

Bijwaard, Govert; Schluter, Christian

Working Paper

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

CRAM Discussion Paper Series, No. 20/16

Provided in Cooperation with:

Rockwool Foundation Berlin (RF Berlin)

Suggested Citation: Bijwaard, Govert; Schluter, Christian (2016) : Interdependent Hazards, Local Interactions, and the Return Decision of Recent Migrants, CRAM Discussion Paper Series, No. 20/16, Centre for Research & Analysis of Migration (CRAM), Department of Economics, University College London, London

This Version is available at:

<https://hdl.handle.net/10419/295547>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



CREAM
Centre for Research &
Analysis of Migration

Discussion Paper Series

CPD 20/16

- ▶ **Interdependent Hazards, Local Interactions,
and the Return Decision of Recent Migrants**
- ▶ Govert E. Bijwaard and Christian Schluter

Centre for Research and Analysis of Migration
Department of Economics, University College London
Drayton House, 30 Gordon Street, London WC1H 0AX

www.cream-migration.org

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

Govert E. Bijwaard *

Netherlands Interdisciplinary Demographic Institute (NIDI)

CReAM and IZA Bonn

Christian Schluter[†]

Aix-Marseille Université (Aix Marseille School of Economics), CNRS & EHESS

University of Southampton and CReAM

August 2016

Abstract

Consider the duration of stay of migrants in a host country. We propose a statistical model of locally interdependent hazards in order to examine whether interactions at the level of the neighbourhood are present and lead to social multipliers. To this end, we propose and study a new two-stage estimation strategy based on an inverted linear rank test statistic. 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 can be substantial.

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

JEL Codes: C41, C10, C31, J61

*bijwaard@nidi.nl. Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW/ University of Groningen), PO Box 11650, 2502 AR The Hague, the Netherlands, +31 70 3565224

[†]christian.schluter@univ-amu.fr GREQAM, Centre de la Vieille Charité, 13002 Marseille, France, and Department of Economics, University of Southampton, Highfield, Southampton, SO17 1BJ, UK.

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, study identification, and propose a method for estimating a reduced-form statistical model of interdependent hazards which arise from such local interactions. In particular, our substantive empirical application concerns the duration of recent labour immigration to the Netherlands. Spatially, immigrants not only cluster but also segregate along ethnic lines. Hence, using Dutch administrative data on individual immigration durations, we test whether an individual’s return hazard is impacted on by the propensity of her peers or co-ethnics in the neighbourhood to return.

This new model of interdependent hazards contributes to the recent literature of modelling of social interactions in duration analysis. However, it differs from the model of interdependent durations of Honoré and de Paula (2010), who consider a complete information 2-agent synchronisation game of optimal switching. We believe that, in the specific context of our empirical setting - local social interactions at the level of the neighbourhood - information is necessarily incomplete while the number of peers is much larger and ties are weaker than in their setting. For the same reason, our model also differs from the hazard model with social interaction of Drepper and Effraimidis (2015). They consider the hazard of first-time drug use in a very small group (siblings), and seek to estimate the impact of the exogenous first transition within the group on other group members. 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 are often parametrised using copulas (see e.g. Goethals et al. (2008) for a discussion). In our model, hazards are directly modelled as interdependent, and this interdependence arises from local interactions. In terms of the spatial dimension of the analysis, we also contribute methodologically and empirically to a growing literature that confirms the importance of the effects of local interactions and of the neighbourhood.¹

Specifically, our substantive empirical application concerns the staying durations of immigrants in a host country, a key question in the migration literature (see e.g. the survey by Dustmann and Görlach (2016)). The spatial clustering² of immigrants

¹For 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.

²See 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,

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. 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 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. Using the same data, Bijwaard, Schluter and Wahba (2014) have established the considerable incidence of return migration, and have examined the individual-specific drivers of the return decision. The size of our population data of recent labour migrants permits us to consider specific groups. As in Adda et 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).

The spatial dimension of our data allows us now to examine the extent to which local social interactions affect the duration of stay. In particular, the 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 immigrant groups (Turks and three others for comparison) is extensively

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.

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 evidence suggests that there is scope for local interactions resulting in interdependent return hazards.

To this end and within this empirical context, we propose a model of interdependent hazards that arise from local interactions. First, we study identification. Identification and estimation strategies for the linear model³ are not available to us since duration models are inherently non-linear. Moreover, the objects of interest, migration duration hazards, are not directly observable, nor can we use classic identification results for mixed-proportional hazard models (MPH), since the reduced-form of our model of locally interdependent hazards is not of the MPH form. Despite these challenges, we demonstrate non-parametric identification.

Our estimation strategy is based on Tsiatis' (1990) inverted linear rank (LR) test (see also Bijwaard et al. (2013)), that is developed further here for the spatial setting. The LR test statistic is based on a weighted 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. Under the null of no covariate effect, the covariate does not influence the hazard, and the expected difference is thus zero. Using Gehan weights ensures the monotonicity of the estimating function. The linear rank estimator (LRE) is the root of the analogue sample moment condition. Since the object of interest, the return hazard, is not directly observable, we follow the approach adopted by Pinske and Slade (1998), who use the generalised residuals in a spatial probit model. For durations, the generalised residuals are obtained by applying a transformation model to the durations, i.e. we consider the integrated hazards. We also note that the spatial interaction matrix (W) for our data is very large (a non-sparse $89,000^2$ in our empirical application), so that estimators using an inverse of the spatial interaction matrix are computationally not feasible. In order to solve this computational problem associated with the size of the social interaction matrix, we follow the suggestions of Klier and McMillen (2008), and use a first order approximation of the moment condition in terms of the social local interaction parameter.⁴

³To be precise, consider the model $y = \alpha + x\beta + Wx\delta + \rho Wy + \epsilon$, where $y = [y_i]_{i=1,\dots,n}$ is a n -vector of outcomes, x is the $n \times k$ matrix collecting exogenous characteristics and $W = [w_{ij}]_{i=1,\dots,n;j=1,\dots,n}$ is the $n \times n$ social interaction matrix (indicating whether individual j is a peer of individual i). ρ captures endogenous neighbourhood effect, and δ the exogenous neighbourhood effect. The reflection problem (Manski, 1993) refers to the perfect multicollinearity between Wy and Wx . E.g. Bramoullé et al. (2009) consider a W that exhibits intransitive triads, so (x, Wx, Wy) can be instrumented by (x, Wx, W^2x) .

⁴These econometric problems and solutions are very different from the analysis of *exogenous* spatial effects on unemployment durations examined in e.g. Gobillon et al. (2010) who use proportional hazard models in which baseline hazards are estimated for each location. More specifically, they (i) estimate the covariate coefficients across all municipalities using stratified partial maximum likelihood (SPLE), and then recover the integrated baseline hazard for each municipality, indexed by j , using the Breslow estimator, (ii) the locality specific baseline hazard is assumed to have a multiplicative form, $\theta^j(t) = \alpha^j \theta(t)$, and the coefficients are estimated using minimum distance (MD); (iii) these coefficients α^j are then regressed on locality characteristics Z^j .

The outline of this paper is as follows. In the next section, we present the model of locally interdependent hazards, demonstrate identification, and detail the estimation strategy whose performance is then examined in the context of a simulation study. Section 3 is devoted to the empirical application, the locally interdependent hazards of leaving the host country (the Netherlands) for recent labour immigrants of the largest ethnic group (Turks). We 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. Overall we show that these social multipliers can be substantial. 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 Local social interactions in duration models: Interdependent hazards

We seek to determine the effect of local social interactions on the return migration intensity of immigrants, so the random outcome variable of interest is the time spent in the Netherlands, denoted by T . The observational units are recent labour immigrants in the host country (the Netherlands). For expositional clarity, we present first a restricted model of the migration duration which ignores the endogenous social interaction effect.

We follow common practice in duration analysis (for a survey see e.g. van den Berg (2001), who also observes that the “hazard function is the focal point of econometric duration models” (p.3387)) and express the distribution of the migration duration variate T in terms of the associated hazard, say λ . 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 time alone (and a parameter vector α) and common to all individuals, and a covariate function, $\exp(x(t)\beta)$, which accelerates or decelerates exits: $\lambda(t|\bar{x}(t); \theta) = \lambda_0(t, \alpha) \exp(x(t)\beta)$ with $\theta = (\alpha', \beta)'$ and $\bar{x}(t)$ being the history of the covariate process $x(\cdot)$ up to time 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 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: $\lambda(t|\bar{x}(t), v; \theta) = v\lambda_0(t, \alpha) \exp(x(t)\beta)$. 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 do not need to specify G further. In our empirical application the baseline hazard is modelled as piecewise constant, so

$\lambda_0(t, \alpha) = \exp(\alpha_0 + \alpha'_{-0}A(t))$ with $A(t) = (I_1(t), \dots, I_k(t))'$ denoting the vector of interval indicators.

Social interaction models distinguish between exogenous or contextual effects and endogenous effects. The contextual effects are easily incorporated in the MPH model since the covariate vector $x(t)$ can include pre-determined characteristics of the neighbourhood the individual lives in (we thus avoid increasing the notational burden). The maintained assumption is that the baseline hazard function does not vary across individuals in different neighbourhoods (recall Footnote 4 for an alternative approach).

We model the endogenous local interaction effects as follows. Consider migrant i in a particular neighbourhood. Denote the set of relevant peers (in our case migrants of a particular ethnic group) of migrant i by $N_i(t)$ with $i = 1, \dots, n$. Since our spatial units are small, this set might extend beyond the confines of the neighbourhood, and might include peers from neighbouring neighbourhoods. N_i is a function of time since the migrant might move to a different neighbourhood within our observation window. The number of relevant peers is denoted by $\#N_i(t)$. As is common in spatial and social network econometrics, we collect this information in the $n \times n$ spatial interaction matrix $W(t) = [w_{ij}(t)]_{i=1, \dots, n; j=1, \dots, n}$ 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.

As the outcome of interest is the return hazard, the endogenous local effect refers to the extent to which 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 . Since we are working within the MPH paradigm, it is natural to assume that this effect is proportional, so that it is the geometric mean

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

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

$$\lambda_i(t|\cdot) = v_i \lambda_0(t, \alpha) \exp\left(x_i(t)\beta + \rho w_i(t) \ln\left[\underline{\lambda}(t)\right]\right), \quad (1)$$

with $\underline{\lambda}(t) \equiv [\lambda_i(t|\cdot)]_{i=1, \dots, n}$ denoting the $N \times 1$ vector whose i th element is $\lambda_i(t|\cdot)$ and w_i denoting the i th row of W . ρ is the endogenous local interaction effect which we seek to estimate. The model stipulates that an individual's (return) hazard is impacted on by the *propensity* of her peers to return (alternatively, the hazard could be modelled as a function of the individual's subjective assessment of her peers' return, and (1) is obtained by imposing rational expectations).

Our model of interdependent hazards thus differs from the model of interdependent durations of Honoré and de Paula (2010), who consider a complete information 2-person game of optimal switching. We believe that, in our social local interaction context, information is necessarily incomplete while the number of peers is much larger and ties are weaker than in their setting. For the same reason, our model also differs from the hazard model with social interaction of Drepper and Effraimidis (2012). They consider the hazard of first-time drug use in a very small group (siblings), and seek to estimate the impact of the first transition within the group on other group

members. We believe this timing-of-events framework, which requires the first transition to be exogenous (the so-called no-anticipation hypothesis), to be inappropriate in our specific setting, where it is the coherent probabilistic assessment of peers' returns that affect the individual's hazard, rather than the (timing of the) realisation of one particular first-return event.

We defer an illustration of the spatial effects on the hazards to Section 2.5, where these are considered in two analytically treatable examples. There we also present simulation evidence on the extent of the spatial biases when the spatial effect is ignored.

2.1 The reduced form

Let $\underline{v} = [v_i]_{i=1,\dots,n}$, $X(t) = [x_i(t)]_{i=1,\dots,n}$, 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

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

with $H(s; \rho) = (I - \rho W(s))^{-1}$ and $X^*(s) = (\ln \lambda_0(s); X(s))$. 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 i th element of $\underline{\lambda}$ is given by

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

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

2.2 Identification

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. This dependence renders the task of separating out dynamic sorting from duration dependence more difficult. However, it 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.

2.2.1 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. First, we control for systematic observable variation in neighbourhood characteristics by including contextual effects 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 3 below). In the Netherlands, social housing represents a large stock of accommodation-for-rent in the poorer neighbourhoods. Hence, poor recent labour immigrants are unlikely to be able to choose one particular neighbourhood.

2.3 Generalised residuals and the linear rank estimator

The usual estimation practice in the standard MPH model is to use maximum likelihood based on the duration distributions conditional on v , and then to integrate out the time-invariant heterogeneity v using some specification of G . This approach is not feasible in the local interaction MPH: the joint distribution of the locally interacted v is very complicated and too demanding computationally since the $n \times n$ spatial interaction matrix W is too large to invert. Hence we pursue an alternative estimation strategy based on generalised residuals and orthogonality conditions.

We develop an instrumental variable estimator for our spatial setting with endogenous neighbourhood effects that takes as its departure the inverted linear rank test statistic of Bijwaard et al. (2013). For a recent survey of rank-based estimation methods, see Chung et al. (2013). The key idea is an intuitive orthogonality condition: If a covariate is 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. The sample analogue of this moment condition can then be used as the estimation equation. Of course, the covariate process at survival time t , $x_i(t)$, does affect the hazard $\lambda_i(t|\cdot)$, but applying a transformation model to the durations T that generates the generalised residuals turns out to yield the desired independence.

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

$$\begin{aligned} U_i = h_i(T, \theta, \rho) &= \int_0^T \exp(e_i' H(s; \rho) X^*(s) \theta) ds \\ &= \int_0^T \lambda_i(s|\cdot) ds / \psi_i ds \end{aligned} \tag{4}$$

which is the integrated hazard except for the function of the unobservable heterogeneity terms, where ψ_i is the i th 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 $\int_0^T \lambda_i(s|\cdot) ds$ 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 (20)). 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 $(T_i, \Delta_i, \bar{x}_i(T_i))$. The transformation model transforms the durations for some θ to $(U_i(\theta), \Delta_i, \bar{x}_i(U_i(\theta)))$. 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

$$S_n(\theta) = \sum_{i=1}^n \nu_i \Delta_i \left[x_i(U_{(i)}) - \tilde{x}(U_{(i)}) \right] \quad (5)$$

where

$$\tilde{x}(U_{(i)}) = \frac{\sum_{j=i}^n I(U_{(j)} \geq U_{(i)}) x_j(U_{(i)})}{\sum_{j=i}^n I(U_{(j)} \geq U_{(i)})} \quad (6)$$

is the sample mean of the covariates for survivors at the transformed survival time, and ν_i is a weighting function.

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,

$$(\hat{\theta}, \hat{\rho}) = \underset{\theta \in \Theta, \rho \in \Theta_W}{\operatorname{argmin}} S_n(\theta, \rho)' S_n(\theta, \rho), \quad (7)$$

since the sample moment condition $S_n(\theta, \rho)$ is a step function. As S_n is thus not differentiable everywhere, 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) . This device yields the following results:

Theorem 2 $(\hat{\theta}, \hat{\rho})$ is consistent, and is distributed asymptotically as a normal variate.⁵

⁵The 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.

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

2.4 An approximation for large spatial interaction matrices

The size of the spatial interaction matrix $W(t)$ renders the minimisation of the criterion function in (7) computationally infeasible. 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 (7) in the absence of spatial interactions, $\hat{\theta}_1 = \operatorname{argmin}_{\theta \in \Theta} S_n(\theta, 0)' S_n(\theta, 0)$.

The resulting linear approximation of the spatial rank-functions is

$$S(\theta_0, \rho_0) \approx S(\hat{\theta}_1, 0) + G(\hat{\theta}_1, 0) \times \begin{pmatrix} \hat{\theta}_1 - \theta \\ 0 - \rho \end{pmatrix} \quad (8)$$

where $G(\theta, \rho) = (\partial S / \partial \theta, \partial S / \partial \rho)$, which is stated explicitly in the appendix (equation (32)). 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 ,

$$\begin{pmatrix} \hat{\theta} \\ \hat{\rho} \end{pmatrix} = \begin{pmatrix} \hat{\theta}_1 \\ 0 \end{pmatrix} + (G'G)^{-1} G' S(\hat{\theta}_1, 0) \quad (9)$$

To summarise, we propose a two-stage estimation strategy: In the first stage, obtain $\hat{\theta}_1$ from the minimisation of (7) 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 (9).

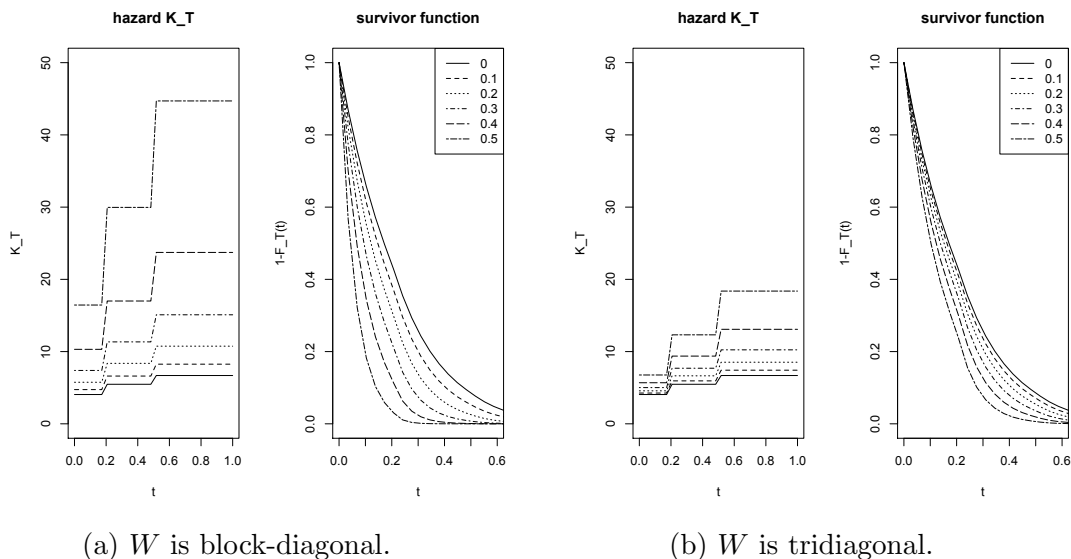
2.5 Simulation evidence

We conduct a simulation study in order to (i) illustrate the effect of social interactions on return hazards, (ii) reveal the resulting estimation biases when the social interaction effect is ignored (which also permits us to investigate the performance of the linear rank estimator), and (iii) investigate the performance of the one-step procedure (9) based on the approximation (8).

The simulation design is as follows. We consider one time-invariant covariate x that is standard uniformly distributed, with population coefficient $\beta_0 = 2$. The baseline hazard has three-pieces, defined over the intervals $[0, .2)$, $[.2, .5)$ and $[.5, \infty)$, with associated coefficients $\alpha_0 = -.6$, and $\alpha_{-0} = (.3, .5)'$. Unobserved heterogeneity v_i follows a two-point mass distribution, with mass points e^{-1} and e^1 , and selection probability $\Pr\{v_i = e^{-1}\} = 1/2$.

As regards the spatial structure, we consider two scenarios in which $H = (I - \rho W)^{-1}$ can be analytically computed. In the first structure, W is block-diagonal, and each block consists of two individuals. This corresponds to a situation in which individuals from different blocks do not interact. In the second setting, W is tridiagonal, so that individuals interact across blocks (and intransitive triads are thus present).

Figure 1: Spatial impacts on the hazard κ_T and the survivor function $1 - F_T(t)$.



Notes. ρ varies from .1 to .5. W is 4×4 in this illustration, and we consider individual 3. For these figures $x_i = .5$ for each i , and $v_1 = v_2 = e$ and $v_3 = v_4 = e^{-1}$.

2.5.1 Illustrations: Spatial impacts on hazards and survival probabilities

We begin by illustrating the spatial impacts in an analytically tractable setting of 4 individuals, so W is 4×4 . We focus on individual 3, and consider the hazard $\kappa_T(t) \equiv [\lambda_{i=3}(t|\cdot)]$ as given by (3), as well as the survivor function $1 - F_T(t)$ for the two spatial structures as ρ increases from 0 to .5. Figure 1 depicts the results.

It is evident that local social interactions raise the hazards and thus decrease survival probabilities. The spatial impacts are increasing in ρ , and the resulting effects can be substantial. For instance, for the block diagonal structure, at $t = .2$ the return hazard has increased relative to its value for $\rho = 0$ by a factor of 5.5 when $\rho = .5$. In terms of the survival probability, we observe an associated fall from 0.42 to .003. The spatial effects depend, of course, on the structure of the spatial interactions. Relative to the block diagonal structure, if W is tridiagonal, the spatial impacts in this example are smaller. At $t = .2$, $\kappa_T(.2; 0)/\kappa_T(.2; .5) = 2.25$, and the associated survival probability falls from .42 to .238.⁶

2.5.2 Simulation results: Parameter estimates

We turn to the simulation results, having drawn samples of size $N = 1,000$, repeated the experiment 200 times, and consider different values of ρ .

We start with the results for the first-stage estimation, reported in Table 1 in order to (i) quantify the extent of the spatial biases, and (ii) assess the performance

⁶In Appendix C, we also illustrate for completeness the spatial effects in terms of the hazard of U_0 , κ_{U_0} , in the transformation model (4). There we also illustrate the quality of the first order approximation of the hazard of the transformed variate U_0 .

of the linear rank estimator in the first step of the estimation. The two particular spatial structures (block and tri-diagonality) permit the explicit computation of the spatially biases that arise when the spatial interactions are wrongly ignored. These theoretical values, denoted by θ^* , are calculated as follows. The reduced form of the hazard (2) can be re-written as

$$\log(\lambda(t|\theta, \rho, \bar{X}(t), \bar{W}(t), \underline{v})) = \alpha_0^*e + \alpha_{-0,1}^*I_2(t)e + \alpha_{-0,2}^*I_3(t)e + \beta Hx + H \log \underline{v}$$

where e is the $N \times 1$ vector of ones, with $\alpha_0^* = \alpha_0/(1-\rho)$ and $\alpha_{-0,k}^* = \alpha_{-0,k}/(1-\rho)$. The bias factor $(1-\rho)^{-1}$ follows since W is row-normalised, $We = e$, so $He = (1-\rho)^{-1}e$ because $e = HH^{-1}e = H(1-\rho)e$. As regards the bias factor of the covariate coefficient β , when W is block diagonal with blocks of size 2, $[H_{i,i}] = (1-\rho^2)^{-1}$. In the triangular case, $[H_{i,i}]$ varies slightly over i , but a good approximation is $(4-3\rho^2)/(4-5\rho^2+\rho^4)$. We denote by $\theta^* = (\alpha_{-0}^*, \beta H[i, i])$ these theoretical values.

First, we consider in Table 1 the effect on the theoretical biased values θ^* as ρ ranges from 0 to .5. The induced biases are positive, substantial and increase in ρ . For instance, when $\rho = .5$, the value for α doubles irrespective of the spatial structure, and β increases from 2 to 2.67 in the block-diagonal structure, and to 2.31 in the tri-diagonal structure.

Next, we assess the performance of the linear rank estimator in the first step by comparing the estimates $\hat{\theta}_1 = \operatorname{argmin}_{\theta \in \Theta} S_n(\theta, 0)'S_n(\theta, 0)$ to their theoretical biased counterparts $\theta^* = (\alpha_{-0}^*, \beta H[i, i])$. For both spatial structures, the first stage (biased) mean estimates are close their theoretical biased counterparts θ^* for all ρ considered, and the estimates are fairly precise, as indicated by the reported standard deviations. Hence we conclude that the linear rank estimator performs well throughout.

Finally, we turn to the second stage estimation using the linear approximation of the spatial rank function in (9). In Table 2, we report the updated estimates $\hat{\theta}$ of θ , as well as the estimate of the spatial interaction parameter ρ . In the block-diagonal structure, the mean second stage estimate of β and α is close to the population value. For the highest value of ρ the estimates are slightly higher, but the population values are contained in a 95% confidence interval. For the tri-diagonal structure, the covariate coefficient β is well estimated throughout. The bias in the duration dependence parameters α persists. Turning to the estimate of ρ , for low values the estimates are good, albeit exhibiting some variability, and the 95% confidence intervals include the population value. For the highest value of ρ , the estimate is fairly precise but underestimates the population value. This should not be a surprise, since the one-step estimator is based on a linear approximation around 0.

To summarise, the linear rank estimator works reliability in the first stage estimation. The one-step estimator successfully removes the spatial bias in the estimate of the important covariate coefficient vector; whether it also succeeds for the duration dependence function depends on the spatial structure. Irrespective of the spatial structure, ρ is estimated well for values of ρ up to .3. For higher values of ρ , the one-step estimator underestimates the population value. For practical work this implies that an estimated high value of ρ should be interpreted as a lower bound.

Table 1: First-stage results and spatial biases

		W 2-block diagonal								
		$\rho_0 = 0$	$\rho_0 = 0.1$		$\rho_0 = 0.2$		$\rho_0 = 0.3$		$\rho_0 = 0.5$	
θ	$\hat{\theta}$	θ^*	$\hat{\theta}$	θ^*	$\hat{\theta}$	θ^*	$\hat{\theta}$	θ^*	$\hat{\theta}$	
$\beta = 2$	2.00	2.02	2.01	2.08	2.09	2.20	2.19	2.67	2.65	
	(.21)		(.20)		(.21)		(.21)		(.26)	
$\alpha_{-0,1} = .3$.30	.33	.32	.37	.39	.43	.43	.60	.59	
	(.15)		(.14)		(.15)		(.16)		(.16)	
$\alpha_{-0,2} = .5$.50	.56	.55	.63	.64	.71	.71	1.00	.97	
	(.23)		(.24)		(.24)		(.25)		(.25)	
		W tridiagonal								
		$\rho_0 = 0$	$\rho_0 = 0.1$		$\rho_0 = 0.2$		$\rho_0 = 0.3$		$\rho_0 = 0.5$	
θ	$\hat{\theta}$	θ^*	$\hat{\theta}$	θ^*	$\hat{\theta}$	θ^*	$\hat{\theta}$	θ^*	$\hat{\theta}$	
$\beta = 2$	1.98	2.01	2.01	2.05	2.05	2.10	2.10	2.31	2.30	
	(.21)		(.20)		(.21)		(.21)		(.23)	
$\alpha_{-0,1} = .3$.31	.33	.33	.38	.38	.43	.43	.60	.59	
	(.15)		(.15)		(.16)		(.15)		(.16)	
$\alpha_{-0,2} = .5$.51	.57	.57	.63	.63	.71	.74	1.00	.98	
	(.23)		(.24)		(.26)		(.25)		(.24)	

Notes: Based on samples of size $N = 1,000$ and 200 repetitions. $\theta = (\beta, \alpha)$ is the population value, $\theta^* = (\beta H[i, i], \alpha_{-0}^*)$ the theoretical spatially biased value, and $\hat{\theta} = (\hat{\beta}, \hat{\alpha}_{-0})$ is the mean of the (first-stage) estimates obtained by imposing $\rho = 0$. SDs in parenthesis.

Table 2: Second stage spatial estimates

	W 2-block diagonal				
	$\rho_0 = 0$	$\rho_0 = 0.1$	$\rho_0 = 0.2$	$\rho_0 = 0.3$	$\rho_0 = 0.5$
θ	$\hat{\theta}$	$\hat{\theta}$	$\hat{\theta}$	$\hat{\theta}$	$\hat{\theta}$
$\beta = 2$	1.94 (.20)	1.94 (.21)	1.97 (.21)	1.98 (.22)	2.34 (.24)
$\alpha_{-0,1} = .3$.28 (.14)	.34 (.15)	.36 (.16)	.34 (.17)	.43 (.16)
$\alpha_{-0,2} = .5$.45 (.23)	.53 (.23)	.56 (.24)	.47 (.25)	.55 (.34)
ρ	.00 (.10)	.06 (.10)	.12 (.11)	.15 (.11)	.26 (.07)
W tridiagonal					
$\beta = 2$	1.96 (.21)	1.98 (.20)	1.99 (.21)	2.01 (.20)	2.12 (.22)
$\alpha_{-0,1} = .3$.29 (.16)	.33 (.15)	.38 (.15)	.42 (.14)	.52 (.14)
$\alpha_{-0,2} = .5$.48 (.25)	.56 (.25)	.63 (.26)	.70 (.23)	.85 (.24)
ρ	.00 (.10)	.06 (.10)	.10 (.11)	.18 (.11)	.33 (.11)

Notes: Based on samples of size $N = 1,000$ and 200 repetitions. $\hat{\theta} = (\hat{\beta}, \hat{\alpha}_{-0}, \hat{\rho})$ is the mean of the spatial estimates using (9). SDs in parenthesis.

3 Empirical application

3.1 The data: Recent Turkish labour migrants

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 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 3: Summary statistics by neighbourhood concentration: Recent Turkish labour migrants

	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
<i>Income at entry</i> [%]					
0 < income p.m. < €1000	77	65	73	78	77
<i>Neighbourhood average :</i>					
% Turks	10	42	33	35	2
perc unemployed	3.5	3.2	3.3	3.4	3.5
average income (neigh) ^a	11.4	8.4	8.9	9.4	12.3
<i>global :</i>					
quaterly unempl. rate at entry	2.9	3.1	3.0	3.0	2.9
<i>Length of stay at return migration [%]</i>					
< 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 ^b	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 Average neighbourhood income in €1000.

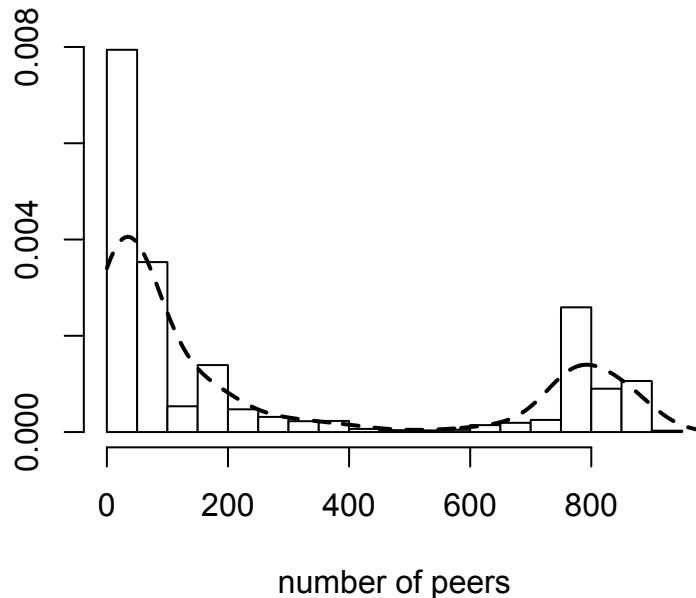
^b Migrants who remain in the country until the end of the observation period.

Table 3 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 by Figure 2, where we plot the histogram and the kernel density estimate of the number of peers (given, for each individual, by the number of non-negative row elements in the spatial interaction matrix W). 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 fairly poor, as the vast majority earn less than €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 EUR).

Figure 2: Histogram of the number of peers in neighbourhood

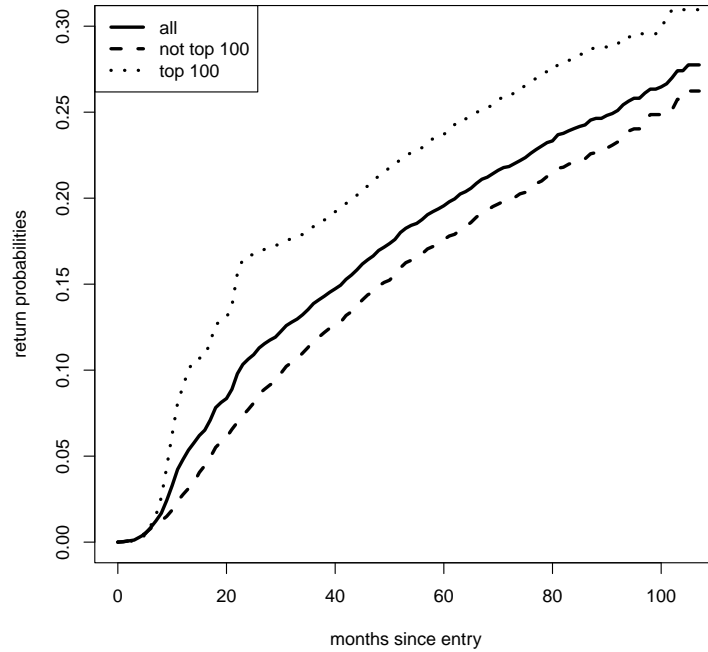


Notes. Also included is the kernel density estimate (dashed line).

Next, we turn to the outcome of interest, namely the time spent in the host country. Table 3 indicates that while censoring is high, the incidence of return 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. Moreover, 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 3 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 difference, but cannot distinguish between systematic differences in neighbourhood characteristics (contextual effects) and endogenous local interaction. We turn therefore to the role of endogenous local social interactions (captured by the parameter ρ) in explaining these difference, while controlling for differences in the the contextual effects.

Figure 3: Kaplan Meier estimates of the return probabilities by neighbourhoods



Notes. KM estimates by neighbourhoods: All (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).

3.2 Empirical results

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).

We estimate the reduced form of model (1) using our two-stage estimation strategy. Table 4 reports the results. For greater transparency, we report both the constrained and spatially biased first stage result, as well as the one-step spatial update based on (9). However, since the local interaction parameter ρ is statistically highly significant, we conclude that the latter is preferred.⁷ The second-stage coefficient point estimates are typically different from the spatially biased first-stage estimates; in particular, the magnitudes of the individual specific effects and the baseline hazard are larger for the former, giving a first illustration the importance of the spatial approach. The effect of such differences between the first and second stage estimates in terms of biases to return probabilities is briefly discussed in Section 3.3 below.

Here our discussion of the estimates concentrates on the second-stage results. The individual characteristics have the expected signs: More tied immigrants (females, married, having children) have lower return hazards, as have immigrants who are more successful on the local labour market (measured by the initial earnings' indicator). This earnings' effect is partly compensated by the average neighbourhood income effect. As regards the contextual effects, unemployment at the local level of the neighbourhood plays a much larger role than at the national level. Immigrants in neighbourhoods with high unemployment rates have significantly higher return hazards. The baseline hazard function, although imprecisely estimated, suggests positive duration dependence, so return probabilities increase in survival times which appears in line with the Kaplein Meier estimates discussed above.

Turning to the estimate of the key local interaction parameter ρ , this is estimated to be positive, fairly large, and statistically significant. This leads to the qualitative inference that, for recent Turkish labour immigrants, the propensity of one's peers to return increases one's own return hazard, which, in turn, accelerates the return of one's peers. In order to interpret better the quantitative effect of ρ , we consider the outcome of interest -the return probabilities- for the average Turkish labour immigrant (setting all covariates to their means), and juxtapose the predicted outcome for $\rho = 0$ (no social interactions) and $\rho = \hat{\rho}$. Figure 4 reports the results. For low durations, there are no appreciable differences in the return probabilities. However, as the durations of stay increase, the gap between the two functions becomes ever greater. For instance, at month 50 after arrival, social interactions imply an increase in the return probability of about 10 percentage points.

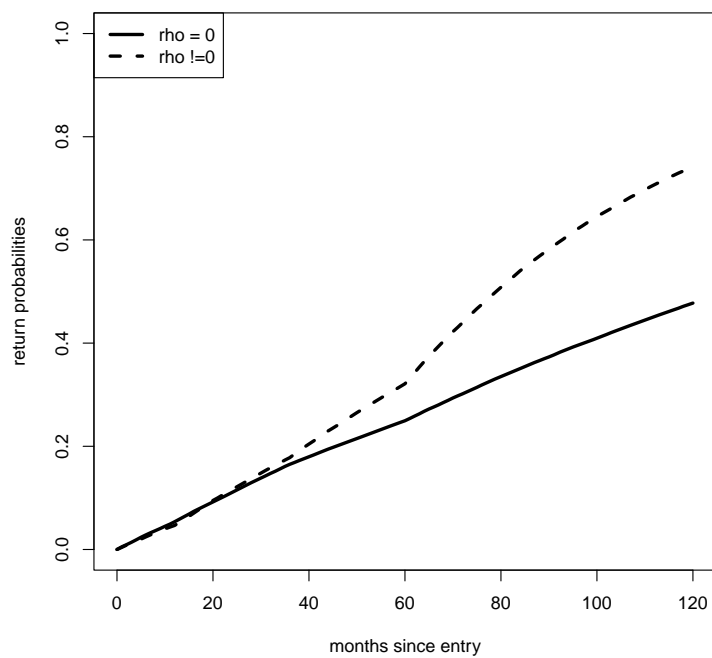
⁷In the light of this, unsurprisingly, the second stage also provides a better goodness-of-fit than the first stage, using a chi-squared criterion discussed in Heckman and Walker (1987), that considers the within-sample counting process implications (i.e. predicted and observed exits) of the duration data.

Table 4: First-stage results and spatial return hazard of Turks

	First stage		Spatial	
income at entry >1000	-1.127**	(0.084)	-1.183**	(0.080)
Female	-0.279**	(0.085)	-0.313**	(0.087)
married	-0.827**	(0.098)	-0.913**	(0.093)
divorced	-0.062	(0.123)	-0.106	(0.133)
number of children	-0.248**	(0.035)	-0.265**	(0.030)
<i>neighbourhood average:</i>				
$W \cdot (\text{income} > 1000)$	-0.910 ⁺	(0.397)	-0.137	(0.419)
$W \cdot \text{Female}$	1.283*	(0.457)	1.098 ⁺	(0.460)
$W \cdot \text{Married}$	0.624	(0.255)	0.600 ⁺	(0.282)
$W \cdot \text{divorced}$	2.507**	(0.841)	0.709	(0.856)
$W \cdot (\text{num of children})$	-0.241	(0.175)	-0.266	(0.184)
perc unemployed	1.571	(1.569)	3.833 ⁺	(1.874)
average income (neigh)	0.095**	(0.033)	0.067*	(0.023)
Unemployment (nat)	0.110 ⁺	(0.055)	0.024	(0.054)
U at entry (quart.)	-0.419*	(0.149)	-0.430*	(0.156)
<i>year of entry:</i>				
2000	0.004	(0.152)	-0.000	(0.149)
2001	-0.088	(0.210)	-0.113	(0.200)
2002	0.039	(0.221)	0.126	(0.197)
2003	1.138**	(0.365)	1.147**	(0.313)
2004	1.214**	(0.401)	1.302**	(0.377)
2005	1.255**	(0.398)	1.321**	(0.378)
2006	1.330**	(0.382)	1.243**	(0.346)
2007	1.051*	(0.445)	0.926 ⁺	(0.422)
<i>baseline hazard:</i>				
α_2 (1-3 yr)	0.003	(0.158)	0.211	(0.159)
α_3 (3-5 yr)	-0.231	(0.274)	0.255	(0.250)
α_4 (> 5 yr)	0.030	(0.346)	0.595	(0.311)
<i>social interaction:</i>				
ρ			0.431*	(0.168)

Notes: “First stage”: $\hat{\theta}_1$ is obtained from the minimisation of (7) ignoring spatial interactions, $S_n(\theta, \rho = 0)$. “Spatial” is the one-step update based on (9). The estimation includes time effects (year of entry). ‘+’ significant at the 5%, ‘*’ at the 1%, and ‘**’ at the .1% level. SE in parenthesis.

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



Notes. Covariates set at their mean in the population. Coefficients as per Table 4. Solid line ($\rho = 0$), dashed line ($\rho = \hat{\rho}$).

3.3 Counterfactual scenarios: The scope for the social multiplier

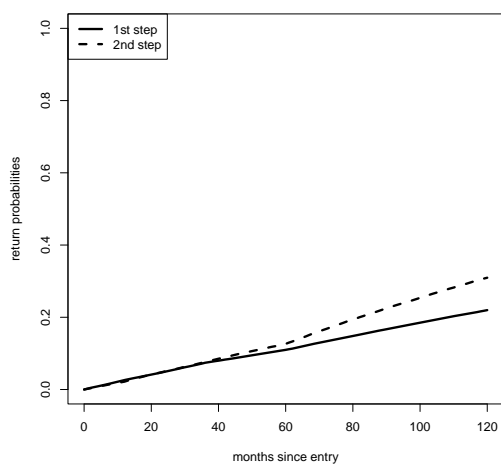
In order to explore and quantify how return probabilities, our outcome of interest, are amplified by local social interactions and the resulting social multipliers, we consider several counterfactual scenarios that capture different pull and push factors and immigrant profiles. Some of these scenarios could be thought 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. The type of counterfactual experiment considered here is, of course, constrained by the covariates included in the empirical model. Throughout we take the social interaction matrix W as given, and vary the immigrant profiles.

In the experiment (a), we consider a situation in only those higher skilled immigrants enter that have incomes above € 1000 p.m. in their first job after entry (labelled “higher incomes”), while in scenario (b) only female Turkish labour immigrants enter the country (labelled “all female”). Recall that both females and higher earners have lower return hazards. In experiment three we increase the Dutch national rate of unemployment in the quarter of arrival to 8% (labelled , “Unemployment”). In experiment four, we assume that immigrants arrive in a wave in 2006 (labelled “Entry 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.

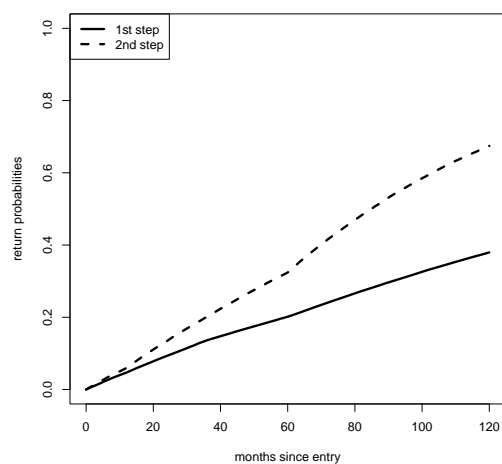
Before considering the quantification of the social multipliers, we briefly revisit the issue of the spatial bias that arises when local social interactions are wrongly ignored. In particular, we compute scenario-specific return probabilities based on the estimated model in the first and the second stage, the difference between the two defining the spatial bias. The results are depicted in Figure 5. In all but scenario (a), the magnitude of the bias is substantial reflecting the respective differences in the point estimates, while in all but scenario (d) the differences are increasing in the duration. Throughout, the spatially biased first stage under-predicts the actual return probability.

We turn to the quantification of the social multiplier by considering now the prediction based on the second stage estimates. In particular, we contrast a situation in which social interactions are present, to one in which they are absent by imposing $\rho = 0$. Figure 6 depicts the results. In the first two experiments that change a particular characteristic of the labour immigrant, the social multiplier effect appears to be fairly small. In experiment (a) the negative individual and neighbourhood average effects reinforce each other, and get further amplified by the social interaction effect ρ , leading to a relative fall in the hazard. By contrast, in experiment (b) the net effect from the two coefficients is small, and the amplification only leads to a positive but negligible effect. By contrast, in the event-based counterfactual scenarios (c) and (d), the social multiplier leads to substantial positive differences. For instance, at month 60, the absolute differences in the return probabilities are across experiments (a)-(d) .0035, .053, .079, and .223.

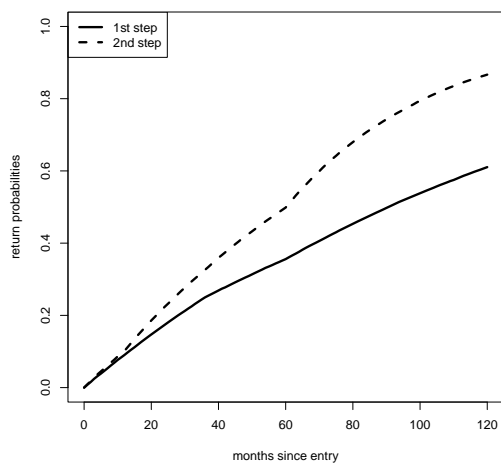
Figure 5: Spatial biases of the implied return probabilities in counterfactual scenarios



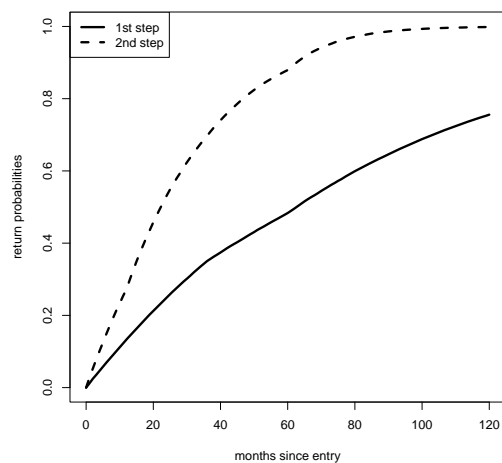
(a) Higher incomes



(b) All females



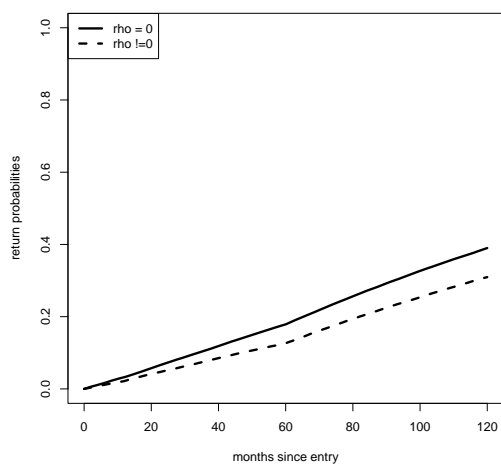
(c) Unemployment



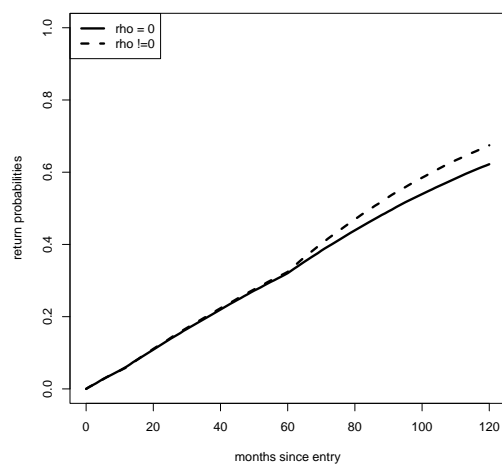
(d) Entry in 2006

Notes. For given counterfactual scenario (described in main text), (spatially biased) first-step prediction (solid lines) and second-step prediction (dashed lines).

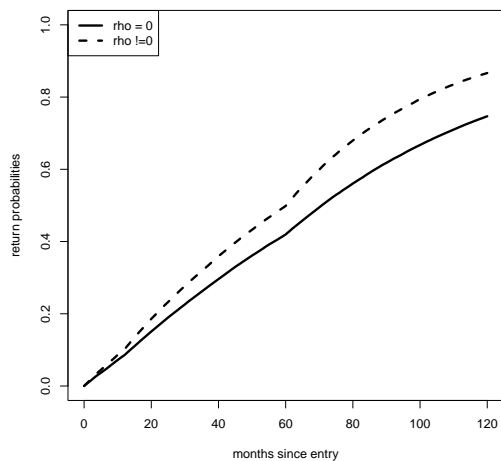
Figure 6: Social multiplier effects on return probabilities in counterfactual scenarios



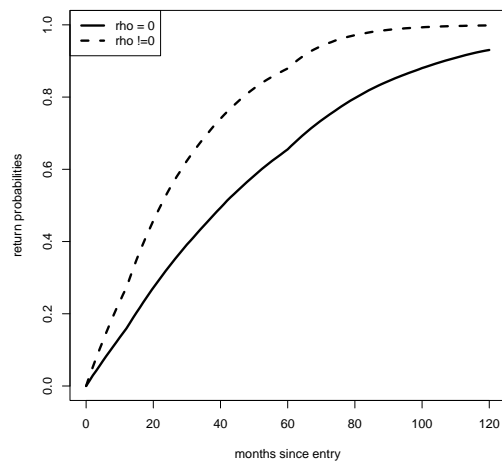
(a) Higher Incomes



(b) All females



(c) Unemployment



(d) Entry in 2006

Notes. For given counterfactual scenario (described in main text), imposing no social interaction ($\rho = 0$) (solid line) v. estimates under social interactions (ρ estimate as per Table 4, dashed lines).

4 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, 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. The illustrations quantifying this social interaction effect reveal that, albeit negligible for short durations, this effect increases substantially for longer durations.

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

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

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

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

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

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

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 v s are identically distributed). We have $H(\rho) = I + \rho W + O(\rho^2)$, 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

$$\bar{F}_{T_1}(t) = \int_{v_2} \cdots \int_{v_N} \mathcal{L}\left(z_0(t) \exp\left(\beta \sum_j H_{1j} x_j\right) \prod_{j \neq 1} v_j^{H_{1j}}\right) dG(v_2) \cdots dG(v_N) \quad (10)$$

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

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

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

$$\lim_{t \downarrow 0} -f_{T_1}(t) = E(v) \exp\left(\beta \sum_j H_{1j} x_j\right) \prod_{j \neq 1} \mu_{1j} \lim_{t \downarrow 0} [\lambda_0(t, \alpha_0)]^{H_{1\Sigma}} \quad (11)$$

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

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

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

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

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

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

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 (10), we have, say,

$$z_0(t) = \left[\exp\left(\beta \sum_j H_{1j} x_j\right) \right]^{-1} \Psi(\bar{F}_{T_1}) \quad (12)$$

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

$$\tilde{\Psi}(s) = \frac{\tilde{C}}{C} \Psi(s) \quad (13)$$

with $s = \bar{F}_{T_1}$ and $t = K(s)$. Note that for $s = 1$, $t = 0$, $z_0(0) = 0$, so $\Psi(1) = 0$. We can differentiate (13) several times under standard regularity conditions

$$\tilde{\Psi}'(s) = \frac{\tilde{C}}{C} \Psi'(s), \tilde{\Psi}''(s) = \frac{\tilde{C}}{C} \Psi''(s), \dots, \quad (14)$$

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 (10). 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

$$\frac{\tilde{C}}{C} = \frac{\tilde{\Psi}'(1)}{\Psi'(1)} = \frac{\int v^{H_{12}} dG(v)}{\int v^{\tilde{H}_{12}} d\tilde{G}(v)} \quad (15)$$

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

$$\frac{\tilde{C}}{C} = \frac{\tilde{\Psi}''(1)}{\Psi''(1)} = \frac{Var(\tilde{v}) + \mu^2}{Var(v) + \mu^2} \left[\frac{\int v^{\tilde{H}_{12}} d\tilde{G}(v)}{\int v^{H_{12}} dG(v)} \right]^{-2} \frac{\int v^{2\tilde{H}_{12}} d\tilde{G}(v)}{\int v^{2H_{12}} dG(v)} \quad (16)$$

and equalising with (15) 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 (4), $U_i = h_i(T, \bar{\theta})$, 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

$$U_i = h_i(h_{i,0}^{-1}(U_{i,0}), \bar{\theta}) \quad (17)$$

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

Lemma 3

$$\begin{aligned} F_{U_i}(u) &= F_{U_{i,0}}(h_{i,0}(h_i^{-1}(u, \bar{\theta}))), \\ \kappa_{U_i}(u, \bar{\theta}) &= \kappa_{i,0}(h_{i,0}(h_i^{-1}(u, \bar{\theta}))) \frac{h'_{i,0}(h_i^{-1}(u, \bar{\theta}))}{h'_i(h_i^{-1}(u, \bar{\theta}), \bar{\theta})} \end{aligned} \quad (18)$$

Proof. We have

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

The second claim follows by direct computation. ■

Simplifying (18) using (4) yields

$$\begin{aligned} \kappa_{U_i}(u, \bar{\theta}) &= \exp(e'_i H_0(h_i^{-1}(u, \bar{\theta})) X^*(h_i^{-1}(u, \bar{\theta})) \theta_0 - e'_i H(h_i^{-1}(u, \bar{\theta}); \rho) X^*(h_i^{-1}(u, \bar{\theta})) \theta) \\ &\cdot \kappa_{i,0}(h_{i,0}(h_i^{-1}(u, \bar{\theta}))). \end{aligned} \quad (19)$$

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 i th row of H)

$$\begin{aligned}
\kappa_{i,0}(u) &= E_{\underline{v}}\{\exp(H_i.(\rho_0) \log \underline{v}) | T_i \geq h_0^{-1}(u)\} \\
&= \int_{\underline{v}} \exp(H_i.(\rho_0) \log v) \exp(-u \times \exp(H_i.(\rho_0) \log \underline{v})) dG_{\underline{v}}(v) \\
&\times \left[\int_{\underline{v}} \exp(-u \times \exp(H_i.(\rho_0) \log \underline{v})) dG_{\underline{v}}(v) \right]^{-1} \\
&= \int_{v_1} \cdots \int_{v_n} \prod_j v_j^{H_{ij}(\rho_0)} \exp(-u \times \prod_j v_j^{H_{ij}(\rho_0)}) dG(v_n) \cdots dG(v_1) \\
&\times \left[\int_{v_1} \cdots \int_{v_n} \exp(-u \times \prod_j v_j^{H_{ij}(\rho_0)}) dG(v_n) \cdots dG(v_1) \right]^{-1} \tag{20}
\end{aligned}$$

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 4 *Under the stated assumptions*

$$\left| \kappa_U(u, \bar{\theta}) - \kappa_0(u) - \frac{\partial \kappa_U}{\partial \bar{\theta}'}(u, \bar{\theta}_0)(\bar{\theta} - \bar{\theta}_0) \right| \leq \|\bar{\theta} - \bar{\theta}_0\|^2 \eta(u) \tag{21}$$

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

Proof. The assumptions that $0 < |\partial^2 \lambda(t, \alpha) / \partial \alpha \partial \alpha'| < \infty$ 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 $g_u(\bar{\theta}) = h_i^{-1}(u, \bar{\theta})$ for ease of notation:

Lemma 5

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u, \bar{\theta})}{\partial \theta} &= [\kappa_{i,0}(h_{i,0}(h_i^{-1}(u, \bar{\theta})))] \\
&\times \exp(H_{0,i}(g_u(\bar{\theta}))X^*(g_u(\bar{\theta}))\theta_0 - H_i(g_u(\bar{\theta}); \rho)X^*(g_u(\bar{\theta}))\theta) \\
&\times (-1)H_i(g_u(\bar{\theta}); \rho)X^*(g_u(\bar{\theta})) \\
&+ \exp(H_{0,i}(g_u(\bar{\theta}))X^*(g_u(\bar{\theta}))\theta_0 - H_i(g_u(\bar{\theta}); \rho)X^*(g_u(\bar{\theta}))\theta) \\
&\times \kappa'_{i,0}(h_{i,0}(g_u(\bar{\theta}))) \\
&\times \exp(H_{0,i}X^*(g_u(\bar{\theta}))\theta_0) \\
&\times \exp(-H_i(\rho)X^*(g_u(\bar{\theta}))\theta) \\
&(-1) \int_0^T \exp(H_i(\rho)X^*(s)\theta)H_i(\rho)X^*(s)ds
\end{aligned} \tag{22}$$

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u, \bar{\theta})}{\partial \rho} &= -\exp(H_{0,i}(g_u(\bar{\theta}))X^*(g_u(\bar{\theta}))\theta_0 - H_i(g_u(\bar{\theta}); \rho)X^*(g_u(\bar{\theta}))\theta) \\
&\times H_iWHX^*(g_u(\bar{\theta}))\theta \\
&\times \kappa_{i,0}(h_{i,0}(g_u(\bar{\theta}))) \\
&- \exp(H_{0,i}(g_u(\bar{\theta}))X^*(g_u(\bar{\theta}))\theta_0 - H(g_u(\bar{\theta}); \rho)X^*(g_u(\bar{\theta}))\theta) \\
&\times \kappa'_{i,0}(h_{i,0}(g_u(\bar{\theta}))) \\
&\times \exp(H_{0,i}X^*(g_u(\bar{\theta}))\theta_0) \\
&\times \exp(-H_i(\rho)X^*(g_u(\bar{\theta}))\theta) \\
&\times \int_0^{g_u(\bar{\theta})} \exp(H_i(\rho)X^*(s)\theta)HWHX^*(s)\theta ds
\end{aligned} \tag{23}$$

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 $h_{i,0}^{-1}(u) = s$, (22) and (23) simplify to

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u, \theta_0, \rho_0)}{\partial \theta} &= -\kappa_{i,0}(u)(e'_i H_0(h_{i,0}^{-1}(u))X^*(h_{i,0}^{-1}(u))) \\
&- \kappa'_{i,0}(u) \int_0^u e'_i H_0(h_{i,0}^{-1}(s))X^*(h_{i,0}^{-1}(s))ds
\end{aligned} \tag{24}$$

$$\begin{aligned}
\frac{\partial \kappa_{U_i}(u, \theta_0, \rho_0)}{\partial \rho} &= -e'_i H_0WH_0X^*(h_{i,0}^{-1}(s))\theta_0 \times \kappa_{i,0}(u) \\
&- \kappa'_{i,0}(u) \int_0^u e'_i H_0(h_{i,0}^{-1}(s))W(h_{i,0}^{-1}(s))H_0(h_{i,0}^{-1}(s))X^*(h_{i,0}^{-1}(s))\theta_0 ds.
\end{aligned} \tag{25}$$

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

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

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 $\bar{Y}(t) = [\bar{y}_i(t)]_{i=1,\dots,n}$ denote the history of the observation indicators. We then have

Lemma 6

$$\Pr(dN^{U_i}(u, \bar{\theta}) = 1 | \bar{X}^{U_i}(u, \bar{\theta}), \bar{Y}^{U_i}(u, \bar{\theta}), \bar{W}^{U_i}(u, \bar{\theta})) = y^U(u, \bar{\theta}) \kappa_{U_i}(u; \bar{\theta}) du \quad (26)$$

with κ_{U_i} given by (18). The associated Doob-Meyer decomposition is

$$dN^{U_i}(u, \bar{\theta}) = y_i^U(u, \bar{\theta}) \kappa_{U_i}(u; \bar{\theta}) du + dM^{U_i}(u, \bar{\theta}) \quad (27)$$

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 (27), the estimation function can be written as

$$\begin{aligned} S_n(\bar{\theta}) &= \sum_{i=1}^n \Delta_i \left[x_i(U_{(i)}) - \tilde{x}(U_{(i)}) \right] \\ &= \sum_{i=1}^n \int_0^\tau (x(u) - \tilde{x}(u)) dN^{U_i}(u, \bar{\theta}). \end{aligned} \quad (28)$$

The transformed durations are observed up to time $\tau < \infty$. Evaluating the estimation function (28) at the population parameters, we have

$$\begin{aligned} S_n(\bar{\theta}_0) &= \sum_{i=1}^n \int_0^\tau (x(u) - \tilde{x}(u)) dM^{U_i,0}(u) \\ &\quad + \sum_{i=1}^n \int_0^\tau (x(u) - \tilde{x}(u)) y_i^{U_0}(u) \kappa_{U_i}(u; \bar{\theta}_0) du \end{aligned} \quad (29)$$

Lemma 7 *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 (4), $\exp(e'_i H(t; \rho_0) X^*(t) \theta_0) E_{\underline{v}}(\psi_i | T \geq t)$. 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

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

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 (29) is zero, so

$$S_n(\bar{\theta}_0) = \sum_{i=1}^n \int_0^\tau (x(u) - \tilde{x}(u)) dM^{U_i,0}(u). \quad (30)$$

Using again the representation (27) for $S_n(\bar{\theta})$, we obtain the following linearisation

$$\tilde{S}_n(\bar{\theta}) = S_n(\bar{\theta}_0) + G(\bar{\theta}_0) \times (\bar{\theta} - \bar{\theta}_0) \quad (31)$$

with

$$\begin{aligned} G(\theta_0, \rho_0) &\equiv (\partial S / \partial \theta, \partial S / \partial \rho) |_{\bar{\theta}=\bar{\theta}_0} \\ &= \sum_{i=1}^n \int_0^\tau (x(u) - \tilde{x}(u)) \frac{\frac{\partial}{\partial \theta} \kappa_{U_i}(u, \bar{\theta}) |_{\bar{\theta}=\bar{\theta}_0}}{\kappa_{U_i}(u, \theta_0)} dN^{U_i}(u, \bar{\theta}) \end{aligned} \quad (32)$$

where $(\partial / \partial \bar{\theta}) \kappa_{U_i}(u, \bar{\theta})$ is given in Lemma 5 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 8 *Under the stated assumptions for all $c > 0$*

$$\sup_{|\bar{\theta} - \bar{\theta}_0| \leq cn^{-\frac{1}{2}}} n^{-\frac{1}{2}} \left| S_n(\bar{\theta}) - \tilde{S}_n(\bar{\theta}) \right| \xrightarrow{p} 0 \quad (33)$$

Finally, we observe that our estimator is the root of $\tilde{S}_n(\bar{\theta})$. Hence solving (31) for $(\bar{\theta} - \bar{\theta}_0)$, and invoking the asymptotic normality of $S_n(\bar{\theta}_0)$ implied by (30), yields the result stated in Theorem 1: 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
	<i>By City</i>				<i>By Ethnic Group</i>			
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
	<i>Dissimilarity</i>				<i>Isolation</i>			
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.

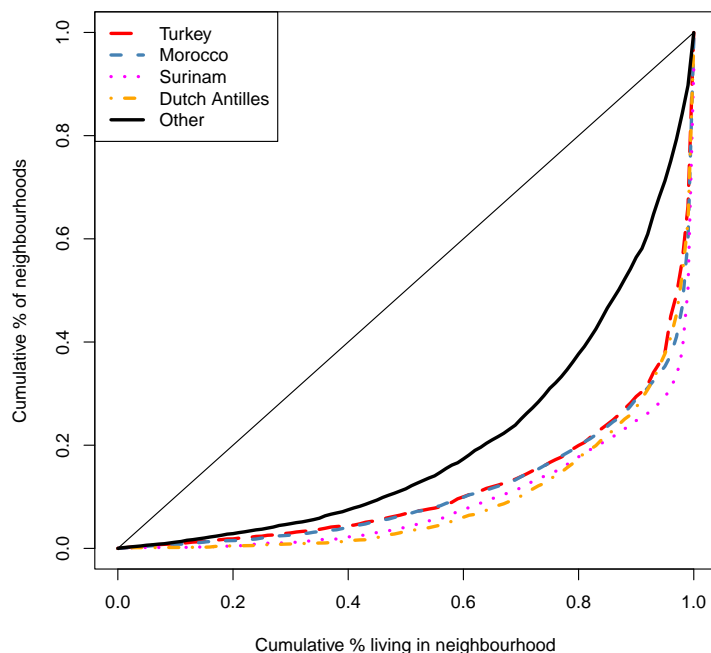
Next, we consider these four cities at the level of the neighbourhood and investigate the extent to which immigrants of a particular ethnic group, such as Turks, (we label them “minority”, min_n in neighbourhood n) differ from natives and other immigrants in this neighbourhood (label this complement 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

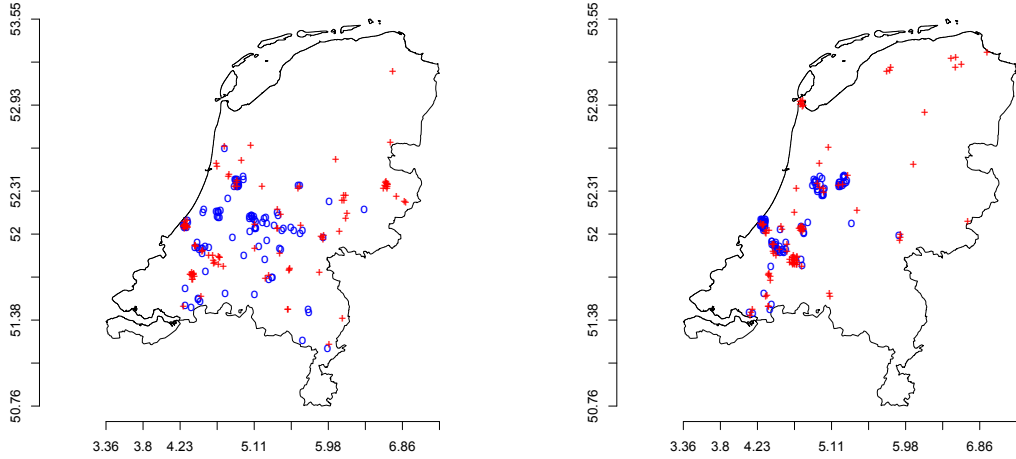
Turning to all neighbourhoods, Figure B.1 depicts the Lorenz curve for spatial concentration. It is evident that most migrants live in a relatively small number of neighbourhood, and this extent of clustering is much larger than for all other immigrants. The Lorenz curve reveals the extent of spatial 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 , $W_g = [w_{g,ij}]_{i,j=1,\dots,100}$, 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 $B(\beta) = (I_{100} - \beta W_g)^{-1} W_g \mathbf{1}_{100} = \sum_{k=0}^{\infty} \beta^k W_g^{k+1} \mathbf{1}_{100}$ where I_{100} is the identity matrix and $\mathbf{1}_{100}$ is a vector of ones. $B(\beta) = [b(\beta)_i]_{i=1,\dots,100}$ 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

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 Antillean neighbourhoods

($\max(B) = 2.05$), 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 ($\max(B) = 1.94$).

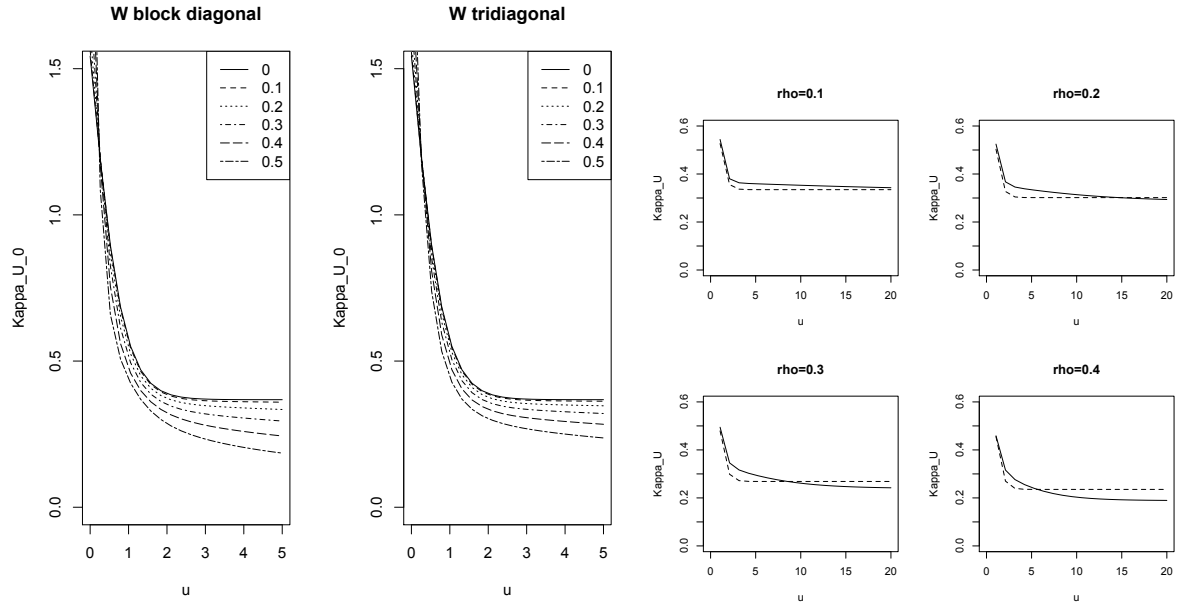
C Further numerical illustrations - Not for publication

In this supplementary appendix we revisit the setting of the simulation study of Section 2.5, and illustrate for completeness the spatial effects in terms of the hazard of U_0 , κ_{U_0} , in the transformation model (4). We also illustrate the quality of the first order approximation of the hazard of the transformed variate U_0 .

Consider then the hazard of U_0 , κ_{U_0} (which is stated explicitly in Appendix equation (20)), and the spatial effects for the blockdiagonal and the tridiagonal spatial structures. This is done in Figure C.3 panels A and B, which correspond to Panels A and B of Figure 1. Note that the associated hazards are now decreasing in ρ . Of course, $\kappa_T(t) = \kappa_{U_0}(h_0(t))h'_0(t)$, but for general covariate and parameter values the sign of the derivative with respect to ρ involves the ambiguous sign of $\partial h'_0(t)/\partial \rho = h'_0(t)H_0WH_0X^*(t)\theta_0$. As in the case of κ_T , the changes in κ_{U_0} are larger in the block diagonal than in the tridiagonal structure, but the magnitude of the changes are ‘compressed’. At $u = 5$, κ_{U_0} with $\rho = .5$ has fallen to 50% of its value when $\rho = 0$ in the block diagonal structure. In the tridiagonal structure the corresponding fall is to 64% of its former value.

Lastly, we consider the quality of the first order approximation of the hazard of

Figure C.3



Notes. Panels A and B: Spatial impacts on the hazard κ_{U_0} of the transform. Panels C and D: The first order approximation to the hazard κ_{U_0} of the transform in the block-diagonal case. The dashed line depicts the first order approximation

the transformed variate U_0 ; this is feasible since the two particular spatial structures permit the analytical calculation of H . For the sake of brevity only the block diagonal case is considered, as the spatial impacts and thus the scope for approximation errors are substantially larger here than in the tridiagonal case. Panels C and D of Figure C.3 shows that the quality of the approximation in the considered setting is good.

REFERENCES

- 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.
- 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.
- 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”, forthcoming *Journal of Economic Literature*.

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.

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.

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.

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.