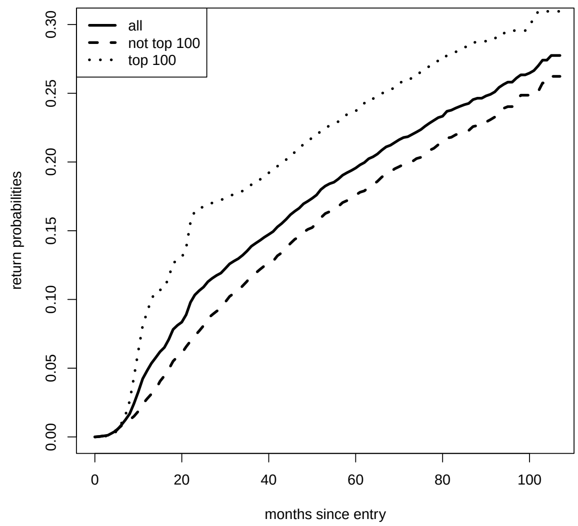
Figure 1: Histogram of the number of peers in neighbourhood



Notes. Also included is the kernel density estimate (dashed line).

Next, we turn to the outcome of interest, namely the time spent in the host country. Table 1 indicates that while censoring is high, the incidence of outmigration (we do not observe destinations) or return ("return" for short) is also substantial. Among Turkish returnees, 24% have left again within one year of arrival. Turks living in more concentrated neighbourhoods leave more often and faster. In order to take into account the censoring of the data, we consider next the non-parametric Kaplan Meier estimates of the return probabilities. As this estimator ignores spatial dependences, the estimator is biased, and used here only as a descriptive benchmark. In order to explore whether spatial difference are in evidence in the data, we juxtapose these Kaplan Meier estimates for recent Turkish labour immigrants residing in and outside the 100 most concentrated neighbourhoods. Figure 2 suggests that for all survival times, Turks in the 100 most concentrated neighbourhoods (22% of our data) have higher probabilities of return. In particular, for all survival times after 20 months, the spatial difference is around 7 percentage points. Hence these comparisons between concentrated and not concentrated neighbourhoods indicate important spatial differences, but cannot distinguish between systematic differences in neighbourhood characteristics (contextual effects) and endogenous local interaction. We disentangle these using a new statistical model that is presented next.

Figure 2: Kaplan Meier estimates of the return probabilities by neighbourhoods



Notes. All Turcs (solid line), Turks in the top 100 most concentrated neighbourhoods in terms of the Turkish population (dotted line), representing 22% of our data, and its complement (dashed line).

3 Local social interactions in duration models

We first present the new statistical model of locally interdependent hazards, before discussing its key aspects in the context of the leading literature. We then study identification, and propose new methods for estimation and statistical inference.

3.1 Interdependent hazards

Since we consider a duration outcome, we follow the literature and focus directly on the hazard function, "the focal point of econometric duration models", our point of departure being a mixed proportional hazard (MPH) specification, as "MPH models are the most popular reduced-form duration models in econometrics" (van den Berg (2001), who provides an extensive survey). The observational units are recent labour immigrants in the host country (the Netherlands), the duration variate, denoted by T, is the time spent in the Netherlands, and the associated hazard is denoted by λ.

For expositional clarity, we present first a restricted model of the migration duration which ignores the endogenous social interaction effect. This non-spatial model will also serve as a natural benchmark in our assessment of spatial biases presented below in Section 4.3.

The proportional hazard (PH) model expresses this return hazard as the product between a baseline hazard, $\lambda_0(t, \alpha)$, which is a function of the duration (or time-to-event) alone (and a parameter vector α) and common to all individuals, and a covariate function, $\exp(x(t)\beta)$, which accelerates or decelerates exits: $\lambda(t|\overline{x}(t);\theta) = \lambda_0(t,\alpha)\exp(x(t)\beta)$ with $\theta = (\alpha', \beta')'$ and $\overline{x}(t)$ being the history of the covariate process $x(.)$ up to duration t. The parameter space Θ is assumed to be convex. The covariate vector $x(t)$ is allowed to change over time, but we assume that their sample paths are piecewise constant, i.e. the derivative with respect to t is 0 almost everywhere, and left continuous. As regards the baseline hazard, we assume that $\lambda(t,\alpha)$ is a positive function, that it is twice differentiable, and that its second derivative is bounded in α and t. In our empirical application the baseline hazard is modelled as piecewise constant, so $\lambda_0(t,\alpha) = \exp\big(\alpha_0 + \alpha'_{-0}A(t)\big)$ with $A(t) = \big(I_1(t),\dots,I_k(t)\big)'$ denoting the vector of interval indicators. In order to accommodate unobserved heterogeneity, the mixed proportional hazard model (MPH) extends the PH model by multiplying it by a time-invariant person-specific positive error term, say v with some distribution G, assumed to be independent of the covariate process: $\lambda(t|\overline{x}(t),v;\theta) = v\lambda_0(t,\alpha)\exp(x(t)\beta)$. It is well known that both baseline hazard and G are non-parametrically identifiable (see e.g. Elbers and Ridder (1982)), so that genuine duration dependence can be distinguished from dynamic sorting, provided that some restrictions are imposed on one of these two objects: either v has a finite mean, or the tail behaviour of G is restricted, or $\lambda(t,\alpha)$ is positive and finite for t close to zero.

We augment this basic MPH specification by including local interactions that lead to a direct interdependence of the hazards. However, it will also turn out that the resulting reduced form is no longer of the MPH form (see (3) below). In order to record whether any two individuals are neighbours, we follow common practice in spatial and social network econometrics and collect this information in a spatial interaction matrix W. Specifically, consider migrant i in a particular neighbourhood. Denote the set of i's co-ethnics in the same immigration cohort residing in the same or close-by neighbourhood (the "peers" for short) by $N_i(t)$ with $i = 1,..,n$. Since individuals might move during the observation window, this set can vary with duration t. The number of relevant peers is denoted by $\#N_i(t)$. The resulting $n \times n$ spatial interaction matrix is $W(t) = [w_{ij}(t)]_{i=1,..,n;j=1,..,n}$ with $w_{ii}(t) = 0$, $w_{ij}(t) = 1/\#N_i(t)$ if $j \in N_i(t)$ and zero otherwise. Hence all peers of i have the same weight, and these weights sum to 1. In our empirical application W is non-sparse, and its size will render it challenging to invert. We address this problem below.

We are now in the position to formalise the idea that an individuals return hazard is impacted on by the propensity of her peers in the neighbourhood to return: Hazards are assumed to be locally interdependent, as the return hazards of peers of migrant i ($\{j \in N_i(t) : \lambda_j\}$) influence and are influenced by the return hazard of migrant i. This interdependence gives rise to the *endogenous* local effect.[6] Since we are working within the MPH paradigm, it is consistent to assume that this effect is proportional, so that it is the geometric mean

$$\left[\prod_{j \in N_i(t)} \lambda_j\right]^{1/\#N_i(t)}$$

[6]Such endogenous effects differ fundamentally from exogenous spatial effect, as modelled in e.g. Gobillon et al. (2010) in their study of spatial effects on unemployment durations, using a PH model in which baseline hazards are estimated for each location.

that impacts on i's hazard λ_i. We thus obtain the following model of interdependent hazards:

$$\lambda_i(t|.) = v_i \lambda_0(t, \alpha) \exp\Big(x_i(t)\beta + \rho w_{i.}(t) \ln\Big[\underline{\lambda}(t)\Big]\Big), \tag{1}$$

with $\underline{\lambda}(t) \equiv [\lambda_i(t|.)]_{i=1,..,n}$ denoting the $n \times 1$ vector whose ith element is $\lambda_i(t|.)$ and $w_{i.}$ denoting the ith row of W. The coefficient ρ captures the strength of the endogenous local interaction effect, and we seek to estimate below this local social interaction parameter. Taking logs of equation (1) shows that the local interaction effect is modelled in a way similar to the standard linear model used in network and spatial econometrics.[7]

The statistical model captures the idea that an individual's return hazard is impacted on by the *propensity* of her peers or co-ethnics in the neighbourhood to return. We believe that this paradigm is appropriate in our setting, in which possibly many peers co-reside in a neighbourhood (recall Figure 1, where the second mode is centered around 800), but no one is completely informed about the decisions of survivors. It is for this reason that, unlike Sirakaya (2006) in a proportional hazard (PH) setting, we do not model the individual hazard to depend on the peers' survival function.

Our statistical model of interdependent hazards has two antecedents. First, Lillard (1993) considers a simultaneous equation model for hazards in an accelerated failure time (AFT) setting in which the log hazard is modelled linearly. Specifically, he studies a statistical model in which the hazard of fertility depends directly on the hazard of marriage dissolution. The second antecedent is the so-called timing-of-events method (Abbring and van den Berg (2003)) in MPH models. There, the hazard of one process depends directly on the duration and thus the hazard of another, and under the no-anticipation assumption the latter is interpreted as a dynamic causal treatment effect. Rosholm and Svarer (2006) combine these two approaches in a study of a model where the hazard rate out of unemployment depends directly on the hazard of entering an active labour market programmes. In our case, the interdependence of the hazards arises because of local social interactions. As in the timing-of-events method we also rule out anticipatory effects, in that the expected hazards at some future duration have no effect on the hazard at the current duration.[8]

Our analysis of locally interdependent hazards contributes to and is proposed as a complement to the recent literature of social interactions in duration analysis that generates interdependent durations. These are obtained structurally in de Paula (2009) and Honoré and de Paula (2010) by considering complete information 2-agent synchronisation games of optimal switching. We believe their modelling assumptions too demanding for our specific empirical setting: here, the numbers of peers is very large in some neighbourhoods, while information is necessarily incomplete, and ties are weak. For this reason, we have opted to allow directly the propensities of peers or co-ethnics to return to impact the individual hazard. For the same reason we believe

[7] This standard linear model is, to be precise, $y = \alpha + x\beta + Wx\delta + \rho Wy + \epsilon$, where $y = [y_i]_{i=1,..,n}$ is a $n-$vector of outcomes, x is the matrix collecting exogenous characteristics and W is the social interaction matrix. As in our case, ρ captures endogenous neighbourhood effect, and δ the exogenous neighbourhood effect. See e.g. Bramoullé et al. (2009) for an analysis of identification in this linear model.

[8] For these reasons, our approach also differs from models in which correlations between hazards arise because of correlated frailties (e.g. Duffie et al. (2009) for a recent model).

that an application of the timing-of-events framework to be problematic in our setting, as this would require the presence of one central or pioneering individual whose duration-to-first-exit influences all other return hazards in the neighbourhood.[9]

3.2 Heterogeneous rational expectations

It is of interest to note that our model of interdependent hazards given by (1) has a structure that is very similar to the binary choice model of social interactions with heterogeneous expectations. The properties of the binary choice model with homogeneous expectations have been extensively analysed in the pioneering work of Brock and Durlauf (2001). Recently, Lee et al. (2014) have considered the case of heterogeneous expectations, and have studied a model of interdependent choice probabilities. There,[10] as in our case, the vector of equilibrium equations depends on network characteristics as well as the characteristics of each of its members, captured by the vector x_i. They obtain conditions under which this mapping is a contraction, leading to a unique solution. We follow this programme in our setting.

Specifically, define the rational expectations equilibrium as the vector $\underline{\lambda}^* = (\lambda_1^*, \cdots, \lambda_n^*)$ such that model (1) is coherent, namely

$$\lambda_i(t|.)^* = v_i \lambda_0(t,\alpha) \exp\Big(x_i(t)\beta + \rho w_{i.}(t) \ln\Big[\underline{\lambda}^*(t)\Big]\Big),$$

for $i = 1, \cdots, n$. This defines a system, $\underline{\lambda} = M(\underline{\lambda})$, where M is $n \times 1$, and the ith element is given by the right hand side of (1). Since this mapping is continuous, the existence of the equilibrium follows from Brouwer's fixed point theorem. Uniqueness follows if M defines a contraction. We follow Lee et al. (2014) and consider the maximum row sum norm, denoted by $||.||_\infty$. The derivative of the ith element of M with respect to λ_j equals $\rho w_{ij} \lambda_i / \lambda_j$. Let $\lambda_{max} = max_{i,j \in N(i)} \lambda_i/\lambda_j \geq 1$. Then the maximum row sum norm of the gradient of M satisfies

$$||\frac{\partial M(\underline{\lambda})}{\partial \underline{\lambda}'}||_\infty \quad = \quad |\rho| \max_{i=1,\ldots,n} \sum_{j=1}^{n} |w_{ij}| \frac{\lambda_i}{\lambda_j} \leq |\rho| \lambda_{max} ||W||_\infty = |\rho| \lambda_{max}$$

since W is row-sum normalised. Hence the mapping M is a contraction for the maximum row sum norm if $|\rho| < \lambda_{max}^{-1} \leq 1$, i.e if the social interaction parameter is sufficiently small.

3.3 The reduced form

Let $\underline{v} = [v_i]_{i=1,..,n}$, $X(t) = [x_i(t)]_{i=1,..,n}$, and $\overline{X}(t)$ and $\overline{W}(t)$ denote the history of the covariate process and the spatial interactions. Then solving (1) yields the reduced form

$$\underline{\lambda}\big(t|\theta, \rho, \overline{X}(t), \overline{W}(t), \underline{v}\big) = \exp\big(H(t;\rho) X^*(t)\theta + H(t;\rho) \ln \underline{v}\big) \tag{2}$$

[9] Hence our setting differs fundamentally from that of e.g. Drepper and Effraimidis (2015) who examine the effect of first-time drug use of one sibling on the hazard of drug taking of the remaining siblings.

[10] Specifically, their model of individual i's choice of $d = 1$ is $p_i = F(x_i\beta + \rho w_{i.}\underline{p})$, with $\underline{p} = (p_1, \cdots, p_n)'$. Brock and Durlauf (2001) show that this choice probability can be rationalised by a model of conformity, where utility is specified as a quadratic distance between an individual's and the expected choices of her peers.

with $H(s;\rho) = (I - \rho W(s))^{-1}$ and $X^*(s) = (\ln \underline{\lambda}_0(s); X(s))$. For notational convenience, we suppress the explicit conditioning on the covariate and the spatial processes $(\overline{X}(t), \overline{W}(t))$. For $(I - \rho W(t))^{-1}$ to be well-defined, we require that ρ be smaller than the inverse of the absolute value of the largest eigenvalue of $W(t)$. As W can change with time, consider the smallest of the upper limits, and define the feasible convex set for ρ thus defined by $\Theta_W \equiv \cap_t \Theta_{W(t)}$. The ith element of $\underline{\lambda}$ is given by

$$\begin{aligned}
\lambda_i(t|\cdot) &= \exp\big(e_i'(H(t;\rho)X^*(t)\theta + H(t;\rho)\ln \underline{v})\big) \\
&= \left[\prod_j v_j^{H_{ij}}\right] \exp\left(\beta \sum_j H_{ij}x_j\right) [\lambda_0(t,\alpha)]^{H_{i\Sigma}} \qquad (3)
\end{aligned}$$

where e_i is a $(n \times 1)$ selection vector that has a one in the ith position and zeros everywhere else, H_{ij} is the $(i,j)^{th}$ element of the matrix $H(s;\rho)$, and $H_{i\Sigma} = \sum_j H_{ij}$.

The correlation structure implied by equation (3) makes clear that we no longer have a MPH model since, depending on the structure of local interactions, unobservables $v_{j\neq i}$ can influence the i's hazard even though all v_j are iid random variables.

3.4 Identification

Identification in MPH models chiefly center on the issue of separating separating out dynamic sorting from duration dependence. It is well known that under some assumptions about the distribution of v identification is achieved (e.g. Elbers and Ridder (1982)). Since our reduced form is no longer a MPH model, we cannot invoke the classic identification results. However, identification is achievable:

Theorem 1 *Assume that unobserved heterogeneity v_i is independent and identically distributed according to G with mean μ. Then the model's parameters are identified.*

Identification is strengthened if not all individuals are neighbours, and the network structure exhibits symmetry (specifically $W_{1n} = W_{n1} = 0$, and the last and first row of W are identical), so local interdependencies in the comparison at time $t = 0$ between individuals 1 and n cancel out; or if there are disconnected neighbourhoods of different sizes. The simulation study reported below in Section 4.3 provides a numerical illustration of the identification of the model.

3.4.1 Threats to identification

If individuals purposefully locate into particular neighbourhoods, it is conceivable that the unobserved heterogeneity terms v are correlated across individuals. While the proof of theorem 1 has imposed the independence assumption, it also suggests how we can relax this independence hypothesis, since we have first exploited systematic within-neighbourhood variation. In particular, we can generalise the empirical model to allow for systematic variation in unobservable heterogeneity across the disconnected neighbourhoods (such as "high" v. "low" mean v neighbourhoods), without materially affecting the proof. Within a neighbourhood, however, we have to maintain the independence assumption.

The threat to identification in our specific empirical context is further lessened by the following two observations that concern our specific empirical setting of recent Turkish labour immigration to the Netherlands. First, we control for systematic observable variation in neighbourhood characteristics by including contextual effects such as the local unemployment rate as regressors. Second, the location choices of our population of interest are severely constrained: Our population of interest are recent migrants, who, because of being recent, do not qualify for social transfer and protection programmes such as social housing. Moreover, at arrival, most Turkish labour immigrants are poor (see Table 1). In the Netherlands, social housing represents a large stock of accommodation-for-rent in the poorer neighbourhoods, which is not available to these immigrants. Hence, poor recent labour immigrants are unlikely to be able to choose one particular neighbourhood because of the lack of affordable housing.

4 Estimation

The estimation of and inference for spatially interdependent hazard models requires new methods: although the statistical model (1) has the structure of a mixed proportional hazard (MPH) model, the reduced form (3) does not. We therefore develop and study two complementary estimation strategies. First, the reduced form suggests naturally the application of maximum likelihood techniques, which yields our spatial mixed proportional hazard (sMPH) estimator. Second, we propose a new spatial linear rank estimator (sLRE) that offers an interesting trade-off for applied work: while this estimator does not require the estimation of the distribution of individual-level unobvserved heterogeneity, it requires that local social interaction parameter ρ be sufficiently small. In Section 4.3 we examine the performance of these two estimation approaches in several Monte Carlos.

4.1 The spatial MPH estimator (sMPH)

Since we have already obtained the individual return hazard in equation (3), it is then natural to consider estimating the model by maximum likelihood, whilst dealing with the distribution of unobserved heterogeneity using the classic techniques proposed in Heckman and Singer (1984). We label the resulting approach the spatial MPH estimator (sMPH).

Specifically, using (3), the implied reduced form survival function for individual i is given by

$$S_i(t|\cdot) = \exp\Big(-\int_0^t \lambda_i(s|\cdot)\,ds\Big) \tag{4}$$

As regards the baseline hazard, we assume here a piecewise constant function, i.e. $\lambda_0(t) = \sum_{r=1}^{R} e^{\alpha_r} I_r(t)$ with $I_r(t) = I(t_{r-1} \leq t < t_r)$ and $t_0 = 0$,$t_R = \infty$, where R denotes the total number of intervals considered. Any duration dependence can be approximated arbitrarily closely by increasing the number of intervals. For identification, we assume that the baseline hazard equals unity in the first interval, i.e. $\alpha_1 = 0$. We approximating the unobserved heterogeneity distribution, as proposed by Heckman and Singer (1984), by a discrete distribution, with $p_k = \Pr(v = e^{V_k})$.

Thus the likelihood contribution of migrant i conditional on the unobserved heterogeneity $v = v_k$ is

$$L_i(v) = \lambda_i^{\Delta_i}(s|\cdot)S_i\big(t|\cdot\big) \tag{5}$$

where Δ_i is the indicator for the event that migrant i's spell is uncensored. Integrating out the unobserved heterogeneity we obtain the likelihood function

$$L = \prod_{i=1}^{n} \int L_i(v)\, dG(v) \tag{6}$$

where $G(v)$ is the (discrete) distribution of the unobserved heterogeneity terms.

Since we have a standard maximum likelihood framework, the distributional theory and hence statistical inference are standard, and not spelt out further for the sake of brevity.

4.2 The spatial linear rank estimator (sLRE)

We propose a new estimation framework that does not require the specification of the (frailty) distribution of unobserved heterogeneity. It extends to our spatial setting a linear rank estimator (LRE), based on ideas in Tsiatis' (1990) developed in the context of a (non-spatial) AFT model and further developed in Bijwaard et al. (2013). The LRE is the root of a sample moment condition, which is based on the comparison between the value of a covariate for individual i at a (transformed and uncensored) duration, and the average value of the covariate for all survivors at this duration. If a covariate were independent of the hazard, then the mean of the covariate among the survivors does not change with the survival time and equals the unconditional mean. Of course, the covariate process at survival time t, $x_i(t)$, does affect the hazard $\lambda_i(t|.)$. But the appropriate transformation, yielding independence and hence a theoretical moment of zero, is the integrated hazard, i.e. the generalised residual (generalised residuals have also been used in a spatial probit model by Pinkse and Slade (1998)).

To this end, consider the transformation model of the random duration variate T given by

$$\begin{aligned} U_i = h_i(T, \theta, \rho) &= \int_0^T \exp\big(e_i' H(s;\rho) X^*(s)\theta\big) ds \\ &= \int_0^T \lambda_i(s|.)ds/\psi_i ds \end{aligned} \tag{7}$$

which is the integrated hazard except for the function of the unobservable heterogeneity terms, where ψ_i is the ith element of $\exp(H(t;\rho)\ln \underline{v})$. U_i is also known as a generalised residual. For the population parameter vector (θ_0, ρ_0) the transformation model is denoted by $U_{i,0} \equiv h_{i,0}(T)$ with $h_{i,0}(T) \equiv h_i(T, \theta_0, \rho_0)$, as is $\psi_{i,0}$ and $H_0(s) \equiv H(s;\rho_0)$. Conditional on $\underline{v}$ and the covariate and the spatial processes, the integrated hazard $\int_0^T \lambda_i(s|.)ds$ has a unit exponential distribution. It follows that $U_{i,0}$ is a positive random variable that is independent of the covariate and the spatial processes and the baseline hazard (this is shown formally in appendix equation (A11)). This independence is the basis of the fundamental moment condition which the linear rank estimator exploits.

In order to accommodate the possibility that some spells are right censored at some predetermined date C (in our case the end of our observation window), assuming that censoring is uninformative, define the observation indicator $\Delta(t) = I(T > t)I(t < C)$.

Consider then the random sample of size n of $(T_i, \Delta_i, \overline{x}_i(T_i))$. The transformation model transforms the durations for some θ to $(U_i(\theta), \Delta_i, \overline{x}_i(U_i(\theta)))$. Rank the transformed durations, and let $U_{(i)}(\theta)$ denote the i's order statistic. The moment condition compares the expected value of the covariates to the expected value for the survivors across all transformed survival times. This population moment condition is zero for the population parameters θ_0 given the above independence result. The sample analogue is

$$S_n(\theta) = \sum_{i=1}^{n} \nu_i \Delta_i \Big[x_i\big(U_{(i)}\big) - \tilde{x}\big(U_{(i)}\big) \Big] \tag{8}$$

where

$$\tilde{x}\big(U_{(i)}\big) = \frac{\sum_{j=i}^{n} I\big(U_{(j)} \geq U_{(i)}\big) x_j\big(U_{(i)}\big)}{\sum_{j=i}^{n} I\big(U_{(j)} \geq U_{(i)}\big)} \tag{9}$$

is the sample mean of the covariates for survivors at the transformed survival time, and ν_i is a weighting function. Using the theory of counting processes (with intensity given by the hazard of $U_{i,0}$), we show in the Appendix Lemma 6 the following result:

Lemma 1 *The counting measure $N^{U_{i,0}}(u)$ does not depend on the covariate and the spatial processes, hence $E(S_n(\theta_0, \rho_0)) = 0$.*

Rather than defining the linear rank estimator as the root of the sample analogue, we define it to be the minimiser of the associated quadratic form,

$$(\hat{\theta}, \hat{\rho}) = \underset{\theta \in \Theta, \rho \in \Theta_W}{\operatorname{argmin}} \; S_n(\theta, \rho)' S_n(\theta, \rho), \tag{10}$$

since the sample moment condition $S_n(\theta, \rho)$ is a step function.

The estimating function is, in general, not monotone in the parameters, but monotonicity ensues using Gehan weights $\nu_i = \sum_{j=i}^{n} I\big(U_{(j)} \geq U_{(i)}\big)$ (Fygenson and Ritov (1994)).

4.2.1 Distributional theory and statistical inference

The sample moment condition $S_n(\theta, \rho)$ is a step function, so S_n is not differentiable everywhere, and the distributional theory for $(\hat{\theta}, \hat{\rho})$ cannot be based on the usual asymptotic analysis which uses a first order expansion. However, applying the arguments in Tsiatis (1990), we can consider an asymptotically equivalent function $\tilde{S}_n(\theta, \rho)$ that is linear in (θ, ρ) in the neighbourhood of (θ_0, ρ_0). In the Appendix, we show using the theory of counting processes the following results:

Theorem 2 *$(\hat{\theta}, \hat{\rho})$ is consistent, and is distributed asymptotically as a normal variate.*[11]

[11] The variance is obtained by the delta method. The theoretical gradient matrix depends on the distribution of U_0 (see appendix), which we approximate, as in Bijwaard (2009), by Hermite polynomials using the exponential distribution as a weighting function. Chung et al. (2013) survey alternative approaches.

4.2.2 An approximation for large spatial interaction matrices

The size of the spatial interaction matrix $W(t)$ renders the minimisation of the criterion function in (10) computationally challenging. We overcome this challenge by following the approach of Klier and McMillen (2008), and consider an approximation of the moment condition about $\rho = 0$ and $\theta_0 = \hat{\theta}_1$ where $\hat{\theta}_1$ is the solution to the minimisation problem in (10) in the absence of spatial interactions, $\hat{\theta}_1 = \text{argmin}_{\theta \in \Theta} S_n(\theta, 0)' S_n(\theta, 0)$.

The resulting linear approximation of the spatial rank-functions is

$$S(\theta_0, \rho_0) \approx S(\hat{\theta}_1, 0) + G(\hat{\theta}_1, 0) \times \begin{pmatrix} \hat{\theta}_1 - \theta \\ 0 - \rho \end{pmatrix} \tag{11}$$

where $G(\theta, \rho) = (\partial S/\partial \theta, \partial S/\partial \rho)$, which is stated explicitly in the appendix (equation (A23)). Setting this linear approximation to zero and solving yields the one-step procedure for the joint estimation of the parameters of the hazard θ, and the spatial dependence ρ,

$$\begin{pmatrix} \hat{\theta} \\ \hat{\rho} \end{pmatrix} = \begin{pmatrix} \hat{\theta}_1 \\ 0 \end{pmatrix} + \left(G'G\right)^{-1} G' S(\hat{\theta}_1, 0) \tag{12}$$

To summarise, we propose a two-stage estimation strategy: In the first stage, obtain $\hat{\theta}_1$ from the minimisation of (10) ignoring spatial interactions, using X^* as instruments. $S(\theta_0, \rho_0)$ is based on the instruments (X^*, WX^*). We update the first-stage estimates using the one-step estimator in (12).

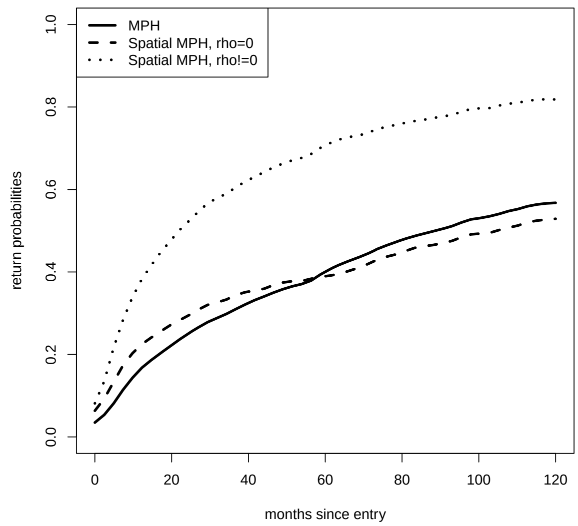
4.3 Simulation evidence

We present performance evidence for our two new spatial estimators.

The Monte Carlo design seeks to replicate several key features of the data used in the empirical application. In particular, in order to generate a realistic pattern of local interactions, the interaction matrix W used in the simulation is the same as in our Dutch data. We also impose a high incidence of censoring, namely of 40% and 70%. Our chosen specification is parsimonious. We assume that individuals are possessed of a time-invariant covariate x that is a random draw from the uniform distribution. Unobserved heterogeneity is modelled as a two-point mass distribution with selection probability .2. The baseline hazard is piecewise constant, with jumps at months 36 and 60. In particular, the covariate coefficient is $\beta = -1$, and the coefficients in the baseline hazard function are $\alpha_0 = (.2, .8)'$. The Monte Carlos are repeated 100 times.

Table 2 presents the results for $\rho \leq .4$ and both 40% and 70% censoring , while Table 3 focuses on higher values of ρ and a high incidence of censoring for brevity. As a benchmark, we also report the results for a standard MPH specification that wrongly ignores the local interactions leading to the interdependence of the hazards. We juxtapose these MPH results, and our two spatial estimators labelled sMPH and sLRE. In order to illustrate the quantitative effect of local social interactions on the outcome of interest, Figure 3 depicts the implied estimates of the return probabilities for one case ($\rho = .3$ and 40% censoring). It is evident that such interaction considerably increase the return probabilities at all durations. The figure also illustrates the considerable spatial bias of the usual MPH approach.

Figure 3: Quantifying the effect of social interactions: $\rho = .3$



Notes. Illustration for $\rho = .3$ and 40% censoring. Dotted line: based on the sMPH estimator. Dashed line: based on the sMPH estimate, setting $\rho = 0$. Solid line: spatially biased estimates based on the usual MPH model. The effect of social interactions is given by the difference between dotted and dashed line.

The results illustrate that the usual MPH estimator suffers from spatial distortions, and these increase in ρ. For instance, consider the case of a censoring incidence of 40%. Although for small spatial interactions ($\rho = .1$) the estimate of the covariate coefficient $\beta = -1$ is very good, the distortions become sizeable as ρ increases. For $\rho = .4$, the mean estimate of β is -.823. The estimated coefficients for the baseline hazard function also progressively worsen.

Our new spatial maximum likelihood estimator (sMPH) captures very well the local social interaction parameter even when ρ is high. For instance, for a population value of $\rho = .4$, the mean estimate is .401 for 40% censoring and .39 for 70% censoring. Throughout, ρ is fairly precisely estimated. The same conclusion applies to the estimate of the covariate coefficient β. Even when $\rho = .7$, the mean estimate is -.973 with a standard deviation of .051. For larger values of ρ the estimates of the coefficients of the baseline hazard exhibit some distortions. Throughout, as censoring increases from 40% to 70%, the performance is only marginally affected. One effect of the loss of information caused by the increase in censoring is increased variability of the estimates.

Our new spatial linear rank estimator (sLRE) is only valid for sufficiently low values of ρ, and the experiments confirm this numerically. For a population value of $\rho = .4$ the mean estimate of ρ is .51 and 40% censoring, for higher population values the distortions increase. For the range $\rho \leq .4$ the estimates of ρ are unbiased. However, compared to the sMPH estimates they exhibits substantially more variability (which, of course, is to be expected). This variability also increases as the incidence of censoring increases. These observations also apply to the estimates of β. sLRE and sMPH perform similarly in terms of the estimates of the baseline hazard coefficients.

We conclude that for sufficiently low values of the interaction parameter ρ the sLRE approach offers a complementary estimator to the sMPH. One the one hand, the sLRE is dominated by the sMPH in all experiments. On the other, its principal attraction is that, by design, the distribution of unobserved heterogeneity does not need to be considered explicitly. In situations in which both sMPH and sLRE are applicable, we therefore suggest that both estimators should be considered.

Table 2: Simulation evidence I ($\rho \leq .4$)

		$\rho = .1$			$\rho = .2$			$\rho = .3$			$\rho = .4$	
	MPH	sMPH	sLRE	MPH	sMPH	sLRE	MPH	sMPH	sLRE	MPH	sMPH	sLRE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						40% censoring						
$\hat{\rho}$		0.104	0.001		0.207	0.189		0.306	0.339		0.401	0.508
		(0.013)	(0.004)		(0.013)	(0.142)		(0.014)	(0.162)		(0.015)	(0.190)
$\hat{\beta}$	−0.988	−0.998	−1.002	−0.949	−0.990	−0.998	−0.891	−0.982	−1.009	−0.823	−0.970	−0.999
	(0.051)	(0.049)	(0.067)	(0.052)	(0.045)	(0.060)	(0.055)	(0.048)	(0.069)	(0.059)	(0.049)	(0.077)
$\hat{\alpha}_2$	0.189	0.167	0.212	0.119	0.140	0.165	0.001	0.122	0.161	−0.141	0.099	0.114
	(0.103)	(0.094)	(0.144)	(0.107)	(0.080)	(0.123)	(0.123)	(0.070)	(0.102)	(0.173)	(0.066)	(0.099)
$\hat{\alpha}_3$	0.839	0.689	0.883	0.794	0.596	0.770	0.683	0.531	0.721	0.562	0.483	0.575
	(0.142)	(0.123)	(0.210)	(0.145)	(0.101)	(0.216)	(0.166)	(0.086)	(0.222)	(0.243)	(0.078)	(0.260)
						70% censoring						
$\hat{\rho}$		0.101	0.005		0.199	0.148		0.296	0.366		0.390	0.546
		(0.014)	(0.042)		(0.015)	(0.183)		(0.015)	(0.216)		(0.015)	(0.276)
$\hat{\beta}$	−1.019	−1.021	−1.007	−0.994	−1.010	−1.007	−0.971	−1.003	−1.017	−0.943	−0.989	−0.993
	(0.051)	(0.068)	(0.126)	(0.075)	(0.068)	(0.108)	(0.089)	(0.062)	(0.116)	(0.081)	(0.058)	(0.100)
$\hat{\alpha}_2$	0.221	0.194	0.194	0.200	0.173	0.178	0.143	0.129	0.127	0.113	0.100	0.073
	(0.155)	(0.150)	(0.296)	(0.163)	(0.126)	(0.223)	(0.189)	(0.111)	(0.189)	(0.185)	(0.094)	(0.158)
$\hat{\alpha}_3$	0.896	0.775	0.848	0.884	0.701	0.764	0.851	0.622	0.614	0.840	0.539	0.418
	(0.156)	(0.181)	(0.434)	(0.210)	(0.152)	(0.335)	(0.259)	(0.135)	(0.351)	(0.243)	(0.113)	(0.343)

The population parameters are $\beta = -1$, $\alpha_2 = .2$, and $\alpha_3 = .8$.

Table 3: Simulation evidence II ($\rho > .4$ and 70% censoring)

	$\rho = .5$		$\rho = .6$		$\rho = .7$	
	MPH	sMPH	MPH	sMPH	MPH	sMPH
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\rho}$		0.484		0.577		0.672
		(0.015)		(0.016)		(0.015)
$\hat{\beta}$	−0.918	−0.975	−0.890	−0.971	−0.841	−0.973
	(0.076)	(0.060)	(0.079)	(0.059)	(0.077)	(0.051)
$\hat{\alpha}_2$	0.064	0.059	0.033	0.037	−0.043	0.007
	(0.147)	(0.075)	(0.134)	(0.066)	(0.131)	(0.051)
$\hat{\alpha}_3$	0.829	0.458	0.830	0.384	0.922	0.360
	(0.218)	(0.105)	(0.205)	(0.091)	(0.194)	(0.068)

5 Empirical analysis: Estimates and counterfactual experiments

We turn to the estimation of our model of locally interdependent return hazards. We consider a parsimonious specification in terms of individual effects. Included in the estimation as contextual effects are not only the neighbourhood averages of these characteristics (WX), but also additional neighbourhood descriptors (the local rate of unemployment and the average income level). Further included are global covariates such as the unemployment rate at the national level in the quarter of arrival, and time effects in terms of year of entry (which control for global push and pull factors). The baseline hazard is piecewise constant, with permitted jumps at month 12, 36, and 60.

In Table 4 we report the coefficient estimates of the usual MPH, and our sMPH and sLRE approaches. The MPH estimates are of course biased, and reported here in order to establish a benchmark, as it is of interest to quantify how distorted in practice the usual MPH approach that ignores local interactions is. In order to better interpret the overall effects, we complement the table with Figure 4, where we plot the ultimate object of interest, namely the predicted return probability $1 - S(t)$, where S denotes the survivor function (given generically by equation (4)) and covariates are set to their mean values.

Our sMPH approach yields an estimate of the local interaction parameter ρ that is fairly high, $\hat{\rho} = .75$, but below unity. This magnitude implies that the alternative sLRE-based estimate is likely to be distorted. The table confirms this, the estimate being .35. Despite this (expected) discrepancy, we interpret the positive and statistically significant estimate as providing complementary evidence on the role of local interactions: the effect is positive, so returns are accelerated by interactions, and sizeable.

The estimated magnitude of the local interaction parameter also implies that the usual MPH estimator will be significantly distorted. The table confirms this, as many of the estimated covariate coefficients differ substantially in magnitude. Figure 4 quantifies these discrepancies in terms of the predicted return probabilities. The MPH prediction (solid line) is substantially below the sMPH based prediction (dotted line). For instance, in months 60 (which corresponds to a new interval in the baseline hazard), the latter is more than three times larger than the former.

Turning to the estimates of the individual-level coefficients, they have the expected signs: more settled migrants (married, with children, females) and more economically successful migrants are less likely to exit. The usual MPH approach, however, substantially over-estimates the role of these individual characteristics. The sLRE-based estimates lie about halfway between the MPH and sMPH estimates. The sMPH-based estimates of the coefficients of the baseline hazard function are monotonically increasing, so there is positive duration dependence: exit proabilities increase as t moves across time intervals.

How large is the social multiplier generated by local interactions ? We quantify it by using the sMPH-based estimates, and generate a return prediction after imposing that ρ be zero. This counterfactual prediction is depicted in Figure 4 by the dashed line. Compared to the factual prediction given by the dotted line, it is evident that for durations less than 60 months, the differences, while present, are fairly small. For higher durations, however, these become substantial, as the sizeable increase in the return probabilities in the counterfactual prediction (driven by the baseline hazard) is amplified by large positive local interactions.[12]

We conclude that social local interactions are substantial, and imply a social multiplier that becomes substantial at higher durations.

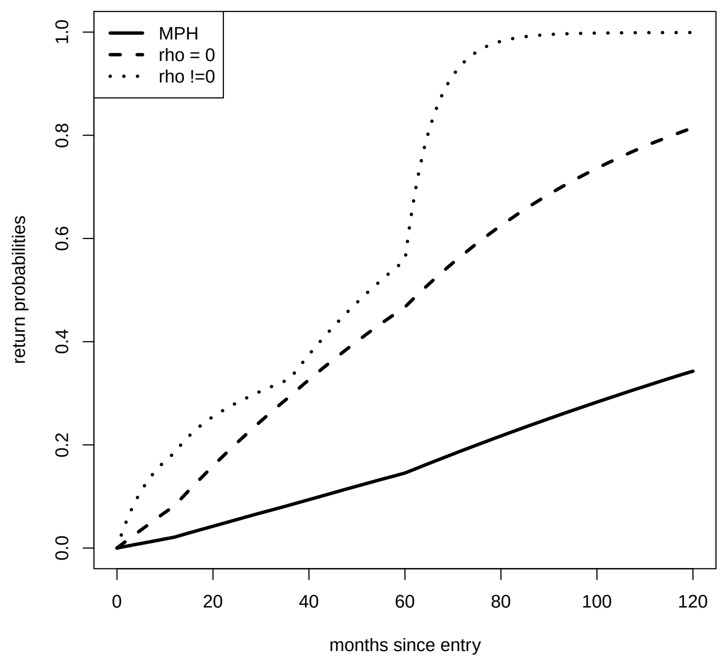
[12]For very high durations (e.g. after 80 months) the differences between predicted return probabilities for the mean individual based on the biased MPH and the unbiased sMPH estimates appear very large. However, in order to place the predictions and their differences into the empirical context recall that we have an observation window of 9 years, only 13.8% of spells are longer than 5 years, and the incidence of censoring in the data is 80% so that many exits are not observed. Of the observed (uncensored) exits, 50.5% are below 2 years. Hence, this discrepancy in the predictions at very high durations is consistent with the presence of large local interactions.

Table 4: Results

	MPH	sMPH	sLRE
ρ		0.750^{**}	0.351^{+}
		(0.013)	0.179
income at entry >1000	-1.424^{**}	-0.496^{**}	-1.105^{**}
	(0.066)	(0.049)	(0.078)
Female	-0.448^{**}	-0.158^{+}	-0.269^{**}
	(0.074)	(0.067)	(0.087)
married	-1.221^{**}	-0.247^{**}	-0.797^{**}
	(0.067)	(0.054)	(0.087)
divorced	-0.275^{+}	0.188^{+}	-0.059
	(0.108)	(0.089)	(0.133)
number of children	-0.339^{**}	-0.165^{**}	-0.247^{**}
	(0.020)	(0.020)	(0.029)
U (national)	0.267^{**}	0.034	0.057
	(0.035)	(0.030)	(0.051)
U (local) at entry	-0.413^{**}	-0.123	-0.487^{**}
	(0.114)	(0.098)	(0.162)
baseline hazard:			
α_2 (1-3 yr)	0.468^{**}	0.333^{**}	-0.014
	(0.086)	(0.042)	(0.139)
α_3 (3-5 yr)	0.544^{**}	0.633^{**}	-0.163
	(0.107)	(0.064)	(0.212)
α_4 ($>$ 5 yr)	0.963^{**}	1.140^{**}	-0.029
	(0.140)	(0.099)	(0.238)

All regressions include year-of-entry fixed effects, and neighbourhood averages measured by Wx. U denotes unemployment. $^{+}p < 0.05, ^{**}p < 0.01$

Figure 4: The effect of local interactions on the return probability for the average Turkish labour immigrant



Notes. Depicted are the predicted return probabilities, where covariates are set at their mean in the population. Coefficients as per Table 4. Solid line: MPH-based (spatially biased) prediction, dotted line: sMPH-based prediction, dashed line: sMPH-based counterfactual prediction after imposing $\rho = 0$.

5.1 Counterfactual scenarios: The scope for the social multiplier

In order to further explore and quantify how return probabilities, our outcome of interest, are amplified by local social interactions and the resulting social multipliers, we consider several counterfactual scenarios that capture different pull and push factors and immigrant profiles. Some of these scenarios could be though of as being under some policy control (e.g. immigrant targeting based on characteristics) while others relate to events largely outside the control of policy makers. Throughout we take the social interaction matrix W as given, and vary the immigrant profiles.

In experiment (a, labelled “higher incomes”), we consider a situation in which only higher skilled immigrants enter the host country (measured here by having incomes above €1000 p.m. in their first job after entry). In scenario (b, labelled “all female”) only female Turkish labour immigrants enter the host country . Recall that both females and higher earners have lower return hazards. In experiment (c, labelled “Unemployment”) we increase the Dutch national rate of unemployment in the quarter of arrival to 8% . In experiment (d, labelled “Entry in 2006”), we assume that immigrants arrive in a wave in 2006. The last two experiments capture push factors of events in the host and source country, while the first two experiments consider the effect on outcomes when the immigrant profile has counterfactually changed. The objective of these experiments is to explore how the magnitude of the social multipliers depend on the scenario and the duration of stay.

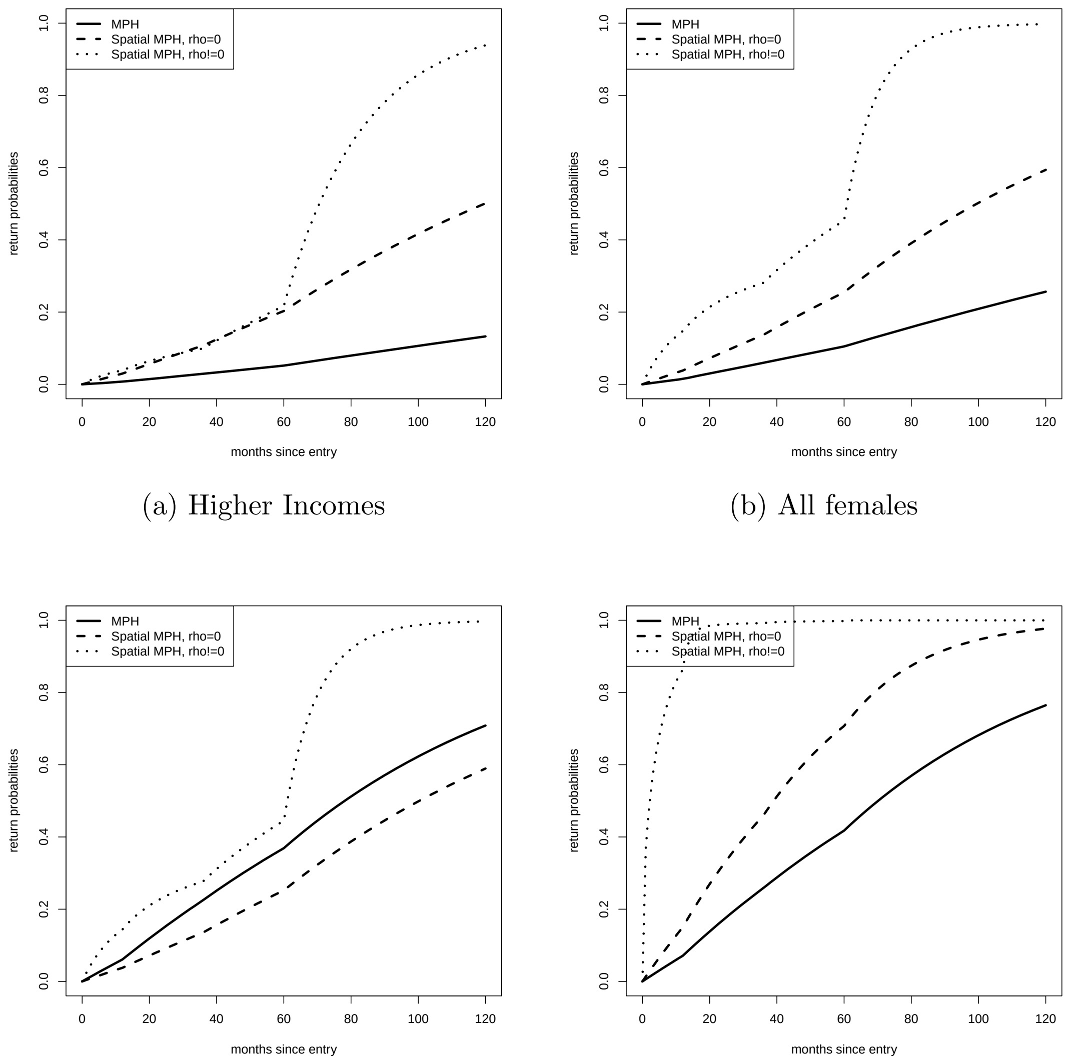
We proceed as in Figure 4, by focussing on the predicted return probabilities. For the sake of completeness, we also report the spatially biased usual MPH-based estimate, in order to quantify the extent of the spatial biases of the usual approach. Throughout, these distortions are substantial for all durations and all scenarios.

The main focus of our analysis is the quantification of the social multiplier. For each counterfactual scenario, we thus compare the sMPH-based prediction with $\rho = .75$ (dotted lines) and $\rho = 0$ (dashed lines). In all settings, the social multiplier is substantial for high durations (in excess of 60 months). Whether the multiplier is also present for short duration depends on the scenario: there is no immediate effect in scenario (a) while it is substantial in scenario (d); scenarios (b) and (c) present some intermediate cases. We conclude that the different scenarios serve to illustrate how strongly local social interactions affect staying durations: social multipliers are substantial.

6 Conclusion

Individuals are distributed across neighbourhoods, cluster, interact locally, and individual specific outcomes might influence and be influenced by the outcomes of one's peers. Focussing on outcomes that are durations, we have studied an econometric model of locally interdependent hazards in terms of identification, estimation, and inference. Our particular empirical application of this general framework is set in the context of recent Turkish labour immigration to The Netherlands, and we have studied, specifically, the impact of local social interactions on the duration of stay and the resulting social multipliers. Using administrative data for this entire (sub)population,

Figure 5: Social multiplier effects on return probabilities in counterfactual scenarios



(a) Higher Incomes

(b) All females

(c) Unemployment

(d) Entry in 2006

Notes. For given counterfactual scenario (described in main text), lines as per Figure 4.

we find strong evidence that the propensity of ones "peers" (i.e. co-ethnics in the same immigration cohort residing in the same or close-by neighbourhood) to return increases ones own return hazard, which, in turn, accelerates the return of ones peers. Our quantifications have revealed that the resulting social multipliers are substantial.

References

Abbring, J. H. and G. J. van den Berg (2003) "The non–parametric identification of treatment effects in duration models", *Econometrica*, 71, 1491–1517.

Adda, J., Dustmann, C., and J-S. Görlach (2015), "Migrant Wages, Human Capital Accumulation and Return Migration", Working Paper UCL.

Andersen, P.K and R. D. Gill, (1982), "Cox's Regression Model for Counting Processes: A Large Sample Study", *Annals of Statistics*, 10, 1100-1120.

Bartel, A.P., (1989), "Where do the new United States immigrants live ?", *Journal of Labour Economics*, 7, 371-391.

Bayer, P., S.L. Ross, and G. Topa (2008), "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes", *The Journal of Political Economy*, 116, 1150-1196.

Bijwaard, G. E. (2009), "Instrumental variable estimation for duration data", in H. Engelhardt, H.-P. Kohler, and A. Furnkranz-Prskawetz (Eds.), *Causal Analysis in Population Studies: Concepts, Methods, Applications*, 111-148. SpringerVerlag.

Bijwaard, G. E. (2010), "Immigrant Migration Dynamics Model for The Netherlands",*Journal of Population Economics*, 23, 1213-1247.

Bijwaard, G.E., G. Ridder and T. Woutersen (2013), "A Simple GMM Estimator for the Semi-parametric Mixed Proportional Hazard Model", *Journal of Econometric Methods*, 2, 1-23.

Bijwaard, G.E., C. Schluter and J. Wahba (2014), "The Impact of Labour Market Dynamics on the Return-Migration of Immigrants", *The Review of Economics and Statistics*, 96, 483-494.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009), "Identification of peer effects through social networks", *Journal of Econometrics*, 150, 41-55.

Brock, W. A., and S.N. Durlauf (2001), "Discrete Choice with Social Interaction," *Review of Economic Studies*, 68, 2, 235-260.

Chung, M. and Long, Q., and B.A. Johnson (2013), "A tutorial on rank-based coefficient estimation for censored data in small- and large-scale problems", *Statistics and Computing*, 23, 601-614.

Clark, K. and S. Drinkwater (2002), "Enclaves, neighbourhood effects and employment outcomes: Ethnic minorities in England and Wales", *Journal of Population Economics*, 15, 5-29.

Cutler, D.M., E.L. Glaeser, and J.L. Vigdor (1999), "The rise and decline of the American Ghetto", *Journal of Political Economy*, 107, 455-506.

de Paula, A. (2006). "Inference in a synchronization game with social interactions," *Journal of Econometrics*, 148, 57-71.

Drepper, B. and G. Effraimidis (2015), “Social Interaction Effects in Duration Models”, Working Paper, University of Mannheim.

Duffie,D., Eckner, A., Horel, G., and L. Saita (2009), “Frailty correlated default”, *Journal of Finance*, 64, 5, 2089-2123.

Dustmann, C. and J.-S. Görlach (2016), “The economics of temporary migrations”, *Journal of Economic Literature*, 54, 98-136.

Elbers, C. and G. Ridder (1982), “True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model”, *The Review of Economic Studies*, 49, 403-409.

Fygenson, M. and, Ritov, Y. (1994), “Monotone estimating equations for censored data”, *Annals of Statistics*, 22, 732-74.

Gobillon, L., T. Magnac, and H. Selod (2010), “The effect of location on finding a job in the Paris region”, *Journal of Applied Econometrics.*

Goethals, K., P. Janssen, and L. Dutchateau (2008) “Frailty models and copulas: Similarities and differences”, *Journal of Applied Statistics*, 35, 1071-1079.

Heckman, J.J. and J.R. Walker (1987). “Using Goodness of Fit and Other Criteria to Choose among Competing Duration Models: A Case Study of Hutterite Data”, *Sociological Methodology*, Vol. 17, 247-307.

Honoré, B. E. and A. De Paula (2010), “Interdependent Durations”, *Review of Economic Studies*, 77, 1138-1163.

Klier, T. and D. P. McMillen (2008), “Clustering of auto supplier plants in the United States: Generalized method of moments spatial logit for large samples”, *Journal of Business & Economic Statistics*, 26, 460-471.

Lee, L-f., Li, J., and X. Lin (2014), “Binary choice models with social network under heterogeneous rational expectations”, *The Review of Economics and Statistics*, 96, 3, 402-417.

Lillard, L. (1993), “Simultaneous equations for hazards”, *Journal of Econometrics*, 56, 189-217.

Logan, J.R., Alba, R.D. and Zwang, W. (2002), “Immigrant enclaves and ethnic communities in New York and Los Angeles”, *American Sociological Review*, 67, 299-322.

Manski, C. (1993), “Identification of endogenous social effects: The reflection problem”, *Review of Economic Studies*, 60, 531-542.

McKenzie, D. and H. Rapoport (2010), “Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks”, *The Review of Economics and Statistics*, 92, 811-821.

Munshi, K. (2003), “Networks in the Modern Economy: Mexican Migrants in the US Labor Market”, *The Quarterly Journal of Economics*, 549-599.

Osikominu A. (2013) “Quick job entry or long-term human capital development? The dynamic effects of alternative training schemes”, *Review of Economic Studies*, 80, 313-342.

Pinkse, J. and M. E. Slade (1998), “Contracting in space: An application of spatial statistics to discrete-choice models”, *Journal of Econometrics*, 85, 125-154.

Rosholm, M. and M. Svarer (2008). “The threat effect of active labour market programmes”, *Scandinavian Journal of Economics*, 110, 2, 385-401.

Sirakaya, S. (2006). “Recidivism and social interactions”, *Journal of the American Statistical Association*, 101, 475, 863-877.

Topa, G. (2001), “Social Interactions, Local Spillovers and Unemployment”, *The Review of Economic Studies*, 68, 261-295.

Tsiatis, A. A. (1990), “Estimating regression parameters using linear rank tests for censored data”, *Annals of Statistics*, 18, 354372.

Van den Berg, G. J. (2001) “Duration models: Specification, identification, and multiple duration”. In: Heckman J. J., Leamer E. (eds.), *Handbook of Econometrics, Vol V*, Chapter 55. Amsterdam: North–Holland, 3381–3460.

Zorlu, A., and C. H. Mulder (2008), “Initial and Subsequent Location Choices of Immigrants to the Netherlands”, *Regional Studies*, 42, 245-264.

A Technical Appendix: Proofs and Derivations

A.1 Proof of Theorem 1

Assume that individual characteristics x_i are time invariant, and scalar. To simplify notation, these will be suppressed in the conditioning statements. Let $H_{i\Sigma} = \sum_j H_{ij}$.

The reduced form model is

$$\begin{aligned}
\lambda_i(t|\underline{v}) &= \exp\left(H_{i.}X^*\theta + H_{i.}\log \underline{v}\right) \\
&= \left[\prod_j v_j^{H_{ij}}\right] \exp\left(\beta \sum_j H_{ij}x_j\right) \left[\lambda_0\left(t,\alpha\right)\right]^{H_{i\Sigma}}
\end{aligned}$$

The survival function of individual i is, conditional on the vector $\underline{v}$,

$$\begin{aligned}
\overline{F}_{T_i}\left(t|\underline{v}\right) &= \exp\left(-\int_0^t \lambda_i(s|\underline{v})ds\right) \\
&= \exp\left(-\left[\prod_j v_j^{H_{ij}}\right] \exp\left(\beta \sum_j H_{ij}x_j\right) z_0\left(t\right)\right)
\end{aligned}$$

with $z_0\left(t\right) = \int_0^t \left[\lambda_0\left(s,\alpha\right)\right]^{H_{i\Sigma}} ds$. Integrating out the unobservable heterogeneity yields

$$\overline{F}_{T_i}\left(t\right) = \int_{v_n} \cdots \int_{v_1} \exp\left(-z_0\left(t\right)\exp\left(\beta \sum_j H_{ij}x_j\right) \prod_j v_j^{H_{ij}}\right) dG\left(v_1\right) \cdots dG\left(v_n\right)$$

Consider the first individual, and the first integration with respect to v_1. Let $\mathcal{L}$ denote the Laplace transform of G (and the subscript on v has been suppressed since vs are identically distributed). We have $H\left(\rho\right) = I + \rho W + O(\rho^2)$, which implies that $H_{i\Sigma} = \sum_j H_{ij} = 1 + \rho + O(\rho^2)$, and $H_{11} = 1 + O(\rho^2)$. The survival function for the first individual is thus

$$\overline{F}_{T_1}\left(t\right) = \int_{v_2} \cdots \int_{v_N} \mathcal{L}\left(z_0\left(t\right)\exp\left(\beta \sum_j H_{1j}x_j\right) \prod_{j\neq 1} v_j^{H_{1j}}\right) dG\left(v_2\right) \cdots dG\left(v_N\right) \quad \text{(A1)}$$

We can then follow ideas first explored in Elbers and Ridder (1982). Differentiating the survival function with respect to time yields

$$\begin{aligned}
-f_{T_1}\left(t\right) &= \int_{v_2} \cdots \int_{v_N} \mathcal{L}'\left(z_0\left(t\right)\exp\left(\beta \sum_j H_{1j}x_j\right) \prod_{j\neq 1} v_j^{H_{1j}}\right) \exp\left(\beta \sum_j H_{1j}x_j\right) \\
&\times \left[\lambda_0\left(t,\alpha_0\right)\right]^{H_{1\Sigma}} \prod_{j\neq 1} v_j^{H_{1j}} dG\left(v_2\right) \cdots dG\left(v_N\right)
\end{aligned}$$

and letting $t \downarrow 0$ yields, since $z_0(t) \to 0$,

$$\lim_{t\downarrow 0} -f_{T_1}\left(t\right) = E\left(v\right)\exp\left(\beta \sum_j H_{1j}x_j\right) \prod_{j\neq 1} \mu_{1j} \lim_{t\downarrow 0} \left[\lambda_0\left(t,\alpha_0\right)\right]^{H_{i\Sigma}} \quad \text{(A2)}$$

with nuisance parameters $\mu_{1j} = \int v^{H_{1j}} dG(v)$. Symmetric expressions obtain for the other individuals. The idea here is to avoid the problem of dynamic sorting (and the systematic change of v in the stock of survivors) by considering the situation at the beginning, i.e. $t \downarrow 0$. We also seek to eliminate the distributional effect of G by comparisons between individuals.

Let's consider different structures of social interactions. Note that W is not necessarily symmetric, since not all individuals might be neighbours. In particular, in a tridiagonal structure, the first and last rows of W will have a different structure, i.e. $W_{12} = 1 = W_{n(n-1)}$. Throughout we will assume that the covariates exhibit sufficient variation ($x_i \neq x_{j\neq i}$).

A.1.1 A neighbourhood of two individuals

We have

$$W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

so $\mu_{12} = \mu_{21}$. We then have

$$\lim_{t\downarrow 0} \frac{f_{T_1}(t)}{f_{T_2}(t)} = \frac{\exp(\beta[H_{11}x_1 + H_{12}x_2])}{\exp(\beta[H_{22}x_2 + H_{12}x_1])} = \exp(\beta(1 - H_{12})(x_1 - x_2))$$

where $H_{12} = \rho + O(\rho^2)$. This implies that β and ρ are jointly identified, but we cannot separate them out yet. This will be done below. Before, we consider how a larger neighbourhood adds identifying information.

A.1.2 A neighbourhood of three individuals

Assume that not all individuals are neighbours, so wlog assume $W_{13} = 0$ but $W_{23} \neq 0$. We have

$$W = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}$$

so $\mu_{13} = \mu_{31}$, $\mu_{12} = \mu_{32}$, and $\mu_{21} = \mu_{23}$. We then have, considering individuals 1 and 3,

$$\lim_{t\downarrow 0} \frac{f_{T_1}(t)}{f_{T_3}(t)} = \exp(\beta[x_1 + (H_{12} - H_{32})x_2 - x_3]) = \exp(\beta(x_1 - x_3))$$

which identifies β. Considering individuals 1 and 2

$$\lim_{t\downarrow 0} \frac{f_{T_1}(t)}{f_{T_2}(t)} = \frac{\mu_{12}}{[\mu_{21}]^2} \exp\left(\beta[(1 - \rho/2)x_1 + (\rho - 1)x_2 - (\rho/2)x_3] + O(\rho^2)\right)$$

If G is identified, then given the identification of β, the identification of ρ follows.

Note that if all individuals were neighbours, than we could only jointly identify β and ρ. In particular, $W_{12} = W_{13} = 1/2$, and we would have $\lim_{t\downarrow 0} f_{T_1}(t)/f_{T_3}(t) = \exp(\beta(1 - \rho/2)(x_1 - x_3) + O(\rho^2))$, and similarly $\lim_{t\downarrow 0} f_{T_1}(t)/f_{T_2}(t) = \exp(\beta(1 - \rho/2)(x_1 - x_2) + O(\rho^2))$. Also note that although there are 3 individuals, we have only 2 independent ratios since $f_{T_2}(t)/f_{T_3}(t) = (f_{T_1}(t)/f_{T_3}(t))/(f_{T_1}(t)/f_{T_2}(t))$.

A.1.3 A neighbourhood of four individuals

Assume that not all individuals are neighbours, and consider the following situation

$$W = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0.5 & 0.5 & 0 \end{pmatrix}$$

As the first and last rows are the same, we have

$$\lim_{t \downarrow 0} \frac{f_{T_1}(t)}{f_{T_4}(t)} = \exp\left(\beta(x_1 - x_4)\right)$$

which identifies β.

By contrast, consider the situation where individual 4 is only connected to individual 2:

$$W = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Note that, compared to the previous structure, the greater isolation of individual 4 does not help identification. Considering individuals 1 and 3 yields

$$\lim_{t \downarrow 0} \frac{f_{T_1}(t)}{f_{T_3}(t)} = \exp\left(\beta(1 - \frac{\rho}{2})(x_1 - x_3)\right)$$

but considering individuals 3 and 4 say

$$\lim_{t \downarrow 0} \frac{f_{T_3}(t)}{f_{T_4}(t)} = \frac{\mu_{31}\mu_{32}}{\mu_4} \exp\left(\beta \left[\frac{\rho}{2}(x_1 - x_2) + (x_3 - x_4)\right]\right)$$

Comparing then across different neighbourhood structures, we find that identification is strengthened by symmetry properties of W: β is already identified if interdependencies cancel out which happens when $W_{1n} = W_{n1} = 0$ and the first and last row are identical. This also happens if the spatial structure W consists of disconnected neighbourhoods of different sizes (e.g. combining the 2-person case with the first three person case).

A.2 Identification of G

Next, we deal with the unknown distribution G. Wlog consider the two-person neighbourhood case. Inverting (A1), we have, say,

$$z_0(t) = \left[\exp\left(\beta \sum_j H_{1j} x_j\right)\right]^{-1} \Psi(\overline{F}_{T_1}) \tag{A3}$$

where the RHS does not depend on x since the LHS does not. This enables us to follow similar steps as in Elbers and Ridder (1982) to yield, for any observationally equivalent structure (denoted by tildes), an equation of the form

$$\tilde{\Psi}(s) = \frac{\tilde{C}}{C} \Psi(s) \tag{A4}$$

with $s = \overline{F}_{T_1}$ and $t = K(s)$. Note that for $s = 1$, $t = 0$, $z_0(0) = 0$, so $\Psi(1) = 0$. We can differentiate (A4) several times under standard regularity conditions

$$\tilde{\Psi}'(s) = \frac{\tilde{C}}{C}\Psi'(s), \tilde{\Psi}''(s) = \frac{\tilde{C}}{C}\Psi''(s), \cdots, \tag{A5}$$

to establish that $\tilde{C} = C$.

We have $\Psi(s) = z_0(K(s)) \exp\left(\beta \sum_j H_{1j}x_j\right)$ and the inverse of Ψ is given, of course, by (A1). Differentiating the latter,

$$\frac{d}{dy}\Psi^{-1}(y) = \int_v \mathcal{L}'(yv^{H_{12}})v^{H_{12}}dG(v)$$

This implies, at $s = 1$ (which implies $y = 0$),

$$\Psi'(s)|_{s=1} = \left[\int \mathcal{L}'(0)v^{H_{12}}dG(v)\right]^{-1} = \frac{1}{E(v)}\left[\int v^{H_{12}}dG(v)\right]^{-1}$$

For the alternative admissible structure (where $\tilde{G}$ has the same mean as G, say $\mu = E(v)$) we have

$$\tilde{\Psi}'(1) = \frac{1}{E(v)}\left[\int v^{\tilde{H}_{12}}d\tilde{G}(v)\right]^{-1}$$

so

$$\frac{\tilde{C}}{C} = \frac{\tilde{\Psi}'(1)}{\Psi'(1)} = \frac{\int v^{H_{12}}dG(v)}{\int v^{\tilde{H}_{12}}d\tilde{G}(v)} \tag{A6}$$

Considering now $\Psi''(s)$ should give us another equation for $\frac{\tilde{C}}{C}$ involving the second moments of v, and equalising with the preceding equality should give us an equation that can only be satisfied if $\tilde{G} = G$. We have

$$\Psi''(s) = (-1)\left[\frac{d}{dy}\Psi^{-1}(y)\right]^{-2} \frac{d}{dy^2}\Psi^{-1}(y)|_{y=\Psi(s)}$$

hence

$$\begin{aligned}
\Psi''(1) &= \frac{1}{E(v)^2}\left[\int v^{H_{12}}dG(v)\right]^{-2} \frac{d}{dy}\int \mathcal{L}'(yv^{H_{12}})v^{H_{12}}dG(v)|_{y=\Psi(s)} \\
&= \frac{1}{E(v)^2}\left[\int v^{H_{12}}dG(v)\right]^{-2} \mathcal{L}''(0)\int v^{2H_{12}}dG(v) \\
&= \frac{Var(v) + E(v)^2}{E(v)^2}\left[\int v^{H_{12}}dG(v)\right]^{-2}\int v^{2H_{12}}dG(v)
\end{aligned}$$

Writing again $\mu = E(v)$, we have

$$\frac{\tilde{C}}{C} = \frac{\tilde{\Psi}''(1)}{\Psi''(1)} = \frac{Var(\tilde{v}) + \mu^2}{Var(v) + \mu^2}\left[\frac{\int v^{\tilde{H}_{12}}d\tilde{G}(v)}{\int v^{H_{12}}dG(v)}\right]^{-2} \frac{\int v^{2\tilde{H}_{12}}d\tilde{G}(v)}{\int v^{2H_{12}}dG(v)} \tag{A7}$$

and equalising with (A6) yields

$$1 = \frac{Var(\tilde{v}) + \mu^2}{Var(v) + \mu^2} \left[\frac{\int v^{\tilde{H}_{12}} d\tilde{G}(v)}{\int v^{H_{12}} dG(v)} \right]^{-1} \frac{\int v^{2\tilde{H}_{12}} d\tilde{G}(v)}{\int v^{2H_{12}} dG(v)}$$

which can only hold with equality if $\tilde{G} = G$ and $\tilde{\rho} = \rho$.

This implies that $z_0(t)$ is identified; since we have identified ρ and thus $H_{1,\Sigma}$, it follows that α, the coefficients of the baseline hazard function, are identified.

A.3 Proof of Theorem 2

Theorem 2 is proved via a series of lemmas. The asymptotic distributional theory for our estimator based on the inverted linear rank test statistic is considerably facilitated by considering the counting process associated with the transformation model: the Doob-Meyer decomposition relates the innovation to the process to a martingale difference, and the asymptotic behaviour of partial sums of martingales are well understood (Rebolledo's martingale central limit theorem, see Andersen and Gill (1982)). To this end, we consider first the intensity of the counting process, given by the hazard, before turning to the Doob-Meyer decomposition itself. For ease of exposition, and wlog, we set the weighting function to unity. In the empirical application, we use Gehan weights.

For notational convenience define $\bar{\theta} \equiv (\theta', \rho)'$ and $\bar{\theta}_0 \equiv (\theta_0', \rho_0)$. Recall the transformation model for duration T given by equation (7), $U_i = h_i(T, \bar{\theta})$, and the transformation model evaluated at the population parameter vector $\bar{\theta}_0$, denoted by $U_{i,0} \equiv h_{i,0}(T)$ with $h_{i,0}(T) \equiv h_i(T, \bar{\theta}_0)$. We associate with the transformed durations U_i and $U_{i,0}$ the hazards $\kappa_i(u, \bar{\theta})$ and $\kappa_{i,0}(u) \equiv \kappa_i(u, \bar{\theta}_0)$ and the CDFs $F_{U,i}$ and $F_{U_{i,0}}$. U_i and $U_{i,0}$ are related by the mapping

$$U_i = h_i\big(h_{i,0}^{-1}(U_{i,0}), \bar{\theta}\big) \tag{A8}$$

where h_i^{-1} denotes the inverse of $h_i(T, \bar{\theta})$ with respect to its first argument. Let also $h_i'(.)$ denote the first derivative with respect to the first argument. The following lemma relates the hazard of U to that of U_0.

Lemma 2

$$\begin{aligned} F_{U_i}(u) &= F_{U_{i,0}}\left(h_{i,0}\left(h_i^{-1}\left(u, \bar{\theta}\right)\right)\right), \\ \kappa_{U_i}(u, \bar{\theta}) &= \kappa_{i,0}\Big(h_{i,0}\big(h_i^{-1}(u, \bar{\theta})\big)\Big) \frac{h_{i,0}'(h_i^{-1}(u, \bar{\theta}))}{h_i'(h_i^{-1}(u, \bar{\theta}), \bar{\theta})} \end{aligned} \tag{A9}$$

Proof. We have

$$F_{U_{i,0}}\left(h_{i,0}\left(h_i^{-1}(u)\right)\right) = \Pr\{h_{i,0}(T) \le h_{i,0}\left(h_i^{-1}\left(u, \bar{\theta}\right)\right) = \Pr\{T \le h_i^{-1}\left(u, \bar{\theta}\right)\} = F_{U_i}(u).$$

The second claim follows by direct computation. ■

Simplifying (A9) using (7) yields

$$\begin{aligned} \kappa_{U_i}(u, \bar{\theta}) = \quad & \exp\left(e_i' H_0(h_i^{-1}(u, \bar{\theta})) X^*(h_i^{-1}(u, \bar{\theta}))\theta_0 - e_i' H(h_i^{-1}(u, \bar{\theta}); \rho) X^*(h_i^{-1}(u, \bar{\theta}))\theta\right) \\ \cdot \quad & \kappa_{i,0}(h_{i,0}(h_i^{-1}(u, \bar{\theta}))). \end{aligned} \tag{A10}$$

For the population parameters, this simplifies to $\kappa_{U_i}(u, \bar{\theta}_0) = \kappa_{i,0}(u)$.

We note that $\kappa_{i,0}(u)$ is neither a function of the parameters θ_0 nor of the distribution of the covariates, nor of the distribution G of unobserved heterogeneity. In particular, we have (letting $H_{i.}$ denote the ith row of H)

$$\begin{aligned}
\kappa_{i,0}(u) &= E_{\underline{v}}\{\exp(H_{i.}(\rho_0)\log \underline{v})|T_i \geq h_0^{-1}(u)\} \\
&= \int_{\underline{v}} \exp(H_{i.}(\rho_0)\log v)\exp(-u \times \exp(H_{i.}(\rho_0)\log \underline{v}))dG_{\underline{v}}(v) \\
&\times \left[\int_{\underline{v}} \exp(-u \times \exp(H_{i.}(\rho_0)\log \underline{v}))dG_{\underline{v}}(v)\right]^{-1} \\
&= \int_{v_1}\cdots\int_{v_n}\prod_j v_j^{H_{ij}(\rho_0)}\exp(-u \times \prod_j v_j^{H_{ij}(\rho_0)})dG(v_n)\cdots dG(v_1) \\
&\times \left[\int_{v_1}\cdots\int_{v_n}\exp(-u \times \prod_j v_j^{H_{ij}(\rho_0)})dG(v_n)\cdots dG(v_1)\right]^{-1} \qquad \text{(A11)}
\end{aligned}$$

This follows from noting that

$$\Pr\{\underline{v} \leq v|T_i \geq h_0^{-1}(u)\} = \frac{\Pr\{T_i \geq h_0^{-1}(u)|\underline{v} \leq v\}\Pr\{\underline{v} \leq v\}}{\Pr\{T_i \geq h_0^{-1}(u)\}}$$

and

$$1 - F_{T_i}(h_0^{-1}(u)|x, \underline{v}) = \exp(-u \times \exp(H_{i.}(\rho_0)\log \underline{v})).$$

If spatial interactions are absent, $\rho_0 = 0$, $H_0 = I$ and $\kappa_{i,0}(u)$ greatly simplifies to $\kappa_{i,0}(u) = \int v dG(v|T \geq h_0^{-1}(u)) = -\mathcal{L}_v'(u)/\mathcal{L}_v(u)$ where $\mathcal{L}_v(u)$ denotes the Laplace transformation of v.

Our study of the estimating function is based on an asymptotically equivalent representation, which involves a first order expansion of κ_U. In the neighbourhood of $\bar{\theta}_0$, $\kappa_U(u, \bar{\theta})$ is asymptotically linear in $\bar{\theta}$:

Lemma 3 *Under the stated assumptions*

$$\left|\kappa_U(u, \bar{\theta}) - \kappa_0(u) - \frac{\partial \kappa_U}{\partial \theta'}(u, \bar{\theta}_0)(\bar{\theta} - \bar{\theta}_0)\right| \leq ||\bar{\theta} - \bar{\theta}_0||^2\eta(u) \qquad \text{(A12)}$$

where $\eta(u)$ is a vector of integrable functions.

Proof. The assumptions that $0 < \left|\partial^2\lambda(t, \alpha)/\partial\alpha\partial\alpha'\right| < \infty$ for all $t \geq 0$ and α in the parameter space, that $x(t)$ is bounded, imply that the second derivatives of $\kappa_U(u, \bar{\theta})$ with respect to $\bar{\theta}$ are bounded for all $u \leq \tau$ and $\bar{\theta} \in (\Theta \times \Theta_W)$. It is then sufficient that the parameter space be convex. ■

The derivatives of $\kappa_U(u; \theta, \rho)$ with respect to θ and ρ evaluated at $\theta = \theta_0$ are given in the following lemma where $g_u\left(\bar{\theta}\right) = h_i^{-1}(u, \bar{\theta})$ for ease of notation:

Lemma 4

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u,\bar{\theta})}{\partial \theta} &= [\kappa_{i,0}(h_{i,0}(h_i^{-1}(u,\bar{\theta}))] \\
&\quad \times \exp\left(H_{0,i.}(g_u\left(\bar{\theta}\right))X^*(g_u\left(\bar{\theta}\right))\theta_0 - H_{i.}(g_u\left(\bar{\theta}\right);\rho)X^*(g_u\left(\bar{\theta}\right))\theta\right) \\
&\quad \times(-1)H_{i.}(g_u\left(\bar{\theta}\right);\rho)X^*(g_u\left(\bar{\theta}\right)) \\
&\quad + \exp\left(H_{0,i.}(g_u\left(\bar{\theta}\right))X^*(g_u\left(\bar{\theta}\right))\theta_0 - H_{i.}(g_u\left(\bar{\theta}\right);\rho)X^*(g_u\left(\bar{\theta}\right))\theta\right) \\
&\quad \times \kappa'_{i,0}(h_{i,0}(g_u\left(\bar{\theta}\right))) \\
&\quad \times \exp\bigl(H_{0,i.}X^*(g_u\left(\bar{\theta}\right))\theta_0\bigr) \\
&\quad \times \exp\bigl(-H_{i.}\left(\rho\right)X^*(g_u\left(\bar{\theta}\right))\theta\bigr) \\
&\quad (-1)\int_0^T \exp\bigl(H_{i.}\left(\rho\right)X^*(s)\theta\bigr)H_{i.}\left(\rho\right)X^*(s)ds \qquad \text{(A13)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u,\bar{\theta})}{\partial \rho} &= -\exp\left(H_{0,i.}(g_u\left(\bar{\theta}\right))X^*(g_u\left(\bar{\theta}\right))\theta_0 - H_{i.}(g_u\left(\bar{\theta}\right);\rho)X^*(g_u\left(\bar{\theta}\right))\theta\right) \\
&\quad \times H_{i.}WHX^*(g_u\left(\bar{\theta}\right)\theta \\
&\quad \times \kappa_{i,0}(h_{i,0}(g_u\left(\bar{\theta}\right))) \\
&\quad - \exp\left(H_{0,i.}(g_u\left(\bar{\theta}\right))X^*(g_u\left(\bar{\theta}\right))\theta_0 - H(g_u\left(\bar{\theta}\right);\rho)X^*(g_u\left(\bar{\theta}\right))\theta\right) \\
&\quad \times \kappa'_{i,0}(h_{i,0}(g_u\left(\bar{\theta}\right))) \\
&\quad \times \exp\bigl(H_{0,i.}X^*(g_u\left(\bar{\theta}\right))\theta_0\bigr) \\
&\quad \times \exp\bigl(-H_{i.}\left(\rho\right)X^*(g_u\left(\bar{\theta}\right))\theta\bigr) \\
&\quad \times \int_0^{g_u\left(\bar{\theta}\right)} \exp\bigl(H_{i.}\left(\rho\right)X^*(s)\theta\bigr)HWHX^*(s)\theta ds \qquad \text{(A14)}
\end{aligned}$$

Proof. The results follow from tedious yet standard computations after noting that $h_i(h_i^{-1}(u;\theta,\rho);\theta,\rho) = u$ implies $(\partial/\partial\theta)h_i^{-1}(u;\theta,\rho) = -(\partial h_i/\partial\theta)/(\partial h_i/\partial s)$ with $s = h_i^{-1}(u;\theta,\rho)$. ■

Evaluated at $\rho = \rho_0$, and using the change of variables $h_{i,0}^{-1}(u) = s$, (A13) and (A14) simplify to

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u,\theta_0,\rho_0)}{\partial \theta} &= -\kappa_{i,0}(u)\bigl(e_i'H_0(h_{i,0}^{-1}(u))X^*(h_{i,0}^{-1}(u)\bigr) \\
&- \kappa'_{i,0}(u))\int_0^u e_i'H_0(h_{i,0}^{-1}(s))X^*(h_{i,0}^{-1}(s))ds \qquad \text{(A15)}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u,\theta_0,\rho_0)}{\partial \rho} &= -e_i'H_0WH_0X^*(h_{i,0}^{-1}(s))\theta_0 \times \kappa_{i,0}(u) \qquad \text{(A16)} \\
&- \kappa'_{i,0}(u)\int_0^u e_i'H_0(h_{i,0}^{-1}(s))W(h_{i,0}^{-1}(s))H_0(h_{i,0}^{-1}(s))X^*(h_{i,0}^{-1}(s))\theta_0 ds.
\end{aligned}$$

Next, we turn to the associated counting processes. For the duration variate T denote by $\{N(t)|t \geq 0\}$ the stochastic process describing the number of exits from the state of interest in the interval $[0, t]$ as time proceeds. Of course, there is at most one exit. For the transformed duration U, we have

$$N^{U_i}(u, \bar{\theta}) = N\big(h_i^{-1}(u, \bar{\theta})\big),$$

and for the population parameters we have $N^{U_{i,0}}(u) \equiv N^{U_i}(u, \bar{\theta}_0))$. It remains to account for censoring of the duration variate. Let $y(t) = I(t \leq T)I(t \leq C)$ denote the observation indicator, where C denotes a non-informative right censoring time. Let $\bar{Y}(t) = [\bar{y}_i(t)]_{i=1,..,n}$ denote the history of the observation indicators. We then have

Lemma 5

$$\Pr\big(dN^{U,i}(u, \bar{\theta}) = 1 \big| \overline{X}^{U_i}(u, \bar{\theta}), \overline{Y}^{U_i}(u, \bar{\theta}), \overline{W}^{U_i}(u, \bar{\theta})\big) = y^U(u, \bar{\theta})\kappa_{U_i}\big(u; \bar{\theta})\big)du \quad \text{(A17)}$$

with κ_{U_i} given by (A9). The associated Doob-Meyer decomposition is

$$dN^{U_i}(u, \bar{\theta}) = y_i^U(u, \bar{\theta})\kappa_{U_i}\big(u; \bar{\theta}\big)\, du + dM^{U_i}(u, \bar{\theta}) \quad \text{(A18)}$$

where M^{U_i} denotes a martingale.

For the population parameters we define $M^{U_{i,0}}(u) \equiv M^{U_i}(u, \bar{\theta}_0)$ and $N^{U_{i,0}}(u) \equiv N^{U_i}(u, \bar{\theta}_0)$. Using this representation (A18), the estimation function can be written as

$$\begin{aligned} S_n(\bar{\theta}) &= \sum_{i=1}^{n} \Delta_i \Big[x_i\big(U_{(i)}\big) - \tilde{x}\big(U_{(i)}\big)\Big] \\ &= \sum_{i=1}^{n} \int_0^{\tau} \Big(x(u) - \tilde{x}(u)\Big) dN^{U_i}(u, \bar{\theta}). \end{aligned} \quad \text{(A19)}$$

The transformed durations are observed up to time $\tau < \infty$. Evaluating the estimation function (A19) at the population parameters, we have

$$\begin{aligned} S_n(\bar{\theta}_0) &= \sum_{i=1}^{n} \int_0^{\tau} \Big(x(u) - \tilde{x}(u)\Big) dM^{U_{i,0}}(u) \\ &+ \sum_{i=1}^{n} \int_0^{\tau} \Big(x(u) - \tilde{x}(u)\Big) y_i^{U_0}(u)\kappa_{U_i}\big(u; \bar{\theta}_0\big)\mathrm{d}u \end{aligned} \quad \text{(A20)}$$

Lemma 6 *The counting measure $N^{U_{i,0}}(u)$ does not depend on the covariate and the spatial processes, hence $E(S_n(\bar{\theta}_0)) = 0$.*

Proof. By definition, we have $Pr\{dN(t) = 1\} = y_i(t)\tilde{\lambda}_i(t|\bar{\theta}_0)dt$. $\tilde{\lambda}_i$ is the expectation of λ_i with respect to $\underline{v}$, which equals, using (7), $\exp\big(e_i' H(t; \rho_0) X^*(t)\theta_0\big) E_{\underline{v}}(\psi_i | T \geq t)$. This probability equals $Pr\{dN^{U_i}(u, \bar{\theta}) = 1\}$ with $du = h'(t, \bar{\theta})dt$, so the intensity of the transformed counting process can be written as

$$\exp\big(e_i'[H(u; \rho_0)X^*(u)\theta_0 - H(u; \rho)X^*(u)\theta]\big) E_{\underline{v}}(\psi_i | U \geq u)$$

Hence, evaluated at the population parameters, we have

$$Pr\{dN^{U_i}(u,\bar{\theta}_0)=1\} = y_i^{U_{i,0}} E_{\underline{v}}(\psi_i | U \geq u)$$

which does not depend on the X^* and H. ■

Since $E(S_N(\bar{\theta}_0)) = 0$, it follows that the second term in (A20) is zero, so

$$S_n(\bar{\theta}_0) \;=\; \sum_{i=1}^{n} \int_0^{\tau} \Big(x(u) - \tilde{x}(u)\Big) dM^{U_{i,0}}(u). \tag{A21}$$

Using again the representation (A18) for $S_n(\bar{\theta})$, we obtain the following linearisation

$$\tilde{S}_n(\bar{\theta}) \;=\; S_n(\bar{\theta}_0) + G(\bar{\theta}_0) \times (\bar{\theta} - \bar{\theta}_0) \tag{A22}$$

with

$$\begin{aligned} G(\theta_0, \rho_0) \;&\equiv\; (\partial S/\partial \theta, \partial S/\partial \rho)|_{\bar{\theta}=\bar{\theta}_0} \\ &=\; \sum_{i=1}^{n} \int_0^{\tau} \Big(x(u) - \tilde{x}(u)\Big) \frac{\frac{\partial}{\partial \bar{\theta}} \kappa_{U_i}(u,\bar{\theta})|_{\bar{\theta}=\bar{\theta}_0}}{\kappa_{U_i}(u,\theta_0)} \mathrm{d}N^{U_i}(u,\bar{\theta}) \end{aligned} \tag{A23}$$

where $(\partial/\partial\bar{\theta})\kappa_{U_i}(u,\bar{\theta})$ is given in Lemma 4 above. The argument in Tsiatis (1990) demonstrates that $\tilde{S}_n(\bar{\theta})$ is asymptotically equivalent to $S_n(\bar{\theta})$ in the neighbourhood of $\bar{\theta}_0$, and this asymptotic equivalence then implies that the estimator is consistent:

Lemma 7 *Under the stated assumptions for all $c > 0$*

$$\sup_{|\bar{\theta}-\bar{\theta}_0| \leq cn^{-\frac{1}{2}}} n^{-\frac{1}{2}} \Big| S_n(\bar{\theta}) - \tilde{S}_n(\bar{\theta}) \Big| \xrightarrow{p} 0 \tag{A24}$$

Finally, we observe that our estimator is the root of $\tilde{S}_n(\bar{\theta})$. Hence solving (A22) for $(\bar{\theta} - \bar{\theta}_0)$, and invoking the asymptotic normality of $S_n(\bar{\theta}_0)$ implied by (A21), yields the result stated in Theorem 2: the estimator is asymptotically normally distributed.

B Data Appendix

B.1 Administrative panel data on the population of recent immigrants to The Netherlands

All legal immigration by non-Dutch citizens to the Netherlands is registered in the Central Register Foreigners (Centraal Register Vreemdelingen, CRV), using information from the Immigration Police (Vreemdelingen Politie) and the Immigration and Naturalisation Service (Immigratie en Naturalisatie Dienst, IND). It is mandatory for every immigrant to notify the local population register immediately after the arrival in the Netherlands if he intends to stay for at least two thirds of the forthcoming six months. Natives as well as immigrants are required to register with their municipality. Our data comprise the entire *population* of immigrants who entered during our observation window of 1999-2007.

The immigration register is linked by Statistics Netherlands to the Municipal Register of Population (Gemeentelijke Basisadministratie, GBA) and to their Social Statistical Database (SSD). The GBA contains basic demographic characteristics of the migrants, such as age, gender, marital status and country of origin. From the SSD we have information (on a monthly basis) on the labour market position, income, industry sector, housing and household situation. Since we consider only new entrants to the Netherlands, most immigrants are not eligible for social benefits such as unemployment insurance payments, since these are conditional on sufficiently long employment or residence durations. Migration and employment durations of specific lengths (e.g. 3 or 5 years) trigger statutory changes in employment and residence rights. However, our earlier work in Bijwaard, Schluter and Wahba (2014) has verified that these do not affect average migration hazards.

In addition to the date of entry and exit, the administration also records the migration motive of the individual. The motive is coded according to the visa status of the immigrant; if not, the immigrant reports the motive upon registration in the population register. Statistics Netherlands distinguishes between the several motives: labour migrants, family migrants (this category include both family unification as well as immigration of foreign born spouses, i.e. family formation), student immigrants, asylum seekers (and refugees), and immigrants for other reasons. Bijwaard (2010) shows that these different immigrant groups differ systematically in terms of return behaviour, labour market attachment, and demographic characteristics. We therefore consider only *labour migrants*, being the group which economists usually are interested in the most. Labour migrants represent about 26% of all non-Dutch immigrants in the age group 18-64. It is possible that the labour migration motive is either miscoded or misreported. Since all Turkish labour migrants require an employment-dependent work visa to immigrate, they should be formally employed not too long after entry. Thus, in order to limit the possibilities of misclassification error of the labour migration motive, we require that immigrants be employed in the Netherlands within three months of their entry.

This selection by immigration motive yields an administrative population of recent labour immigrants of 94,270 individuals. This size of our population data permits us to consider specific groups. Such stratification also controls for important differences in language ability, as these could influence assimilation and are thus important for the studies focussing on return. For instance, the (mean) language deficits of Turks and Moroccans in the Netherlands are well known. As in Adda et al. (2015) in the context of Germany, we consider in the main text the largest ethnic group of labour immigrants, namely Turks (about 8000 individuals). In the next subsection, we document their spatial clustering as well as their spatial segregation from other principal immigrant groups.

B.2 The spatial dimension: Neighbourhoods

The neighbourhood is often argued to be the spatial unit at which local social interactions take place. A further special feature of our data is that we know the neighbourhood the immigrant lives in, defined by Statistics Netherlands as an area of approximately 2,000 households. The Netherlands is thus partitioned into about 14,000 neighbourhoods.

In order to document the spatial clustering and segregation along ethnic lines among the principal immigration groups, and Turks in particular, we use publicly available population data produced by Statistics Netherlands for *all* immigrants (recent and established, labour and non-labour immigrants) for the year 2007. The size of this data permits a reliable description of the spatial settlement patterns. In order to establish some benchmarks, we contrast Turkish immigrants with immigrants from the next three largest groups, i.e. immigrants from Moroccans, and immigrants from the former Caribbean colonies of Surinam and the Dutch Antilles. The four groups represent about 11% of the total population of the Netherlands.

We start by documenting the spatial clustering in the four largest cities, then consider the distinct neighbourhoods in these four largest cities (the number of neighbourhoods by city are 92 in Amsterdam, 78 in Rotterdam, 107 in the Hague and 96 in Utrecht). As about only 12.8% of the total population of the Netherlands resides in these four cities, we then turn to all 14,000 neighbourhoods.

B.2.1 Concentration, isolation, and dissimilarity in the four largest cities

Table B.1: Ethnic concentration in the four largest cities

	Tur	Mor	Sur	Ant	Tur	Mor	Sur	Ant
	By City				*By Ethnic Group*			
Amsterdam	5.2	9.0	9.2	1.6	9.3	18.5	22.3	10.0
Rotterdam	7.8	6.4	8.9	3.3	11.1	10.3	16.4	16.6
The Hague	7.0	5.3	9.5	2.3	7.6	6.8	14.0	8.9
Utrecht	4.4	8.8	2.6	0.9	3.0	6.8	2.2	2.0

Notes. Immigrant groups: Tur(ks), Mor(occans), Sur(inamese), Ant(illians). Panel A: Indices are by cities, so Turkish concentration in Amsterdam is the share of the Amsterdam population that is Turkish. Panel B: The Turkish concentration in Amsterdam is the share of the Turkish population that lives in Amsterdam. Data for 2007.

The extent of clustering of immigrants along ethnic lines is illustrated in Table B.1 in the year 2007 by city. The four largest cities are home to a large share of the immigrant population. For instance, 9% of the Amsterdam population are Moroccans, 5% are Turks, and nearly 25% of the population of Amsterdam is from the four principal immigrant groups. Panel B of Table B.1 considers the four largest cities in terms of the total immigrant populations. The proportion of Turks living in the these four cities equals 31%, the Antillians' share is 38%, the Moroccan share is 42%, and the Suriname share is 55%. Hence the spatial analysis needs to extend beyond these four principal cities.

Table B.2: The four largest cities

	Tur	Mor	Sur	Ant	Tur	Mor	Sur	Ant
	Dissimilarity				*Isolation*			
Amsterdam	.446	.429	.344	.308	.106	.172	.181	.031
Rotterdam	.417	.386	.217	.276	.160	.110	.109	.048
The Hague	.523	.504	.348	.291	.183	.133	.147	.036
Utrecht	.441	.482	.248	.202	.096	.219	.034	.012

Notes. Immigrant groups: Tur(ks), Mor(occans), Sur(inamese), Ant(illians). Indices are by cities, so Turkish concentration in Amsterdam is the share of the Amsterdam population that is Turkish. Data for 2007.

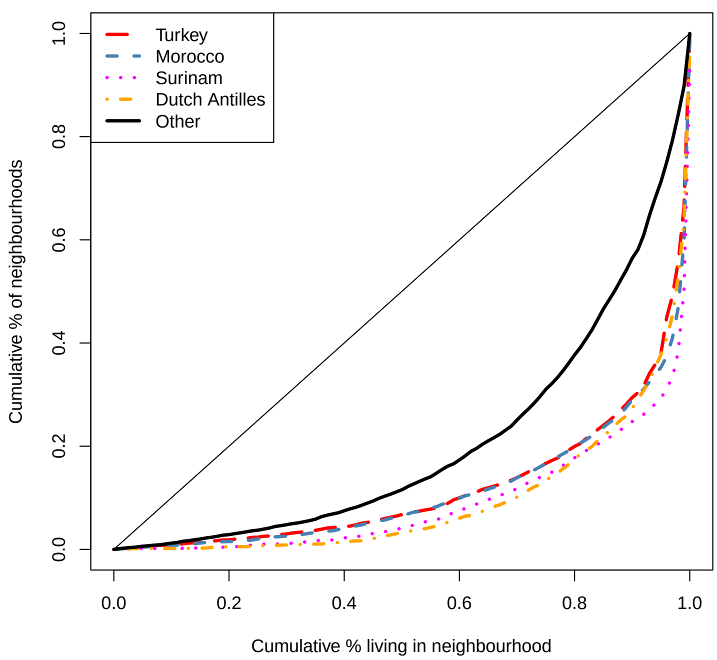
Next, we consider these four cities at the level of the neighbourhood and investigate the extent to which immigrants of a particular ethnic group, such as Turks, (we label them "minority", min_n in neighbourhood n) differ from natives and other immigrants in this neighbourhood (label this complement to the minority the "majority", maj_n). Summing over all neighbourhoods in a city yields the subpopulation totals min_{total} and maj_{total}. Two standard descriptors are the following indices of dissimilarity and isolation (see e.g. Cutler et al. (1999)). The dissimilarity for neighbourhood n is often measured by comparing same-group population shares, and summing over neighbourhoods yields the dissimilarity index $0.5 \sum_n |min_n/min_{total} - maj_n/maj_{total}|$. This dissimilarity index is a measure of imbalance and quantifies the extent to which group g immigrants are unevenly distributed across neighbourhoods. The magnitudes of the estimates reported in Table B.2 confirm that the four principal immigrant groups are unequally distributed across the cities' neighbourhoods. For each immigrant group, the dissimilarity index is similar across the four cities. Comparing the immigrant groups, dissimilarity for Turks and Moroccans is substantial larger than for Surinamese and Antilleans.

We also consider the measure of isolation or exposure given by $\sum_n(min_n/min_{total} \times min_n/(min_n+maj_n))$ which weights the own-group population share of the neighbourhood (or concentration) by the its population share in that neighbourhood. Except for Antilleans, Table B.2 suggests that isolation is fairly high by European standards. Moreover, Turks are the most isolated in The Hague, Moroccans in Utrecht, Surinamese in Amsterdam, and Antilleans in Rotterdam. Overall, we conclude that the extent of clustering and segregation among the four principal immigrant groups in the four largest cities is substantial.

B.3 All neighbourhoods: Clustering and segregation

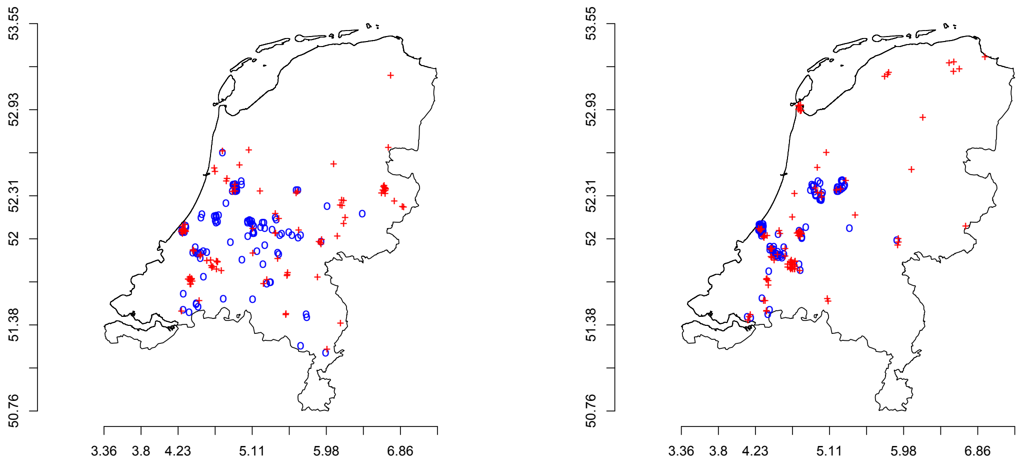
Turning to all neighbourhoods, Figure B.1 depicts the Lorenz curve for spatial concentration. It is evident that most migrants live in a relatively small number of neighbourhood, and this extent of clustering is much larger than for all other immigrants. The Lorenz curve reveals the extent of spatial concentration, but cannot reveal the geographic distribution. This is done in Figure B.2 where we map, for different ethnic groups, the 100 most concentrated neighbourhoods. The map shows the extent of segregation as there is little overlap across ethnic lines between the neighbourhoods.

Figure B.1: Lorenz curves of spatial concentration



The links between these neighbourhoods, and thus the scope for social interactions, can be examined using standard tools from social network analysis. To this end, consider the adjacency matrix of the 100 most concentrated neighbourhoods for ethnic group g, $W_g = [w_{g,ij}]_{i,j=1,..,100}$, where the binary $w_{g,ij}$ equals zero unless neighbourhoods i and j are within, say, 5km distance of each other and one otherwise. To examine the connectedness or centrality of a neighbourhood, Bonacich (1987) has proposed the measure $B(\beta) = (I_{100} - \beta W_g)^{-1} W_g 1_{100} = \sum_{k=0}^{\infty} \beta^k W_g^{k+1} 1_{100}$ where I_{100} is the identity matrix and 1_{100} is a vector of ones. $B(\beta) = [b(\beta)_i]_{i=1,..,100}$ equals the weighted sum of direct and indirect links between neighbourhoods. Setting the subjective weight $\beta = 1/33$ (to satisfy the parameter's eigenvalue constraint across all ethnic groups, see discussion of equation (2) above), the Bonacich measure reveals for Moroccans some neighbourhoods in Amsterdam to be most central ($\max(B) = 2.05$),

Figure B.2: The 100 most concentrated neighbourhoods by ethnic group.



Notes: Panel A: "o" depicts Turkish, and "+" depicts Moroccan neighbourhoods; Panel B: "o" depicts Surinamese, and "+" depicts Antillian neighbourhoods

whereas for Turks the most central neighbourhoods are in Schilderswijk (inside The Hague, $\max(B) = 1.66$). For Surinamese the maximum is attained in different neighbourhoods of the Hague ($\max(B) = 1.71$), and for Antillians the most central neighbourhoods are in Rotterdam ($\max(B) = 1.94$).