CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 342

Understanding employment decentralization by estimating the spatial scope of agglomeration economies

Paul Verstraten
Gerard Verweij
Peter Zwaneveld

Understanding employment decentralization by estimating the spatial scope of agglomeration economies*

Paul Verstraten, Gerard Verweij, and Peter Zwaneveld†

February 2017

Abstract

This paper argues that the spatial decay function of agglomeration economies is much more complex than is often assumed in the agglomeration literature. We provide insight into this issue by analyzing a nationwide and spatially rich wage panel. The key finding is that wages and urbanization are not significantly related on short distances (<5km), while strongly and positively related on medium distances (5-10km). This positive effect attenuates across geographic space and becomes insignificant after 40-80km. This non-monotone relation between wages and distance to economic mass is in line with recently observed trend towards employment decentralization, because it suggests that agglomeration economies on short distances, i.e. in city centers, are offset by congestion externalities. Additionally, this paper finds no evidence that foreign economic mass affects wages in the Netherlands, which suggests that national borders are still a substantial barrier for economic interaction.

JEL Codes: R12; J31

Keywords: Agglomeration economies; Employment decentralization; Panel data analysis; Spatial wage disparities; Urbanization

* We are grateful to Steven Brakman, Joost van Gemeren, Stefan Groot, Mark Kattenberg, Clemens Kool, Hans Koster, and participants at the CPB seminar for their comments on an earlier version of this paper. Furthermore, we thank Thomas Tichelbäcker for his assistance in gathering data on German historical population censuses. The usual disclaimer applies.

† All authors are affiliated with the CPB Netherlands Bureau for Economic Policy Analysis. Views expressed in this paper are our own and do not necessarily reflect those of the CPB. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector. Corresponding author: Paul Verstraten (p.verstraten@cpb.nl).

1. Introduction

Although the agglomeration literature has reached no consensus on the maximum spatial range of agglomeration economies[1], most studies do agree on how these agglomeration externalities decay across geographic space (Rice et al., 2006; Arzaghi and Henderson, 2008; Di Addario and Patacchini, 2008; Rosenthal and Strange, 2008; Koster, 2013; Ahlfeldt et al., 2015). In general, the literature concludes that the relation between the spatial concentration of employment and productivity is strongest on short distances and decays rapidly across geographic space. For some studies, this consistent finding in the literature has even been a reason to assume a priori that agglomeration externalities decay monotonically across space (e.g. Rice et al., 2006; Koster, 2013).[2]

This paper argues that the spatial decay function of agglomeration economies is much more complex than is often assumed. Although there are sound arguments to think that individual mechanisms underlying agglomeration externalities[3] can be described by a monotonically declining distance decay function, it is a misconception to assume the same for the net effect of all these externalities, especially when individual mechanisms work in opposite directions. This paper provides empirical evidence for this complexity by showing that the gains from urbanization exhibit a remarkable spatial pattern. On relatively short distances (<5km) the degree of urbanization does not significantly affect wages, whereas on medium distances (5-10km) it has a strong and significant positive effect. This effect attenuates across geographic space and becomes insignificant after 40-80km.

The results of our study contribute to the understanding of the recently observed trend towards employment decentralization and urban sprawl. These empirical phenomena, which have been strikingly omnipresent at the end of the 20th century in the United States (Glaeser and Kahn, 2001) and have, more recently, also shaped urban growth in the Netherlands (Nabielek et al., 2012), suggest that close proximity to a Central Business District is becoming less attractive. A non-monotonic distance decay function is in line with decentralized employment and urban sprawl,

[1] Estimates range from 40-80km (Rosenthal and Strange, 2008) to only a few kilometers (Arzaghi and Henderson, 2008; Ahlfeldt et al., 2015), and everything in between (e.g. Di Addario and Patacchini, 2008). This lack of consensus can however be attributed to differences regarding the area under scope. For instance, it is self-evident that studies that analyze data within one particular city (e.g. Arzaghi and Henderson, 2008; Ahlfeldt et al., 2015) are well-suited for identifying agglomeration economies with a narrow spatial scope, although they are, by construction, unable to identify agglomeration economies that stretch beyond city borders. The opposite goes for studies that use nationwide data at a highly aggregated spatial level (e.g. Rosenthal and Strange, 2008; Rice et al., 2006; Di Addario and Patacchini, 2008).

[2] Studies using a market-potential function based on Harris (1954) also assume there is a monotone relation between measures of distance and the weight attached to another region's economic mass, e.g. Hanson (2005) and Combes et al. (2008).

[3] For instance, it is hypothesized that dense urban surroundings foster the diffusion of ideas (learning), promote efficient labor market coordination (matching), and increase the availability of differentiated intermediate inputs (sharing). See Duranton and Puga (2004) for an extensive theoretical overview of these individual mechanisms. As opposed to the gains from agglomeration economies, there also exist mechanisms that lead to congestion costs, e.g. traffic congestion, pollution, and small lot sizes.

because it suggests that agglomeration economies are offset by urban congestion costs on short distances, i.e. in city centers, whereas the gains from urbanization dominate the costs on longer distances, i.e. at the outskirts of cities. This interplay between agglomeration economies and urban congestion at various distances gives an adequate explanation for the observed decentralized employment. Interestingly, our estimates are consistent with the observation of Glaeser and Kahn (2004) that per capita income in US metropolitan areas correlates positively with the share of jobs more than 5km from the city center.

In order to reveal the complexities underlying the spatial decay function, this study employs a nationwide wage panel with a high level of geographic detail: Dutch ZIP codes with a mean area of only 9km^2. The use of this dataset has two advantages compared to earlier work. Firstly, the spatial richness of the dataset enables us to construct concentric ring variables which are very narrow compared to earlier studies. This high level of spatial detail is an important prerequisite to disentangle the effects of urbanization on very short distances (e.g. within 2.5km) from those on longer distances. Similar studies on the spatial scope of agglomeration economies have relied on spatial units that are much larger than the Dutch ZIP code: e.g. 6,522km^2 (Rosenthal and Strange, 2008), 1,394km^2 (Rice et al., 2006) and 889km^2 (Di Addario and Patacchini, 2008). It is evident that this lack of spatial detail in most other studies makes it difficult to identify the effects of urbanization on various short distances.[4]

The second key advantage relates to the longitudinal nature of the wage data. By following workers over time, we are able to control for both observed and unobserved worker characteristics. This is crucial for the identification of agglomeration economies, since it is well known that a considerable part of the urban wage premium is driven by the sorting of high-skilled workers into urban areas (Combes et al., 2008). Other studies on the spatial scope of agglomeration economies using wage data (e.g. Rosenthal and Strange, 2008; Di Addario and Patacchini, 2008; Rice et al., 2006), however, have controlled for observed worker characteristics only. Therefore, these studies run the risk of omitting important unobserved differences in labor quality. This spatial sorting of workers by skills is also the main reason why we prefer the use of wages over a TFP-based approach, which has difficulties in accounting for regional differences in labor quality.

Our finding that agglomeration economies stretch across a relatively large distance also raises questions about the role of foreign urbanization in domestic wage formation. After all, the Netherlands is a small country, part of the European Single Market, and shares a common language

[4] In fact, some of the few studies that have employed a dataset with a high level of spatial detail to examine the spatial scope of agglomeration economies, do find evidence against the idea that agglomeration economies decline monotonically across space (Rosenthal and Strange, 2003; Duranton and Overman, 2005). These studies analyze the birth-rate of new establishments and location patterns of industries. Arzaghi and Henderson (2008) and Ahlfeldt et al. (2015) do not find evidence against a monotonically declining spatial decay function, despite the high level of spatial detail in these studies. It should be noted, however, that these studies examine relatively small areas (Manhattan and Berlin, respectively), which makes it practically impossible to detect agglomeration economies with a larger spatial extent, see also footnote 1.

with the Northern part of Belgium. Therefore, in order to assess the influence of foreign urbanization, we have constructed a unique dataset containing information on the current spatial distribution of employment and historical (19^{th} century) population censuses for both Belgium and Germany. Despite the openness of the Dutch economy, our analysis provides no compelling evidence that foreign economic mass affects wages in the Netherlands. This result is nevertheless consistent with the bulk of the literature, which finds substantial border barriers (e.g. Brakman et al., 2002). These substantial border effects are relevant for policymakers in the Netherlands, due to its implication that productivity levels in the Netherlands, and especially in border regions, can be fostered by reducing border barriers.

The remainder of this paper is structured as follows. Section 2 discusses the micro-econometric specification, which enables us to estimate the spatial scope of agglomeration economies. In Section 3 we describe the wage data and the process of constructing concentric ring variables. Section 4 reports the results and robustness checks, whereas Section 5 provides a discussion of the results and how to reconcile these with findings in related literature. Section 6 concludes.

2. Methodology

In order to analyze the relation between wages and urbanization, this paper employs a two-stage estimation approach as proposed by Combes et al. (2008). In the first stage of this approach, we estimate area fixed-effects using a Mincerian wage equation. These area fixed-effects can be interpreted as area-specific productivity indices. Then, in the second stage, we explain these area fixed-effects using concentric ring variables that measure the employment levels at various distances, as proposed by Rosenthal and Strange (2003). An important benefit of this two-stage estimation approach is the elegant solution of the dependent disturbances within the regional units.[5] We will further elaborate on this two-stage approach in the remainder of this section.

Two-stage estimation approach

In spatial and competitive equilibrium, a profit-maximizing competitive firm in area r, industry k and year t pays wages equal to the marginal product of labor. Hence, following Combes et al. (2008), the hourly wage of worker i in year t can be described as:

$$\log w_{i,t} = \beta X_{i,t} + \delta_r R_{r(i,t)} + \delta_k K_{k(i,t)} + \delta_t T_{t(i,t)} + \varepsilon_{i,t} \quad (1)$$

[5] Non-independent disturbances may arise because observations sharing the same geographic space might influence each other and/or might be subject to the same local shocks. Neglecting this dependence often leads to downward biased standard errors (Moulton, 1990).The standard solution of calculating cluster robust standard errors assumes nesting of the workers within the same regional cluster. However, our study relies on workers who change their working location, making the default use of cluster robust standard errors not applicable.

where the log-transformed hourly wage $w_{i,t}$ is explained by a vector of worker characteristics $X_{i,t}$ and productivity effects unrelated to worker characteristics. The latter consists of a vector of area-dummies $R_{r(i,t)}$ indicating the individual's place of work, a vector of industry-dummies $K_{k(i,t)}$, and a vector of year-dummies $T_{t(i,t)}$. The vectors β, δ_r, δ_k and δ_t contain the parameters to be estimated, and $\varepsilon_{i,t}$ is a random error term.[6]

It is commonly acknowledged in the agglomeration literature that the urban wage premium might be driven by the sorting of high-skilled workers into urban areas. This implies that $\text{cov}\left(X_{i,t}, R_{r(i,t)}\right) \neq 0$. In order to identify the area-specific productivity effects under the presence of sorting, it is necessary to include variables that capture all relevant worker characteristics $X_{i,t}$. To this end, all studies to date that examine the geographic scope of agglomeration economies using wage data (e.g. Rosenthal and Strange, 2008; Di Addario and Patacchini, 2008; Rice et al., 2006), have employed observed characteristics to control for worker heterogeneity. However, these studies run the risk of having omitted some worker characteristics that correlate with the region-specific productivity effects. In contrast to these earlier works, our study relies on worker fixed-effects to control for all time-invariant worker characteristics. The age of workers and its square are used as a proxy for worker experience. The regression equation then becomes:

$$\log w_{i,t} = \theta_i + \beta_1 \widetilde{age}_{i,t} + \beta_2 \widetilde{age}_{i,t}^2 + \delta_r R_{r(i,t)} + \delta_k K_{k(i,t)} + \delta_t T_{t(i,t)} + \varepsilon_{i,t} \qquad (2)$$

where θ_i represents a worker fixed-effect and $\widetilde{age}_{i,t}$ denotes the age of a worker. The worker's age is centered around its industry-average to account for the fact that some industries tend to hire older/younger workers (Combes et al., 2008). The squared term captures any concave effects of age on wages.

It should be noted that, with this specification, the area-specific effects on wages are assumed to be static. This means that the model ignores potential area-specific wage growth effects, like dynamic agglomeration economies. Although these dynamic effects can potentially bias the estimates of the static effect, we know from De la Roca and Puga (2017) that standard worker fixed-effects estimates of the static gains from urbanization are, under reasonable circumstances, insensitive to the existence of dynamic effects. Given these considerations, we will rely on the standard fixed-effects model of Equation (2).

[6] In this specification, we ignore potential interactions between the area-, industry- and time-specific productivity effects. This is for practical reasons, as estimating the full interaction set ($\delta_{r,k,t} R_{r(i,t)} K_{k(i,t)} T_{t(i,t)}$) would require the inclusion of roughly 2.3 million fixed effects, besides the 2.3 million worker fixed-effects. Equation 1, on the other hand, would require only 3,800 additional fixed effects. Since our main interest is ultimately in the effect of urbanization on wages, it is not strictly necessary to include an area-year interaction, because the spatial distribution of economic mass does not vary much over time (on the ZIP code-level, the correlation coefficient between the number of jobs in 2006 and 2014 equals 0.982). Nevertheless we will show in the robustness section of this paper that our results are robust to the inclusion of an area-year interaction.

The area fixed-effect estimates that are obtained from Equation (2), reflect regional differences in productivity. The equation below describes how these productivity differences are the result of a complex interplay between agglomeration economies and urban congestion costs at various distances:

$$\delta_r = \phi \sum_e E_e \mathrm{B}(D_{r,e}) - \omega \sum_e E_e \mathrm{C}(D_{r,e}) + \varepsilon_r \tag{3}$$

where δ_r is the area fixed-effect parameter from Equation (2), and E_e denotes total employment at establishment e, which we use as a measure of urbanization.[7] $\mathrm{B}(D_{r,e})$ and $\mathrm{C}(D_{r,e})$ represent the distance decay functions of respectively the agglomeration economies and urban congestion costs. These distance decay functions provide weights to employment at various distances and, without loss of generality, can take any value between 0 and 1, depending on the straight line distance between area r and establishment e ($D_{r,e}$).[8] The parameters ϕ and ω represent the wage effect of the spatially weighted agglomeration measures, and ε_r is a random error term.

Estimating Equation (3) is, however, not possible because both the agglomeration economies and congestion costs stem from the same source (E_e), which makes it virtually impossible to disentangle these two effects. It is for this reason that this study, as the other studies in this field of research, estimates the net effect of urbanization:

$$\delta_r = \gamma \sum_e E_e \mathrm{N}(D_{r,e}) + \varepsilon_r \tag{4}$$

Estimating the net effect of urbanization has one crucial implication for our setting. Even when the partial effects of agglomeration, $\mathrm{B}(D_{r,e})$ and $\mathrm{C}(D_{r,e})$, decay monotonically across space, there is no a priori reason to expect that the net effect decays monotonically as well. In fact, the decay

[7] Much debate exists in the literature about whether agglomeration economies arise from the concentration of industries (localization) or from the overall size of the market (urbanization). In this paper, our main interest lies in the effect of urbanization, measured in terms of total employment. This is in line with most studies that examine the spatial extent of agglomeration economies (e.g. Rice et al., 2006; Rosenthal and Strange, 2008; Di Addario and Patacchini, 2008; Koster, 2013; Rice et al., 2006). It should however be noted that the patterns of spatial decay can differ between localization and urbanization economies (Rosenthal and Strange, 2003).

[8] Ideally, we would use measures of effective distance, taking into account natural barriers and transport infrastructure, rather than straight line distance. Yet, the use of effective distance would inevitably lead to problems of reverse causality, because the more productive and dense regions tend to have more and better transport connections than less productive areas. For this reason, we prefer the use of straight line distance, which is arguably a more exogenous measure than effective distance. The results of this paper are nevertheless unlikely to be very different when using travel time on roads as a measure of effective distance because this is generally highly correlated to straight line distances (Phibbs and Luft, 1995). Based on our own calculations, we find that straight line distances between ZIP codes in the Netherlands can explain around 94% of the variation in travel time on roads as predicted by the LMS National Model System for Traffic and Transport.

function of the net effect is characterized by a wide variety of functional forms. Figure A.1 in the Appendix shows one set of possible functional forms for the decay function.

In order to estimate Equation (4) we construct a set of concentric ring variables that measure total employment at various distance intervals, e.g. within 2.5km, between 2.5 and 5km, etc.[9] This flexible strategy is preferred over strategies that employ a pre-defined monotonically declining decay function (e.g. Rice et al., 2006; Koster, 2013), because of the aforementioned reason that the net effect of urbanization might decline non-monotonically across space. In line with the work of Rosenthal and Strange (2003), the regression equation of the second stage then becomes:

$$\delta_r = \sum_{D_d} \gamma_d \sum_{D_{r,e} \in D_d} E_e + \varepsilon_r \tag{5}$$

where the first summation is over all concentric rings at various distance intervals D_d. The second summation term aggregates all employment that falls within that particular distance interval. In this specification, the estimated parameters of the concentric ring variables (γ_d) give the percentage wage effect of an additional unit of employment within a particular distance interval. A numerical comparison of these parameters provides information on how the net agglomeration economies differ across geographic space.[10]

When determining the width of the distance intervals, we encounter practical limitations. Although the construction of very narrow distance intervals will, in theory, render a detailed distance decay pattern, it will also lead to serious multicollinearity problems. In particular, this problem tends to become more severe as the distance from area r gets larger.[11] Therefore, in order to avoid problems of multicollinearity, the distance intervals must be somewhat wider on longer distances than on shorter distances. We use the following regular set of cutoff values for our distance intervals: 2½, 5, 10, 20, 40, 80, and 120 kilometer.[12]

---

[9] Note that this approach implies that we use a Riemann summation as an approximation of the true distance decay function.

[10] Throughout this paper we use standard OLS and IV regressions for the second-stage estimation, although this will generally lead to biased and inefficient estimates as Combes et al. (2008) point out. The size of this bias and inefficiency depends on the standard error of the estimated area fixed-effects in the first-stage. We have re-estimated the model with the correct equations provided by Gobillon (2004) for an unbiased estimate of the standard errors in the second-stage and with an efficient FGLS estimator. Both estimations provided results comparable to the standard estimation strategy. The difference in estimated parameters and standard errors was generally below 10% (results are available upon request). We therefore conclude that the influence of estimation errors of the area fixed-effects from the first-stage can be neglected during the second-stage.

[11] The circumference of a concentric ring increases proportionally to its radius, which makes that more spatial detail is lost when employment is aggregated at longer distances from a particular area. For that reason, the bivariate correlations tend to be larger between rings on long distances than between those on short distances.

[12] We want the first concentric ring to cover most of the ZIP code's own area. Given the size of most ZIP code areas, this implies that the width of the first concentric ring cannot be smaller than approximately 2½ kilometer. Then we keep doubling the cutoff values, in order to avoid multicollinearity problems, until 120km. As we will see in Section 4 of this paper, 120km is far enough to get an insignificant estimate on long distances. Compared to earlier studies that employ concentric ring variables to explain the urban wage premium, we use a rather

Instrumental variable approach

Finally, a word on one of the classical problems in the agglomeration literature: endogeneity of the agglomeration measure. This issue of endogeneity means that the estimated relationship between urbanization and wages might be driven by omitted variables, like (non-)human local endowments, and/or reverse causality. To tackle this endogeneity problem, the literature has suggested several approaches; see Rosenthal and Strange (2004) and Combes et al. (2010) for an extensive discussion.

This paper applies the instrumental variable (IV) approach with two sets of instruments.[13] Firstly, we compute concentric ring variables that measure historical (year 1840) population counts. This set of variables will be used as an instrument for the concentric ring variables that measure current employment. The assumption underlying this IV is that (non-)human local endowments that have influenced the spatial distribution of population until the mid-19th century, are no longer important for productivity in a modern, 21st century economy, except through their influence on current employment. Historical population censuses are a relevant IV because the spatial distribution of population is strongly autocorrelated over time, for example due to path-dependency caused by self-reinforcing spillovers from agglomeration (Bleakley and Lin, 2012).

The second instrumental variable is the distance to the nearest railway station in 1870 (Koster, 2013). This instrument is correlated to current employment levels because the opening of railway stations during the 19th century drastically increased the area's accessibility and therefore triggered the formation of urban areas. Nowadays, however, these railway stations are only one of the many links in the infrastructure network. Hence, railway stations that have opened before 1870 are not expected to influence labor productivity today. In fact, almost half of these stations are no longer operational.

3. Data description

Our empirical model requires three key datasets. Firstly, we use wage data containing individual information for all employees in the Netherlands[14] on pretax wages and other financial rewards, hours worked, date of birth, sectoral classification of the employer (two-digit NACE), place of work at the ZIP code level, and job-type. This dataset is based on own calculations using non-public microdata from Statistics Netherlands (CBS): fiscal data (*Polisadministratie*), census data (*Sociaal Statistisch*

---

narrow and comprehensive set of concentric ring variables. For example, Rosenthal and Strange (2008) use cutoff values 8, 40, 80 and 160km (they use terrestrial miles as their unit of lenght which corresponds to cutoff values of 5, 25, 50 and 100 miles), whereas Di Addario and Patacchini (2008) choose 4, 8, 12 and 16km.

[13] Other instruments have been considered but are found to be inappropriate for our setting. For example, Combes et al. (2010) propose geology as a source of exogenous variation for population. It is however unlikely that geological variables would qualify as exogenous variables in the Netherlands, because the Dutch have turned many bodies of water into land during the past few centuries, see Figure A.3. Moreover, the Netherland's capital Amsterdam is largely built on 11 million wooden poles because its soil, consisting of fen and clay, is actually rather unsuitable for construction.

[14] This dataset does not contain information on the self-employed workers. The self-employed comprise between 10 to 15 % of the total Dutch working population.

Bestand), and firm data (*Algemeen Bedrijven Register*). Based on this information we construct a panel (2006-2014) with yearly observations for each individual.

The wage data do not only contain regular pre-tax wages, but also overtime payments, paid holidays, bonuses, thirteenth salaries and company cars. The reported number of hours worked consists of both regular and overtime hours. Dividing the sum of these annual financial rewards by the number of hours worked and deflating them with the consumer price index, provides an adequate approximation of the total hourly labor costs of each employee in a particular year. Due to limitations of the dataset, this calculation of total hourly labor costs is prone to measurement errors when a worker has not been employed for the full year at the same employer. For this reason we drop observations that are not based on a complete year of work at the same employer. We present a robustness analysis (see next section) using only regular hourly pre-tax wages. This alternative wage definition permits the inclusion of these dropped observations.

The data are further restricted as follows. We excluded all workers under 18 and above 65 years old. Also, jobs with less than 12 hours of work per week, the official definition by Statistics Netherlands for being employed, are excluded from the sample. In order to limit the influence of non-regular workers, we decided to drop the following job-types: owner-director, intern, temporary worker, and WSW-worker[15]. Jobs in agriculture and the fishing industry are excluded from the sample, because we do not expect substantial agglomeration benefits to occur in these sectors. Also the public sectors are excluded because it is improbable that these sectors meet our underlying assumption that employers are competitive and profit maximizing. Jobs provided by a firm with more than one establishment, could not be assigned geographically and had to be removed from the sample. Furthermore, for those people with more than one job during a year, we restrict the analysis to the job with the highest number of hours worked during that particular year. Outliers are defined as hourly wages below the legal minimum wage and above 20 times this minimum wage, and they are removed. After cleaning the data, over 1 million observations per year remain. Table A.1 summarizes the data remaining for estimation in the years 2006, 2010 and 2014. The left panel of Figure A.2 shows the mean hourly wages in euros per four-digit ZIP code.

The second key dataset contains information on the spatial distribution of both current[16] employment and historical population in the Netherlands and neighboring countries. We constructed this dataset by combining several data sources, which are listed in Table A.2. As can be seen from this table, our spatial unit of analysis, the four-digit ZIP code in the Netherlands, is rather small with an average area of only 8.86km^2. This high level of spatial detail allows us to examine the decay pattern of agglomeration economies on short as well as long distances.

[15] The WSW is a Dutch law aimed to foster the employment of persons with disabilities.

[16] Because we estimate area fixed-effects for the period 2006-2014, we take the middle year (2010) as our measure of current employment.

The right panel of Figure A.2 shows the spatial distribution of employment in the Netherlands. Comparing the left and right panel reveals that areas with relatively high employment densities also tend to have relatively high mean wages. This observation is of course merely descriptive and does not imply a causal relation between wages and urbanization. Figure A.3 gives an overview of the domestic and foreign spatial distribution of current employment (left panel) and historical population counts (right panel).

Using GIS-software, we construct concentric ring variables that measure the current employment levels and historical population counts within particular distance intervals. Firstly, we draw concentric rings around the geographic centroid of the ZIP codes and then calculate for each geographic unit in our sample which percentage of its area falls within the concentric ring. As previously discussed, we choose a total of seven concentric rings with a respective radius of 2½, 5, 10, 20, 40, 80, and 120 kilometer. Then we assume that, within geographic units, employment and population are homogeneously distributed across space, which enables us to approximate the level of current employment and historical population within each concentric ring. Finally, we first-difference the concentric ring variables in order to get total employment and population within particular distance intervals. The concentric ring variables are graphically presented in Figures A.4 and A.5.

Table A.3 contains a correlation matrix of the ring variables that measure employment, and those that measure historical population in brackets. This table shows that, although the ring variables are mutually correlated, this correlation is limited due to the increasing distance intervals. Hence, it appears that concerns regarding multicollinearity of the exogenous regressors will be limited.

The third and last key dataset contains coordinates of all railway stations that have been operational during the year 1870. This amounts to a total of 235 railway stations, of which 106 stations were no longer operational by the year 2006 (the first year of our wage data). This dataset is used to calculate for each ZIP code the straight line distance to the nearest railway station in 1870, see Figure A.6.

4. Results

Since the outcomes of the first stage regressions are not directly relevant for our paper[17], this section presents the outcomes of the second stage regressions only. Column (2) of Table 1 shows the results of the second stage IV estimates with the full set of concentric ring variables (Equation 5). From this we conclude that employment within 5km distance does not significantly affect wages. Between 5 and 10km we observe a relatively strong effect of employment on wages. More specifically, wages increase by 0.84% when employment between 5 to 10km distance increases by 100,000.[18] After 10km

[17] The parameters of the age variables are as one would expect: significantly positive for the linear variable and significantly negative for the quadratic variable.
[18] Note that employment within concentric rings is expressed as the total number of jobs in millions. A change of 100,000 jobs within 5 to 10km is equal to 0.8 standard deviation.

the net benefits of urbanization decrease, although the effect still remains significant at the 1%-level until at least 40km. After 80km we find no significant effect of employment on wages. A graphical representation of these results can be seen in Figure A.8.[19]

Table 1
Second-stage regressions (Eq. 5) – concentric ring variables

Dependent variable: first-stage area fixed-effects	OLS	IV						
	(1) All rings	(2) All rings	(3) Six rings	(4) Five rings	(5) Four rings	(6) Three rings	(7) Two rings	(8) One ring
Employment 0 to 2.5 km	0.1278** (0.0530)	0.0389 (0.0723)	0.0388 (0.0724)	0.0388 (0.0722)	0.0236 (0.0743)	0.0914 (0.0728)	-0.0132 (0.0697)	0.3339*** (0.0391)
Employment 2.5 to 5 km	-0.0043 (0.0430)	0.0079 (0.0538)	0.0080 (0.0539)	0.0077 (0.0536)	0.0212 (0.0553)	-0.0280 (0.0537)	0.2368*** (0.0404)	
Employment 5 to 10 km	0.0701*** (0.0171)	0.0837*** (0.0220)	0.0829*** (0.0219)	0.0856*** (0.0218)	0.0816*** (0.0220)	0.1532*** (0.0181)		
Employment 10 to 20 km	0.0219*** (0.0071)	0.0208** (0.0097)	0.0199** (0.0095)	0.0219** (0.0095)	0.0458*** (0.0077)			
Employment 20 to 40 km	0.0116*** (0.0026)	0.0123*** (0.0032)	0.0118*** (0.0031)	0.0163*** (0.0027)				
Employment 40 to 80 km	0.0041*** (0.0013)	0.0031** (0.0016)	0.0035** (0.0015)					
Employment 80 to 120 km	0.0003 (0.0016)	0.0010 (0.0018)						
P-value Kleibergen-Paap rk LM statistic		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-value Hansen J statistic		0.6258	0.6533	0.7504	0.4170	0.9613	0.1547	0.0349
Max VIF [Mean VIF]	4.76 [2.74]	10.15 [4.52]	10.14 [4.94]	10.02 [5.19]	9.99 [5.78]	9.58 [6.11]	5.01 [5.01]	
R^2	0.0542	0.0537	0.0537	0.0519	0.0430	0.0332	0.0195	0.0135

Notes: 3,722 observations. Robust standard errors are in parentheses. Employment is expressed as the total number of jobs in millions. * p < 0.1, ** p < 0.05, *** p < 0.01

When comparing column (1) and (2) of Table 1, we see that both the OLS and IV estimates are very similar, apart from the first concentric ring variable. This suggests that endogeneity is not a big concern at longer distances, whereas it does play a role at short distances (within 2½km). And

[19] We want to make a note on the Variance Inflation Factor (VIF) as reported in Table 1. Especially for the IV estimates, the VIF turns out to be rather high, and some would argue that this is too high given the frequently used rule-of-thumb that a VIF larger than 10 indicates serious multicollinearity. O'Brien (2007), however, rejects the common practice of relying blindly on these general rules, and suggests that researchers should also motivate whether the inclusion or exclusion of variables makes sense from a theoretical perspective. As there are good reasons to expect that agglomeration economies behave differently on various distances, especially on short distances, we prefer this extensive set of concentric ring variables. Nevertheless, we will show in the robustness analyses of this paper that by merging the first and second concentric ring (0 to 2½ and 2½ to 5km), the VIF values can be greatly reduced, though the results remain similar.

indeed, according to the endogeneity test[20], the data reject the null hypothesis that the first concentric ring variable can be treated as an exogenous regressor. Furthermore, the Kleibergen-Paap underidentification test and the Hansen J overidentification test confirm that the instruments are both relevant and valid. Hence, in what follows, we will primarily focus on the IV regressions.

It is revealing to see what happens when the model contains only a limited set of concentric ring variables and thus ignores employment at further distances. Therefore, Table 1 reports a total of seven IV regressions; each containing one additional concentric ring variable. By looking only at column (8), we may conclude that employment within 2½km affects wages positively and significantly. However, this estimate suffers from an omitted variable bias, even despite the fact that we have included instrumental variables. To see how this works, we must take a look at Table A.3. This table shows that each concentric ring variable correlates with adjacent concentric ring variables, which will be a source of omitted variable bias if we do not include the full set of concentric ring variables. The instrumental variables are inappropriate IV's to tackle this kind of endogeneity, since they are themselves also correlated with adjacent concentric rings and therefore also with the error term. And indeed, we can see from column (8) that the Hansen J test of overidentification rejects the null-hypothesis that the instruments are valid (i.e. uncorrelated with the error term).

The most efficient way to deal with this omitted variable bias is to add other concentric ring variables to the model. Column (7) shows that the coefficient of the first concentric ring becomes insignificant when an adjacent concentric ring is included. Hence, the previous model, containing only one concentric ring, was indeed confounded by an omitted variable bias. The Hansen J test now marginally fails to reject the null-hypothesis at a 10%-level. When adding a third concentric ring variable to the model, the coefficient of the second concentric ring variable becomes insignificant as well. When we continue this process of adding additional ring variables to the model, the results remain quite stable.

So far, the results indicate that agglomeration economies have a wide spatial scope. For a small open economy like the Netherlands, this raises questions about the role of foreign economic mass in shaping domestic wages. Furthermore, if foreign economic mass does in fact affect wages, then the results of Table 1 may be confounded by an omitted variable bias. Table 2 shows the result of this analysis. We do not find compelling evidence that foreign economic mass influences domestic wages. As can be seen from Table 2, all concentric ring variables measuring foreign employment are insignificant until 40km.[21] The other two ring variables, 40-80km and 80-120km, are significant, but have opposite signs, for which we cannot provide an obvious explanation. We conclude from these estimation results that foreign economic mass has, at best, only a limited influence on domestic wages.

[20] We ran a C or, equivalently, GMM distance test (see Baum et al., 2007).
[21] The first three foreign concentric rings are merged to form one variable measuring employment within 10km. We did this because foreign employment is not accurately measured on short distances. Nevertheless we did run the regressions with the full set of concentric ring variables, which lead to similar (insignificant) estimates.

Table 2
Second-stage regressions (Eq. 5) – domestic and foreign concentric ring variables

Dependent variable: first-stage area fixed-effects	IV		OLS	
	(1a) Domestic rings	(1b) Foreign rings	(2a) Domestic rings	(2b) Foreign rings
Employment 0 to 2.5 km	0.0445 (0.0729)		0.1320** (0.0531)	
Employment 2.5 to 5 km	0.0071 (0.0540)	0.2107 (0.2696)	-0.0050 (0.0432)	0.2734* (0.1461)
Employment 5 to 10 km	0.0837*** (0.0221)		0.0693*** (0.0171)	
Employment 10 to 20 km	0.0227** (0.0100)	0.0346 (0.1742)	0.0237*** (0.0073)	-0.0064 (0.0574)
Employment 20 to 40 km	0.0132*** (0.0035)	-0.0414 (0.0578)	0.0130*** (0.0027)	0.0065 (0.0169)
Employment 40 to 80 km	0.0049*** (0.0018)	0.0151** (0.0061)	0.0048*** (0.0014)	0.0042 (0.0031)
Employment 80 to 120 km	0.0017 (0.0023)	-0.0039* (0.0022)	0.0010 (0.0020)	-0.0014 (0.0015)
P-value Kleibergen-Paap rk LM statistic	0.0000			
P-value Hansen J statistic	0.6125			
Max VIF [Mean VIF]	37.09 [9.61]		6.42 [3.41]	
R^2	0.0544		0.0571	

Notes: 3,722 observations. Robust standard errors are in parentheses. Employment is expressed in millions of jobs (domestic) and employed people (foreign). * p < 0.1, ** p < 0.05, *** p < 0.01. The domestic and foreign concentric ring variables are estimated simultaneously using one single regression equation. However, in order to save space, the regression results of the domestic and foreign concentric ring variables are presented next to each other in two distinct columns (a and b). The first three foreign concentric rings are merged to form one concentric ring measuring employment within 10km. We did this because foreign employment is not accurately measured on short distances. Nevertheless we did run the regressions with the full set of concentric ring variables, which lead to similar (insignificant) estimates.

Robustness analysis

We start this robustness section with combining the first and second concentric ring variable into one variable that measures employment within 5km distance. We do this because the relatively high VIF-values reported in Table 1 might raise questions about the accuracy of the estimates, especially the estimates of the first and second ring variable. It can be seen in Table 3 column (1) that the action of combining these ring variables greatly reduces the VIF, although the results remain roughly similar compared to those in Table 1. The most notable difference between these results and those in Table 1 is the size of the standard errors on short distances. The percentage wage effect of employment within 5km is now more precisely estimated; see Figure A.9 for a graphical representation.

Table 3
Second-stage regressions (Eq. 5) – robustness analyses

Dependent variable: first-stage area fixed-effects					
			IV		
	(1) Combine 2 rings	(2) Exclude polders	(3) Nonlinearities	(4) Basic wages	(5) Area-year
Employment 0 to 5 km * highly urbanized			0.0067 (0.0323)		
Employment 0 to 5 km * urbanized			0.0062 (0.1209)		
Employment 0 to 5 km * little urbanized			-0.1327 (0.3416)		
Employment 0 to 5 km * not urbanized			-0.8928 (1.4011)		
Employment 0 to 2.5 km	0.0205 (0.0183)	0.0238 (0.0187)		-0.0485 (0.0555)	0.0591 (0.0719)
Employment 2.5 to 5 km				0.0318 (0.0397)	0.0093 (0.0541)
Employment 5 to 10 km	0.0815*** (0.0195)	0.0752*** (0.0200)	0.0767*** (0.0211)	0.0374** (0.0168)	0.0777*** (0.0219)
Employment 10 to 20 km	0.0210** (0.0097)	0.0188* (0.0099)	0.0196** (0.0099)	0.0174** (0.0076)	0.0236** (0.0092)
Employment 20 to 40 km	0.0123*** (0.0032)	0.0118*** (0.0032)	0.0127*** (0.0032)	0.0060** (0.0024)	0.0113*** (0.0029)
Employment 40 to 80 km	0.0031** (0.0016)	0.0039** (0.0016)	0.0031* (0.0016)	0.0037*** (0.0011)	0.0035** (0.0014)
Employment 80 to 120 km	0.0010 (0.0018)	0.0010 (0.0018)	0.0011 (0.0017)	0.0028** (0.0013)	0.0001 (0.0015)
P-value Kleibergen-Paap rk LM statistic	0.0000	0.0000	0.0000	0.0000	0.0000
P-value Hansen J statistic	0.6381	0.6079	0.5180	0.0462	0.0307
Max VIF [Mean VIF]	4.10 [2.81]	4.03 [2.81]	7.80 [4.58]	10.17 [4.53]	10.09 [3.04]
N	3,722	3,589	3,722	3,746	32,327
R^2	0.0536	0.0519	0.0538	0.0467	0.1230

Notes: Robust standard errors are in parentheses. Employment is expressed as the total number of jobs in millions. * p < 0.1, ** p < 0.05, *** p < 0.01

Secondly, we examine the possibility that the historical population IV is inappropriate when the sample contains polders (land reclaimed from bodies of water). Figure A.3 shows that a considerable part of the Netherlands has been dried since 1840. As a result, our sample contains 133 ZIP codes of which the geographic centroid location was not yet reclaimed from the water in 1840. This could lead to noisy estimates when using historical population within 5km as an instrument for current employment within 5km. We check for this possibility by dropping these 133 ZIP codes from the sample, and then re-estimate the second-stage regression. It can be seen from column (2) in Table 3 that the exclusion of polders does not substantially affect the estimates.

Furthermore, the relationship between urbanization and wages could potentially be nonlinear on rather short distances (within 5km), which would make the main results not applicable to all regions. For instance, non-urbanized areas might not meet a critical threshold to benefit from economies of agglomeration. In this case, we expect to see no effect of employment on wages for the non-urbanized areas, but a positive effect for urbanized areas. Another possibility is that, above a particular point that reflects the optimal level of employment, every additional unit of employment within 5km adds more to total congestion than to total gains. In this particular case, we expect to see a positive effect of employment on wages for the less urbanized regions but not for the most urbanized regions.

To check for these nonlinearities on short distances, we interact the first concentric ring variable (within 5km) with a set dummy-variables that indicate the level of urbanization. To this end, we divide the ZIP codes into subsamples, based on the level of employment within 5km. Because the level of employment within 5km is highly skewed to the right, we choose to split our sample into four unequally sized groups, using the 40^{th}, 70^{th} and 90^{th} percentile as cutoff values. For convenience, we label these subgroups as 'highly urbanized', 'urbanized', 'little urbanized' and 'not urbanized', see Figure A.7. The results in column (3) of Table 3 reject the existence of nonlinearities, as all interaction terms are insignificant. Hence, we conclude that there is no evidence of a nonlinear relation between urbanization and wages on short distances.

The fourth robustness check recalculates the hourly wages, excluding financial rewards other than the worker's regular pre-tax wage. Hence, with this recalculation, the dependent variable does no longer reflect total labor costs as it excludes thirteenth salaries, holiday entitlements, cash bonuses, etc. An advantage of this recalculation, however, is that we can retain those years in which a worker has been employed for less than the full year at the same employer, which increases the number of observations for the first-stage regression.[22] A comparison between the original and newly estimated area fixed-effects shows that this recalculation has substantial implications for the area fixed-effects estimates. First of all, both sets of area fixed-effects are not as strongly correlated as one may expect (the bivariate correlation is 0.63). Secondly, this recalculation of hourly wages reduces the dispersion of the area fixed-effects estimates substantially. To be specific, the variance of these area fixed-effects falls from 0.0045 to 0.0026.

Despite the fact that this alternative definition of hourly wages has a considerable impact on the estimates of the area fixed-effects in the first-stage, we find that the second-stage estimates are still consistent with the original estimates. When comparing column (4) of Table 3 and column (2) of Table 1, we find that the most notable change occurs at the 80-120km ring variable, which turns

[22] This increase in the number of first-stage observations is also the reason why the number of second-stage observations (number of estimated area fixed-effects) increases from 3,679 to 3,730.

significant at the 5%-level. The significance levels of the other parameters remain similar to the original estimates, though the point estimates are somewhat lower. These results support the original finding that urbanization at short distances does not affect wages.

The fifth and last robustness check involves a recalculation of the first-stage equation using an area-year interaction. The inclusion of an area-year interaction captures any variation over time of the area fixed-effects, which provides a more flexible estimator. The downside of including an area-year interaction, is that the number of fixed-effects to be estimated increases by around 30,000, making the calculations very time-consuming with standard regression algorithms.[23] Column (5) of Table 3 shows that this alternative specification yields estimates very similar to the base specification.

5. Discussion of the results and reconciliation with related literature

The previous section has shown that urbanization and wages are not significantly related on short distances (<5km), but strongly and positively related on longer distances (between 5 and 40-80km). Furthermore, the estimates are quite robust to a variety of alternative specifications. We argue that this spatial pattern can arise when the costs of urbanization (e.g. traffic congestion, pollution, and small lot sizes) decay more rapidly than the gains. In this case, the congestion externalities will offset the productivity gains on short distances, whereas the productivity gains will dominate the congestion costs on longer distances.

The estimated non-monotone spatial decay function helps to understand the incentives behind employment decentralization and urban sprawl. On the basis of our estimates, we conclude that agglomeration economies exhibit a relatively wide spatial scope. This implies that decentralized employment is, at least to some extent, able to exploit the gains from economic mass located at city centers. In fact, since we find no causal relation between urbanization and wages on short distances, we may even conclude that decentralization of jobs is economically efficient. The finding that urbanization on 5 to 10km distance has a relatively strong effect on wages suggests that especially the outskirts of cities benefit from economies of agglomeration (see Figure A.4). Our finding also fits the empirical observation of Glaeser and Kahn (2004) that per capita incomes in metropolitan areas are positively correlated to the share of jobs more than 5km from the Central Business District.

Although the results of this paper appear, at first sight, to contradict earlier studies, reconciliation with their results is straightforward. For instance, the apparent contradiction with earlier studies that find a strong and positive relation on short distances (e.g. Arzaghi and Henderson, 2008; Rosenthal and Strange, 2008; Di Addario and Patacchini, 2008; Ahlfeldt et al., 2015), can be attributed to differences in spatial detail of the datasets and the area under scope. To be specific, this study has used Dutch ZIP codes with a mean area of 9km^2 as the geographic unit of analysis, whereas

[23] For this robustness analysis, we used the STATA-module REGHDFE for estimating models with many fixed effects (Correia, 2014).

Rosenthal and Strange (2008) and Di Addario and Patacchini (2008) have used respectively US place-of-work PUMA's with a mean area of 6,522km^2 and Italian local labor markets with a mean area of about 889km^2. Evidently, a high level of spatial detail is necessary to disentangle urbanization wage-effects on short distances. If such a dataset is not available, then problems concerning collinearity between the concentric ring variables and measurement error will bias the results. In contrast, Arzaghi and Henderson (2008) and Ahlfeldt et al. (2015) did analyze a spatially detailed dataset, although their spatial scope is limited. The areas under study were, respectively, Manhattan and Berlin, which makes by definition the detection of agglomeration economies with a large spatial extent impossible. Hence, the crucial feature of this type of study is to analyze a nationwide wage panel with a high level of spatial detail.

The results of our study are consistent with the few studies that did have access to a nationwide spatially detailed dataset. In particular, Rosenthal and Strange (2003), who also use data at a ZIP code level, find that urbanization is not always positively related to the birth-rate of new establishments, especially at short distances. The authors also attribute this finding to the interplay of agglomeration economies and congestion costs. Duranton and Overman (2005), although they focus on localization rather than urbanization, also provide evidence that the location pattern of industries does not always decline monotonically and can be bumpy. Also, we want to touch upon the study of Koster (2013), who used a spatially detailed dataset to analyze the relation between urbanization and rents of commercial property in the Netherlands. Interestingly, Koster finds that commercial rents and urbanization are strongly related on short distances (<5km) and unrelated on longer distances. Although our results appear to be the exact opposite of the findings of Koster, it should be noted that the two studies estimate two distinct phenomena: the effect of urbanization on the marginal productivity of respectively land and labor.[24]

Finally, we draw two other conclusions from this study. Firstly, this paper shows that models including only employment at short distances are confounded by an omitted variable bias, see Table 1. Therefore, we conclude that studies on the relation between wages and urbanization should always take into account employment at various distances. Instrumental variables cannot adequately tackle these endogeneity issues, because employment at various distances is mutually correlated. Note that 'classical' estimates of the wage-density elasticity, which focus on the relation between urban density and wages at a local level (e.g. at the level of ZIP codes or municipalities), are likely to suffer from this omitted variable bias.

Secondly, we find no compelling evidence that domestic productivity levels are influenced by urbanization abroad, which indicates the existence of substantial border barriers. This finding fits

[24] The results of both studies can be reconciled if there are mechanisms at work that split the gains from urbanization unequally among landowners and workers. For instance, landowners may have more bargaining power when land is scarce, i.e. on short distances to urban areas, whereas workers may have strong bargaining power when land is abundant, i.e. on longer distances to urban areas. We leave it for future research to assess whether this hypothesis is true.

within a large strand of the literature dealing with border effects. Brakman et al. (2002), for instance, also find that market potential stemming from abroad does not affect wages in Germany. The good news is, however, that the estimates of earlier studies, which generally ignore foreign economic mass, are most likely not biased. Since we find a relatively large spatial extent of agglomeration economies, we conclude that productivity levels in Dutch border regions can be fostered by reducing border barriers.

6. Conclusions

The main methodological novelty of this paper is to estimate the spatial scope of agglomeration economies by analyzing a nationwide and spatially rich dataset on individual wages. This nationwide high level of spatial detail, which is absent in similar studies, enables us to analyze the behavior of agglomeration economies on long as well as on short distances. We find an insignificant relation between wages and urbanization on short distances (<5km) and a strong positive relation on medium distances (5-10km). This positive effect attenuates across geographic space and becomes insignificant after 40-80km. This spatial pattern can be explained as the net effect of two externalities with opposite effects: agglomeration economies and congestion costs.

This interplay between agglomeration economies and urban congestion costs at various distances helps us to understand the driving forces behind employment decentralization and urban sprawl. Our estimates indicate that those regions with relatively high employment levels between 5 and 10km coincide with the outskirts of cities (see Figure A.4), which makes them an attractive location for both firms and people. Likewise, we find that, on short distances, urban congestion largely offsets the gains from urbanization, which can explain why firms are pushed away from city centers.

Finally, this paper contributes to strands of the literature dealing with border effects, by showing that foreign economic mass does not affect wages in the Netherlands. Although the Netherlands is a small and open economy, this lack of cross-border diffusion of agglomeration economies suggests that national borders still hinder economic interaction. Hence, policymakers can foster productivity levels in border regions by reducing barriers to cross-border economic interaction.

References

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. Econometrica, 83(6), 2127-2189.

Arzaghi, M., & Henderson, J. V. (2008). Networking off madison avenue. The Review of Economic Studies, 75(4), 1011-1038.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007). Enhanced routines for instrumental variables/GMM estimation and testing. Stata Journal, 7(4), 465-506.

Bleakley, H., & Lin, J. (2012). Portage and path dependence. The quarterly journal of economics, 127(2), 587.

Brakman, S., Garretsen, H., & Schramm, M. (2002). The final frontier? Border effects and German regional wages (No. 197). Hamburg Institute of International Economics (HWWA).

Combes, P. P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters!. Journal of Urban Economics, 63(2), 723-742.

Combes, P. P., Duranton, G., Gobillon, L., & Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In Agglomeration Economics (pp. 15-66). University of Chicago Press.

Correia, S. (2015). REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components.

De la Roca, J., & Puga, D. (2017). Learning byWorking in Big Cities. The Review of Economic Studies, 84(1), 106-142.

Di Addario, S., & Patacchini, E. (2008). Wages and the city. Evidence from Italy. Labour Economics, 15(5), 1040-1061.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. The Review of Economic Studies, 72(4), 1077-1106.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. Handbook of regional and urban economics, 4, 2063-2117.

Glaeser, E. L., & Kahn, M. E. (2001). Decentralized employment and the transformation of the American city (No. w8117). National Bureau of Economic Research.

Glaeser, E. L., & Kahn, M. E. (2004). Sprawl and urban growth. Handbook of regional and urban economics, 4, 2481-2527.

Gobillon, L. (2004). The estimation of cluster effects in linear panel models. Processed, ined.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. Journal of international economics, 67(1), 1-24.

Harris, C. D. (1954). The, Market as a Factor in the Localization of Industry in the United States. Annals of the association of American geographers, 44(4), 315-348.

Koster, H. R. (2013). Rocketing rents The magnitude and attenuation of agglomeration economies in the commercial property market.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. The review of Economics and Statistics, 334-338.

Nabielek, K., Boschman, S., Harbers, A., Piek, M., & Vlonk, A. (2012). Stedelijke verdichting: een ruimtelijke verkenning van binnenstedelijk wonen en werken. The Hague, Netherlands: PBL Netherlands Environmental Assessment Agency.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41(5), 673-690.

Phibbs, C. S., & Luft, H. S. (1995). Correlation of travel time on roads versus straight line distance. Medical Care Research and Review, 52(4), 532-542.

Rice, P., Venables, A. J., & Patacchini, E. (2006). Spatial determinants of productivity: analysis for the regions of Great Britain. Regional Science and Urban Economics, 36(6), 727-752.

Rosenthal, S. S., & Strange, W. C. (2003). Geography, industrial organization, and agglomeration. review of Economics and Statistics, 85(2), 377-393.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. Handbook of regional and urban economics, 4, 2119-2171.

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. Journal of Urban Economics, 64(2), 373-389.

Appendix

Table A.1
Summary statistics of the longitudinal wage data

	2006	**2010**	**2014**
Number of workers	1,456,067	1,192,499	1,115,038
Hourly wages in euro's (price level 2006)			
Mean (standard deviation)	20.6 (11.5)	21.6 (12.4)	21.4 (12.9)
Median	17.8	18.4	18.0
1st percentile	8.3	8.6	8.4
99th percentile	67.2	71.8	74.4
Age			
Mean (standard deviation)	40.0 (10.7)	40.9 (10.9)	41.8 (11.1)
Median	39.3	40.8	42.2
1st percentile	20.5	20.6	21
99th percentile	61.4	62.7	63.25
Industrial composition (in percentages)			
Manufacturing	23.2	22.1	22.7
Construction	11.3	10.4	8.1
Logistics	7.2	7.6	7.6
Wholesale	15.1	15.4	16.2
Retail	7.0	7.4	6.8
Consumer services	3.2	3.5	3.5
Hospitality industry	4.2	4.7	5.0
ICT	4.7	5.4	6.3
Financial services	3.6	3.0	3.0
Business services	20.4	20.7	21.0

Table A.2
Data sources for constructing the concentric ring variables

	Geographic unit	Number of geographic units	Year	Mean area in km^2	Data source
Current employment					
Netherlands	Four-digit ZIP code	3950	2010	8.86	LISA
Germany*	Municipality	1445	2010	57.00	Statistik der Bundesagentur für Arbeit
Belgium	Municipality	589	2010	52.13	Vlaamse Arbeidsrekening
Historical population					
Netherlands	Municipality	1232	1840	26.38	CBS (Volkstellingen)
Germany*	Municipality	1445	1867	57.00	See on this page below
Belgium	Municipality	589	1846	52.13	Statistics Belgium

* Our dataset does not contain all German municipalities, but only those that belong to the Bundesländer Lower Saxony, Bremen and North Rhine-Westphalia. This is sufficient for our analysis.

Data sources of the German historical population counts

Statistisches Bureau Preussen (1874). Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: Nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871 (11): Die Gemeinden und Gutsbezirke der Rheinprovinz und ihrer Bevölkerung: nebst einem Anhange, betreffend die Hohenzollerschen Lande. Berlin, Verlag des königlichen Statistischen Bureaus.

Statistisches Bureau Preussen (1873). Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: Nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871 (8): Die Gemeinden und Gutsbezirke der Provinz Hannover. Berlin, Verlag des königlichen Statistischen Bureaus.

Statistisches Bureau Preussen (1874). Die Gemeinden und Gutsbezirke des Preussischen Staates und ihre Bevölkerung: Nach den Urmaterialien der allgemeinen Volkszählung vom 1. December 1871 (9): Die Gemeinden und Gutsbezirke der Provinz Westfalen und ihrer Bevölkerung: nebst einem Anhange, betreffend die Fürstenthümer Waldeck und Pyrmont. Berlin, Verlag des königlichen Statistischen Bureaus.

Statistisches Bureau (1871). Statistische Nachrichten über das Grossherzogtum Oldenburg (12): Ergebnisse der Volkszählung vom 3. December 1867. Oldenburg, Gerhard Stalling.

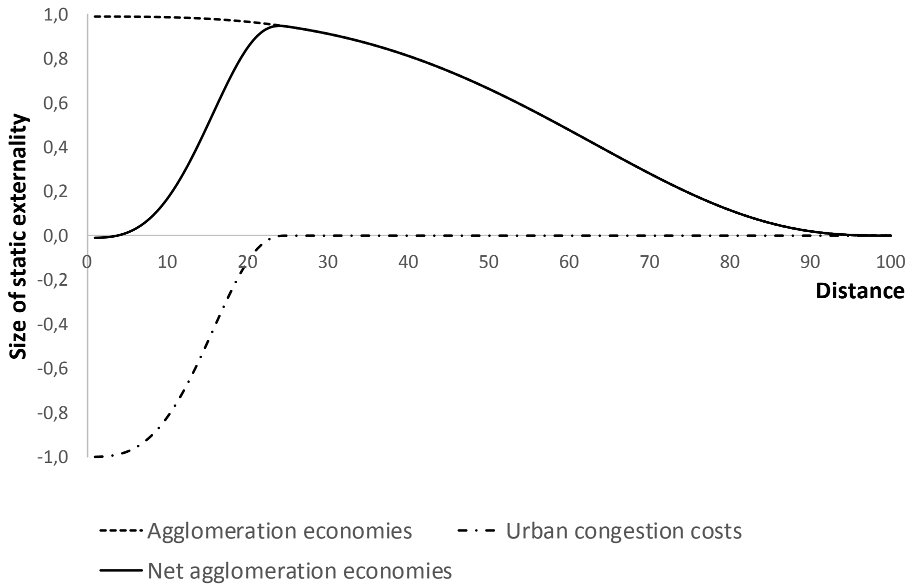
Kraus, Antje (1980). In Köllmann, Wolfgang (Ed.), Quellen zur Bevölkerungs-, Sozial- und Wirtschaftsstatistik Deutschlands: 1815-1875, Bd. 1. Quellen zur Bevölkerungsstatistik Deutschlands 1815-1875, p. 329-335. Boppard am Rhein, Boldt.

Table A.3
Correlation matrix of domestic concentric ring variables

	0-2½km	2½-5km	5-10km	10-20km	20-40km	40-80km	80-120km
0-2½km	1.0000 [1.0000]						
2½-5km	0.8083 [0.6085]	1.0000 [1.0000]					
5-10km	0.4837 [0.0778]	0.7096 [0.3450]	1.0000 [1.0000]				
10-20km	0.2918 [0.1366]	0.3928 [0.1838]	0.6320 [0.3567]	1.0000 [1.0000]			
20-40km	0.1969 [0.1069]	0.2501 [0.1484]	0.3582 [0.2424]	0.5890 [0.4461]	1.0000 [1.0000]		
40-80km	0.2198 [0.0919]	0.2809 [0.1400]	0.3853 [0.2156]	0.5173 [0.3718]	0.7014 [0.6143]	1.0000 [1.0000]	
80-120km	-0.0571 [-0.0472]	-0.0762 [-0.0567]	-0.1202 [-0.0877]	-0.1420 [-0.1100]	-0.0634 [-0.0096]	0.1867 [0.2524]	1.0000 [1.0000]

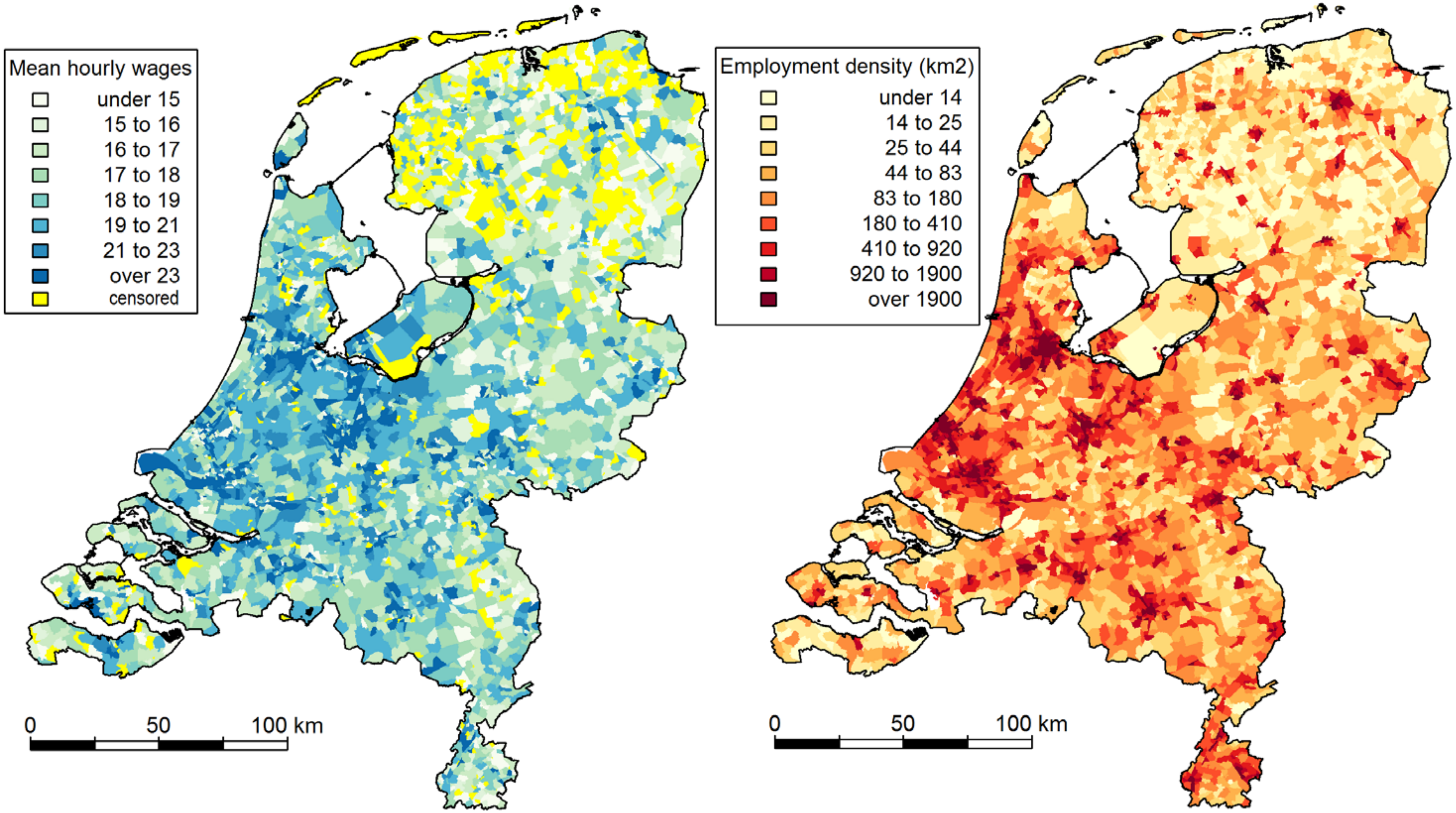
The correlation coefficients of concentric ring variables measuring historical population are in brackets.

Figure A.1
Graphical representation of the net effect of two monotone distance decay functions



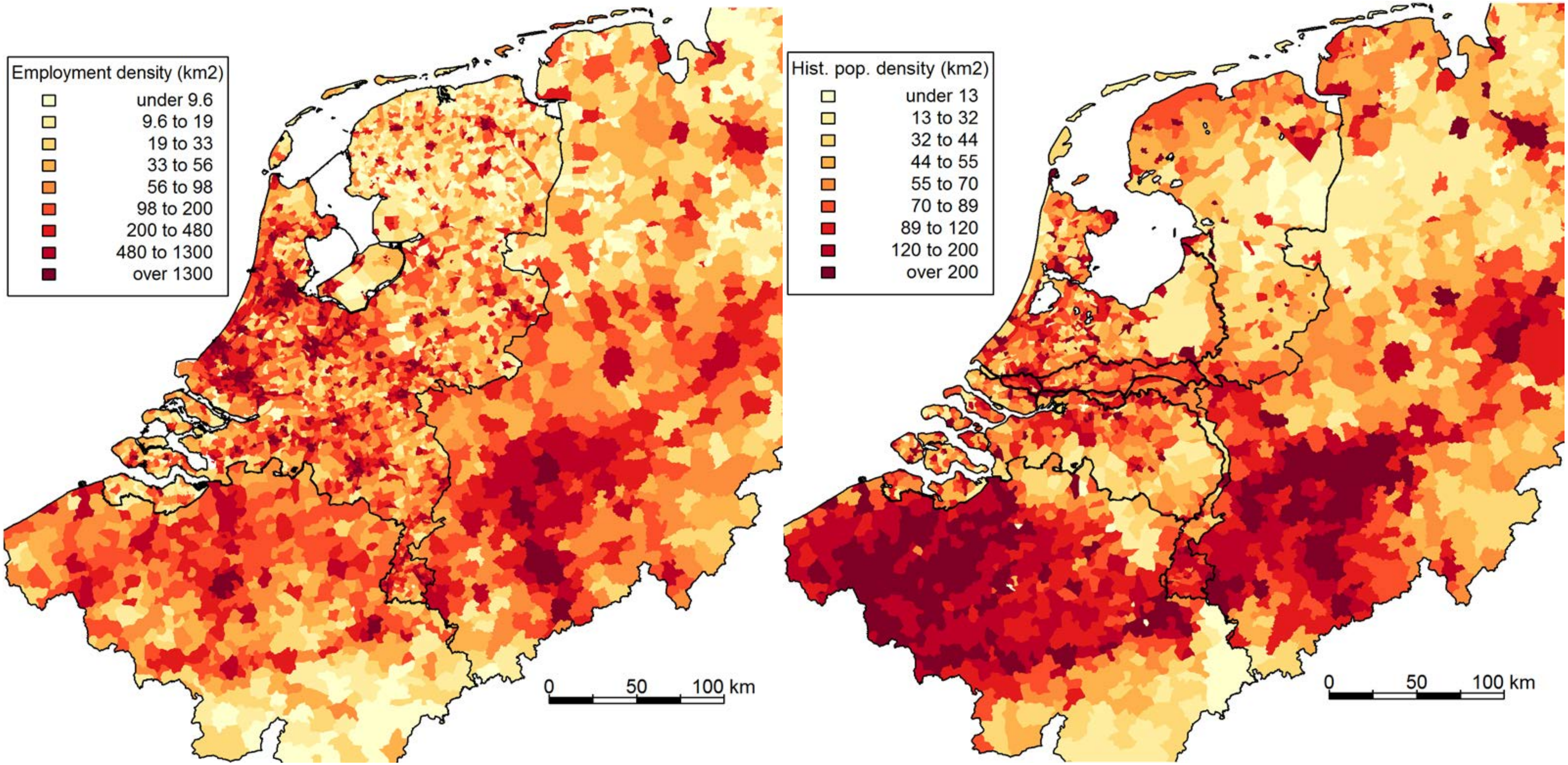
Note: the purpose of this figure is to show how a non-monotonically declining distance decay pattern might be the net outcome of two monotonically declining distance decay functions with opposite effects. Of course many other functional forms are possible.

Figure A.2
Mean hourly wages in euros per four-digit ZIP code (left panel) and employment density (right panel)



In the left figure, approximately 17% of the four-digit ZIP codes is censored because these ZIP codes contain less than 10 workers in our dataset.

Figure A.3
Current employment density in the Netherlands, Germany and Belgium (left panel) and historical population density (right panel)



The figures above are purely descriptive and should be interpreted with caution. For example, the legend displays 9 classes based on quantiles, which may not accurately reflect the fact that the densities are actually highly skewed to the right. Also, the map of the Netherlands in the right panel is based on the 1840 division of municipalities (source: Dr. O.W.A. Boonstra (2007): NLGis shapefiles. DANS. http://dx.doi.org/10.17026/dans-xb9-t677), whereas the German and Belgian maps are based on the current municipal division. As a consequence, the Dutch municipalities are on average only half the size of the Belgian and German municipalities, which affects the graphical representation. It is however true that Belgium was more densely populated than the Netherlands during the 19^{th} century. According to population censuses, the Netherlands' total population transcended Belgium's total population around 1930.

Figure A.4
Graphical representation of the concentric ring variables measuring domestic employment

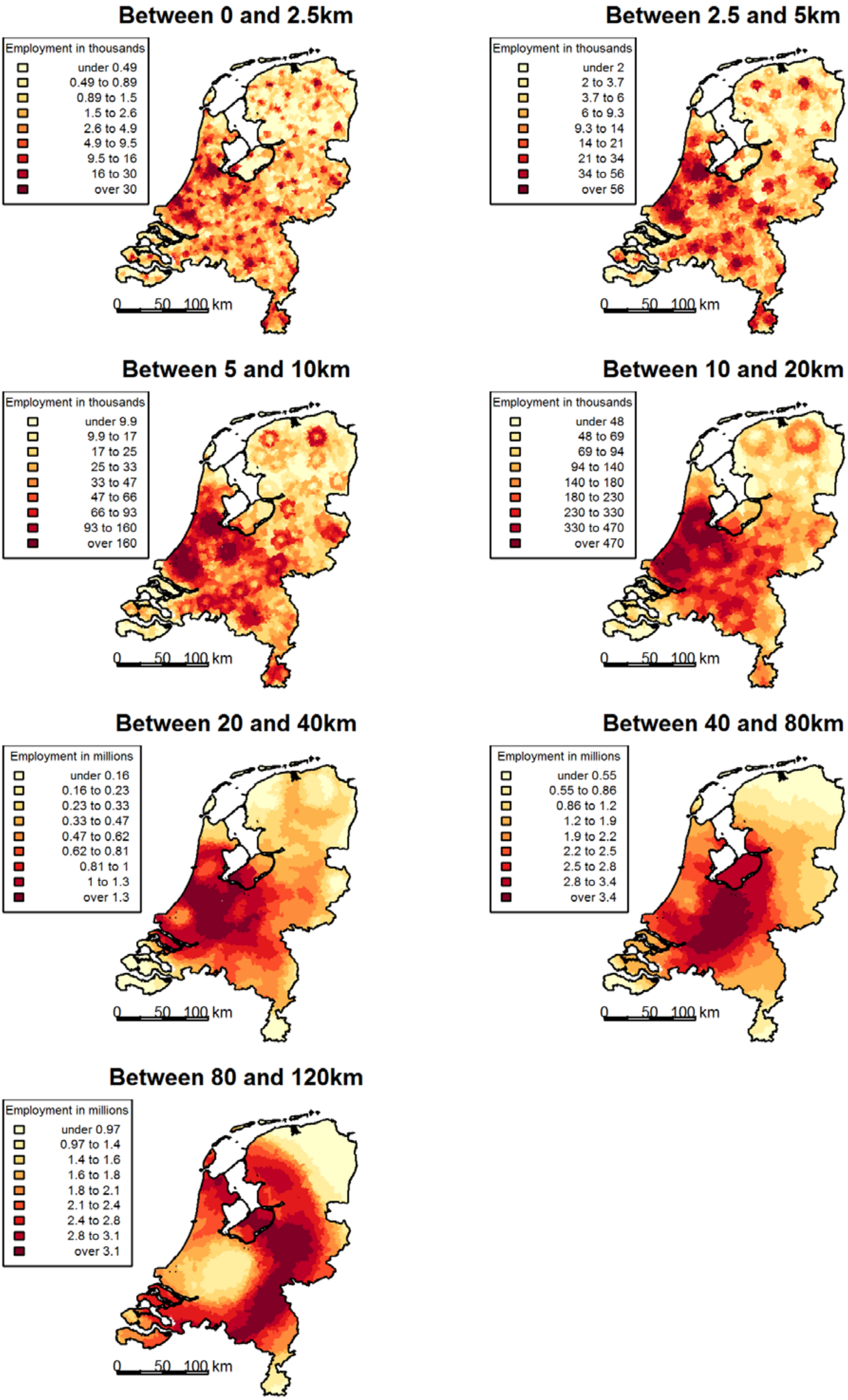


Figure A.5
Graphical representation of the concentric ring variables measuring foreign employment

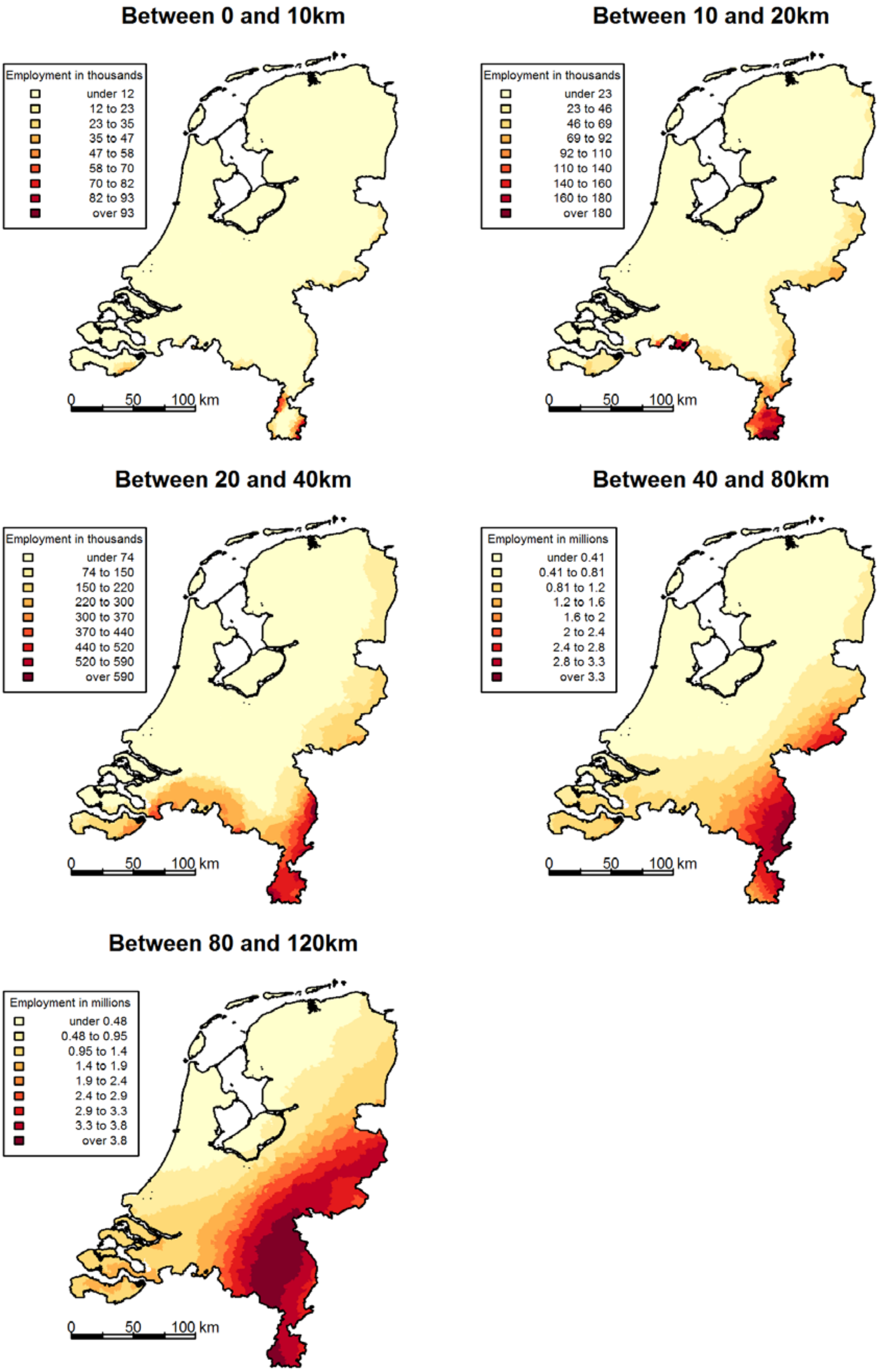


Figure A.6
Straight line distance to nearest railway station in 1870

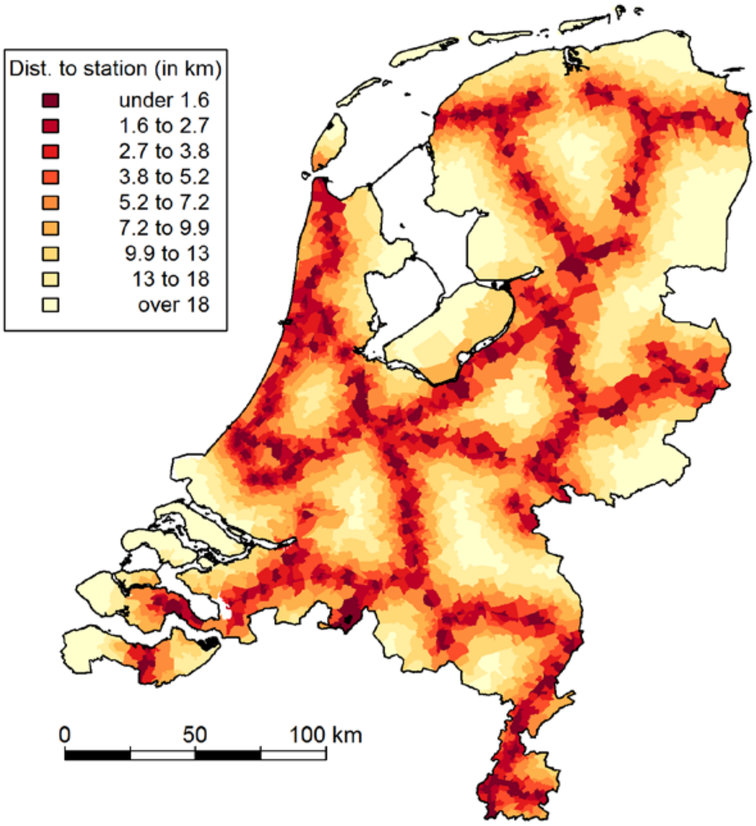


Figure A.7
ZIP codes divided into four subgroups based on employment within 5km

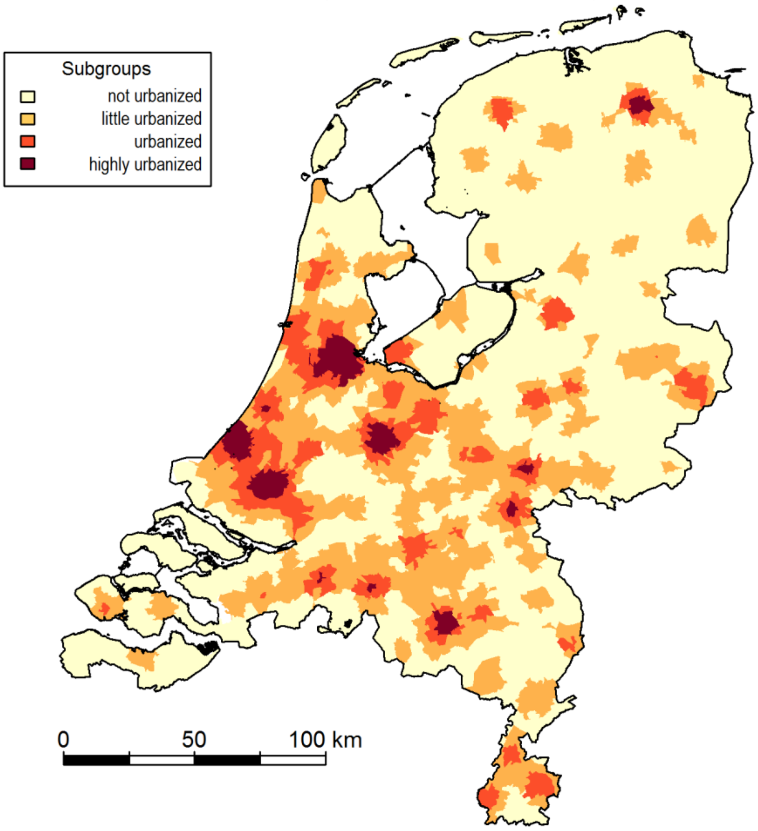
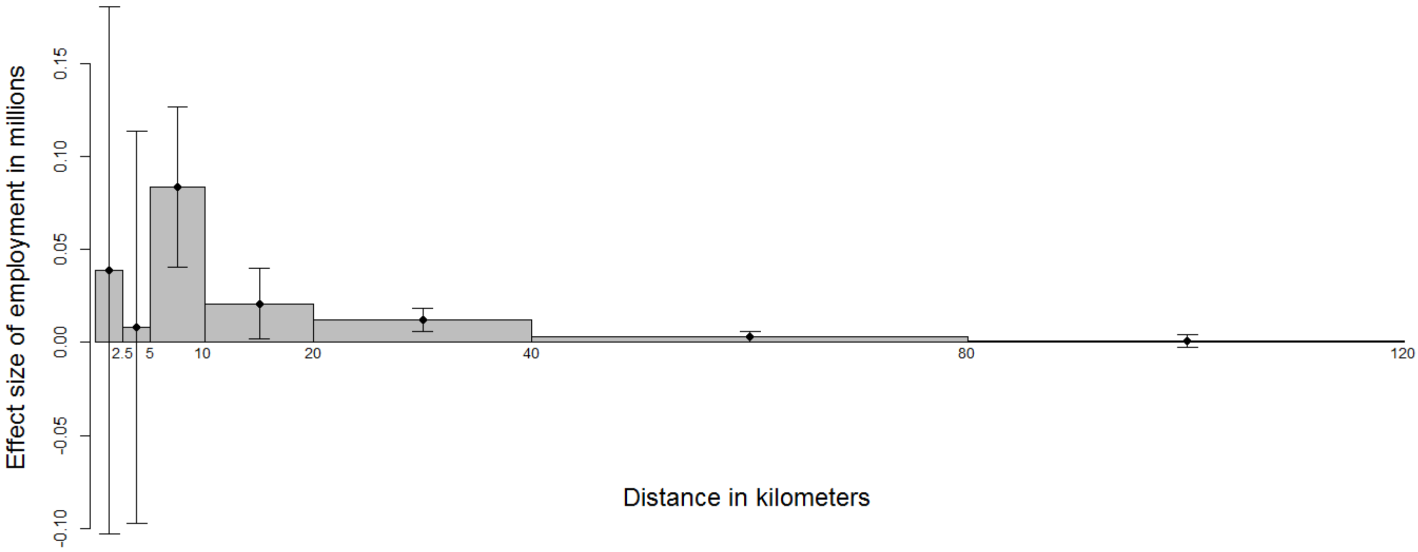
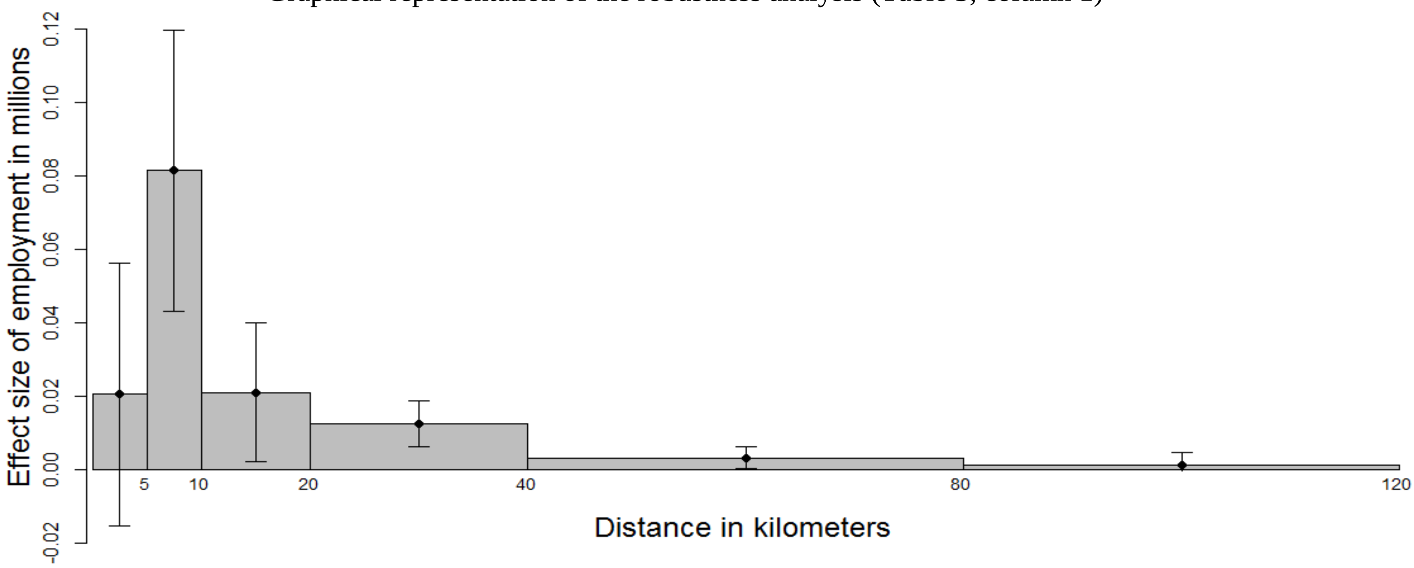


Figure A.8
Graphical representation of the main results (Table 1, column 2)



The height of the bars represents the point estimates of the concentric ring's parameter (see Table 1, column 2). The bar's width represents the corresponding distance intervals. The error bars show the 95% confidence interval.

Figure A.9
Graphical representation of the robustness analysis (Table 3, column 1)



The height of the bars represents the point estimates of the concentric ring's parameter (see Table 3, column 1). The bar's width represents the corresponding distance intervals. The error bars show the 95% confidence interval.

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (088) 984 60 00

February 2017 | ISBN 978-90-5833-761-0