DALARNA UNIVERSITY

Working papers in transport, tourism, information technology and microdata analysis

How does employment density influence individuals' wages?

A micro data approach.

Författare 1: Johan Håkansson

Författare 2: Gunnar Isacsson

Editor: Hasan Fleyeh

Nr: 2013:26

Working papers in transport, tourism, information technology and microdata analysis
ISSN: 1650-5581

© Authors

Title: How does employment density influence individuals' wages? A micro data approach.

Authors: Johan Håkansson[1], Gunnar Isacsson[2]

Abstract: We estimate the effect of employment density on wages in Sweden in a large geocoded data set on individuals and workplaces. Employment density is measured in four circular zones around each individual's place of living. The data contains a rich set of control variables that we use in an instrumental variables framework. Results show a relatively strong but rather local positive effect of employment density on wages. Beyond 5 kilometers the effect becomes negative. This might indicate that the effect of agglomeration economies falls faster with distance than the effects of congestion.

[1] Dalarna University
[2] The Swedish National Road and Transport Research Institute (VTI)

1. Introduction

A number of theoretical models has suggested different mechanisms behind a positive relationship between economic mass and wages or productivity (see e.g. Fujita et al. 1999 or Duranton and Puga, 2004). Although some of the theories in this field have received a more solid theoretical underpinning lately, the basic ideas seem to date back to Marshall (1890) who suggested that: (i) producers may gain from locating in a region with good access to suppliers and final demand so that they can save on transportation costs, (ii) in economically dense areas human capital externalities flow easily, (iii) a large labor market may be beneficial to both workers and employers for. It is widely believed that such 'agglomeration economies' are important to understand the spatial structure of modern economies.

A growing body of empirical research also seems to support the idea that agglomeration economies are relevant to understand spatial variation in wages or productivity (e.g. Ciccone and Hall, 1996, Glaeser, 2000, Malmberg et al., 2000, Ciccone, 2002, Hanson, 2004, Eriksson et al., 2008, Åslund et al., 2010, Combes et al., 2010, Potter and Watts, 2011). Here agglomeration economies is sometimes assessed by using employment per square km of land area as in Ciccone and Hall (1996) and Ciccone (2002). We follow previous research and refer to this as employment density in the rest of this paper. In other instances, agglomeration economies refers to a region's market potential (cf. Harris, 1954, and Hanson, 2004) where total income in surrounding regions are weighted by some measure of distance to the region and then summed to an index. In other cases, agglomeration economies is measured by accessibility to some resource, which is similar to the market potential but may refer to a weighted sum of, say, jobs (Eliasson et al., 2003). In this paper we will use employment density to measure agglomeration economies.

A fundamental problem in most empirical research on the relationship between agglomeration economies and wages is the endogeneity of variables intended to measure agglomeration economies in a wage equation (see Combes et al. 2011, for a comprehensive discussion of this issue). Thus, it would be desirable to use an empirical strategy based on an instrumental variable design of some kind to properly identify agglomeration economies. It is, however, difficult to find valid instrumental variables in this context. This may explain why only a few studies in this field are based on an instrumental variables framework. In a pioneering and widely cited study, Ciccone and Hall (1996) use long lags of population density as an instrument when estimating the relationship between productivity and employment density across U.S. states. Ciccone (2002) analyses the effect of employment density on value added per worker in a set of European countries at the so-called NUTS3 regions. He uses land area of each NUTS3 region as an instrumental variable for employment density. A more recent example is Combes et al. (2010) who estimate the relationship between employment density and wages, and between market potential and wages, using aggregated data on French municipalities. They use two different instrumental variables: long lags of population density by municipality and a set of geological variables like water capacity of the subsoil. Rosenthal and Strange (2008) also use geological variables as an instrumental variable for population.

Furthermore, most previous empirical studies in this field use spatially aggregated data for predefined spatial units of measurement. This implies a number of potential problems for the analysis and interpretation of the results. First of all, spatial units may vary substantially in terms of geographical size and shape which suggests that they might suffer from the modifiable area unit problem (Openshaw and Taylor, 1979). Secondly, this implies that most studies are based on relatively few units of observations as compared to much work where the observational unit is an individual or a household. This is specifically relevant to studies that use estimators based on instrumental variables since such estimators are asymptotically motivated (see e.g. Angrist and Pischke, 2009, pp. 205-216 for a discussion). Thirdly, differences between spatial units may be a result of the sorting of heterogeneous micro units like workplaces, individuals or households across space. This sorting may be more difficult to control for when using aggregated data (Ottaviano, 2011). One of the very few previous studies relevant to agglomeration economies based on an instrumental variable estimator and data on micro units is Åslund et al. (2010) who analyze how important proximity to jobs is for recent immigrants' transitions from unemployment to employment in Sweden. Their instrument is based on Swedish refugee dispersal policy so they effectively deal with the problem of individuals' spatial sorting. Their outcome variable is thus different from the previously cited papers on agglomeration economies that focus on wages/productivity. In addition, their analysis is restricted to a small part of the population. Thus, it is unclear how the estimated parameter relates to the corresponding parameter in the total population.

The purpose of the present paper is to estimate the effect of employment density on wages in Sweden using geocoded micro data on all individuals and all workplaces. In so doing, we use coordinates for place of residence and place of work of each individual coupled with a rich set of individual and workplace characteristics. The analyses are based on the framework presented by Ciccone (2002). However, we modify his productivity equation framework to a corresponding wage equation framework. In addition, our data allows us to construct spatial units on which to measure employment density that may be better representations of each individual's local labor market than data based on administrative geographic units like a state in the US or counties or municipalities. We also follow Ciccone (2002) and use the amount of land area as an instrumental variable for employment density.

More specifically, the data contains the entire Swedish population between ages 20 and 64 where the location of each individual's place of residence and the location of each employed individual's workplace is geocoded to at least a precision of one squared km. We use this information to measure employment density within each of four different circles around each individual's place of residence: (i) the employment density within 5 kilometers, (ii) the employment density between 5 kilometers and 25 kilometers, (iii) the employment density between 25 kilometers and 50 kilometers, (iv) the employment density between 50 kilometers and 100 kilometers. Density refers to the number of jobs per squared kilometer of land area within distance band.

The rest of this paper is organized as follows. Section 2 presents a theoretical framework for the analysis whereas section 3 outlines the empirical models in more detail. Section 4 presents the data and section 5 contains the results and includes a sensitivity analysis. The paper ends with a concluding discussion in section 6.

2. Theoretical framework

We use the same production function proposed by Ciccone (2002) to derive a demand curve for labor. In a static labor demand framework, the demand for labor simply reflects the value of the marginal product and assuming that firms are price takers in the labor market it is easy to show that profit maximization implies that the value of the marginal product for labor should equal the wage rate. Ciccone (2002) proposes a production function where there are constant returns to scale to labor (human capital), capital and land. He then adds the density of output to the production function to capture the effect of agglomeration economies. The density of output is simply output per unit of land in the geographic unit of measurement (the place of location). Thus, the production function may be written as

$$Q_{pr} = \Omega_{pr}\left[{N_{pr}}^{\beta}{K_{pr}}^{1-\beta}\right]^{\alpha}{A_{pr}}^{1-\alpha}\left(\frac{Q_{pr}}{A_{pr}}\right)^{(\lambda-1)/\lambda} \tag{1}$$

where Q_{pr} is output in place p (p = 1, 2, …, P), region r (r=1, 2, …, R); Ω_{pr} is an index of total factor productivity; N_{pr} is labor; K_{pr} is physical capital; A_{pr} is the sum of total land area; α measures returns to capital and labor and β is a parameter related to the shares of capital and labor in total income such that $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$; λ is the central parameter of interest here, since it is supposed to capture the effects of agglomeration economies and if it equals one then there are no effect of agglomeration economies on output. After rearranging this equation and assuming that the rental price of physical is the same within each region (we make this assumption since we lack information on physical capital stocks) we arrive at the following expression for the natural logarithm (*ln*) of the wage

$$lnw_{pr} = C + f_r + \theta_1\left[lnN_{pr} - lnA_{pr}\right] + \theta_2\left[ln\Omega_{pr}\right] \tag{2}$$

where C is a constant that involves the parameters of the production function; f_r is a region specific fixed effect that includes the rental price of physical capital;

$$\theta_1 = \frac{\alpha\lambda - 1}{1 - \alpha\lambda(1-\beta)}$$

and

$$\theta_2 = \frac{\lambda}{1 - \alpha\lambda(1-\beta)}$$

where the former is the parameter we aim to estimate in our empirical model. A positive value of θ_1 thus tells us that $\alpha\lambda > 1$. As noted by Ciccone (2002), α indicates the returns to scale to capital and labor per unit of land, thus a value of α below 1 suggests that there are decreasing returns, which in turn can be interpreted as a congestion effect. So if wages increase with higher employment density, this reflects that the effect of agglomeration economies on output is sufficiently large to dominate the congestion effect. A negative value of θ_1 implies that $\alpha\lambda < 1$ thus suggesting the opposite.[3] A useful case to consider is, of course, when there is no effect of agglomeration economies on output, this corresponds to the situation where $\lambda = 1$. In this case, θ_1 will be negative for all α below 1 and wages will decrease the higher is the employment density.

The last term in equation 2 reflects a set of factors related to total factor productivity that may, *inter alia*, be related to regional observable and unobservable characteristics; e.g. composition of industries and sectors in the region, climatic characteristics, access to natural resources and local policies including municipality tax rates. Furthermore, since workers differ in terms of observable and unobservable factors worker spatial self-selection (sorting) will also cause regional wage differentials (Glaeser & Maré, 2001, Combes et al. 2011, Venables, 2011). In sum, to empirically identify agglomeration economies we need to control for a number of factors and to address remaining endogeneity of employment density in the wage equation, we need a valid instrument.

3. Empirical Models

Empirical investigations of the effect of employment density on wages/productivity are usually based on data on an administrative or functional spatial unit like a state or a metropolitan

[3] Here we rule out the possibility that $\lambda > \alpha^{-1}(1-\beta)^{-1}$; i.e. the parameter intended to capture agglomeration economies is not larger than the inverse of capital's share in GDP.

statistical area in the U.S.A (see for example Ciccone and Hall, 1996, Glaeser and Maré, 2001 or Fallah, et al. 2011) or a NUTS3 region in Europe (Ciccone, 2002) or a municipality (Combes et al., 2011, who use French data) or local labor markets (Eriksson, et al., 2008, who use Swedish data). Sometimes the empirical framework also deals with spillovers between such units (see for example Combes, et al. 2011). This implies that the estimated relationships are based on geographical units that differ substantially in terms of size and shape. This may introduce a kind of heterogeneity in the estimated effect that few studies address and that might be more or less difficult to deal with.

We believe that the geocoded data we use has at least two advantages: we can construct our spatial unit of measurement freely and the size of each circle is the same for every individual. This implies that we reduce potential problems related to the "modifiable area unit problem" (MAUP). We construct four circles around each individual's place of residence and measure employment density within each of the four different circles: (i) the employment density between 0 and 5 kilometers, (ii) the employment density between 5 kilometers and 25 kilometers, (iii) the employment density between 25 kilometers and 50 kilometers, (iv) the employment density between 50 kilometers and 100 kilometers. Employment density refers to the number of jobs per squared kilometer of land area within each distance band.

The motivation for using more than one circle is to allow for a declining wage effect of jobs located further away from the individual rather than allowing the effect to be non-zero for jobs located closer than 5 kilometers and then, a priori, constraining the effect to be zero for jobs located further away than 5 kilometers. This approach also allows us to investigate the distance-decay aspect of the relationship between wages and employment density. Finally, by using more than one circle we also deal with spillovers between locations.

The motivation for the size of each of the 4 circles we use is related to the spatial extension of Swedish cities. The smallest circle may be considered to cover a larger part of the city center including the largest cities in Sweden. The next-to-smallest and next-to-largest circles cover the distances of most commuting trips in Sweden. We considered that agglomeration economies further away than 50 kilometers was unlikely to have any effects on wages but included the largest distance band (50-100 kilometers) to investigate whether this was actually the case. In addition, one-way commuting above 50 kilometers are not unusual in the larger metropolitan areas of Sweden.

Thus, our first empirical model is presented in equation 3 and it provides us with our first estimate of the agglomeration effect on wages

$$lnw_i = \alpha_0 + \alpha_1 lnE_{1i} + \alpha_2 lnE_{2i} + \alpha_3 lnE_{3i} + \alpha_4 lnE_{4i} + \varepsilon_{1i} \qquad (3)$$

where w_i is individual i's (i = 1, 2, …, N) wage; E_1 is the employment density in the first circle, E_2 is the employment density between the first and second circle, E_3 is the employment density between the second and third circles and E_4 is the employment density between the third and fourth circles; and ε_{1i} is the error term of the equation; α_p (p=1, 2, 3, 4) are the key parameters to be estimated in this paper. Agglomeration economies suggest that these four parameters should be positive and if distance-decay is relevant to the relationship we expect that $\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4$.

Clearly, individuals differ in terms of observable characteristics which could affect the individual's productivity and labor supply. Furthermore, since individuals choose where to live it is obvious that spatial sorting on observable characteristics may bias our previous estimator of the agglomeration effect. Hence we introduce a set of control variables for such characteristics in the next model.

$$lnw_i = \alpha_0 + \sum_{p=1}^{4} \alpha_p lnE_{pi} + \sum_{k=1}^{76} \gamma_k X_{ki} + \varepsilon_{2i} \qquad (4)$$

where X_{ki} (k=1, 2, …, 76) are a set of control variables including: gender, age, age-squared, dummy variables for education, dummy variables for industry and sector and separate variables for the number of children in different age groups living at home. The other variables are the same as in equation 3.

Since local conditions vary we add a set of municipality fixed effects to the previous equation and estimate equation 5 to deal with time invariant local variables. This means that "local conditions" pertain to factors that are unique to each municipality. Since the municipality level is the most local level of government in Sweden, the municipality fixed effect include the municipality tax rate; the quality of certain types of publicly funded services like schools and care of elderly people; and the local practice of laws and regulations. By using municipality fixed effects we also reduce the variation in unobserved geographical heterogeneity: for instance, proximity to the coast and accessibility to natural resources like mines and hydro power.

$$lnw_i = \alpha_0 + \sum_{p=1}^{4} \alpha_p lnE_{pi} + \sum_{k=1}^{76} \gamma_k X_{ki} + \sum_{k=1}^{290} \delta_k M_k + \varepsilon_{3i} \qquad (5)$$

where M_k (k= 1, 2, …, 290) denotes a set of dummy variables for the municipality in which the individual lives.

To deal with further endogeneity in the wage equation (5) we use an instrumental variable approach and choose as an instrument for the employment density, the land based area of each distance band. This parallels the instrumental variable used by Ciccone (2002). We, thus, assume that wages are *conditionally independent* of the amount of land area within each circle. Note that we condition on a set of individual specific characteristics, workplace characteristics and municipality fixed effects to obtain our preferred estimate (the number of municipalities are 290).

In addition, we also note that the relationship between employment density and land area may be either negative or positive. First, the more land area there is, the lower is the employment density for a specific number of jobs within the specific distance band. Secondly, since jobs in general cannot be located in water, the larger the land area the more jobs there is within a distance band (all else equal). The former effect will tend to produce a negative correlation between the amount of land area and employment density whereas the latter effect will tend to produce a positive correlation.

Thus, we effectively assume that there will be no direct effect of the amount of land within each circle on the individual's wage after conditioning on the previously mentioned observable characteristics and the municipality fixed effects. The effect of the amount of land within each circle on the individual's wage will only be indirect through its effect on the employment density in each circle. This brings us to our model for employment density which is also the first stage equation of our IV approach,

$$lnE_{pi} = \theta_{p0} + \theta_p lnLand_{pi} + \sum_{k=1}^{76} \mu_{pk} Xk_i + \sum_{k=1}^{290} \pi_{pk} M_k + \varepsilon_{4pi} \quad (6p)$$

where E_p (p=1, 2, 3, 4) is the employment density of the first, second, third and fourth distance bands; $Land_p$ is the amount of land within distance band p; Xk_i and M_k are defined as previously.

The second stage equation of our IV approach corresponds to equation 5 with $\ln E_p$ exchanged for the fitted value from the corresponding equation 6p. We use a two-stage-least squares estimator to conduct the empirical analysis. In addition, even though the number of observations on individuals is large in our data set, there is a smaller number of observations on the employment densities which are defined at the coordinate level. We have, in other words, clustering of observations at the coordinate level for the central independent variables in our data set. Therefore, we adjust the conventional standard errors of the two-stage-least squares estimator for clustering at the coordinate level using the Moulton factor outlined in Angrist and Pischke (2009, pp. 311-312) noting that the intra class correlation of the employment density is equal to one for all individuals living on the same coordinate; i.e. in the same square kilometer or square of a quarter of kilometer.

4. Data

The data used to estimate the four empirical models in this paper are derived from large administrative registers of the entire Swedish population aged 20-64 years in the year 2007. The registers are administered by Statistics Sweden (SCB). The information we use concerns information on a large set of individual and workplace specific characteristics. The information is geocoded so that the location of place of living and workplace can be observed. In the following we outline how we have used this information to arrive at the variables used in our empirical models. For a more detailed presentation of the original data we refer the reader to SCB (2009).

The dependent variable in our model is gross annual wage earnings had the individual worked full time. This reduces potential problems related to labor supply. The variable is constructed from register information on annual labor earnings, which basically consists of all taxable payments from an employer during the year. To convert this information to gross annual wage earnings at full time employment for each year we use information from a large yearly survey by Statistics Sweden of the Swedish wage structure where they collect information about hours work for: all employees in the public sector and approximately 50 percent of all employees in the private sector (SCB, 2012). This survey produces information on the share of full time employment for individuals included in the survey. Since some individuals in our data set is not included in this survey we impute the information on full time employment for these individuals. To this end we first estimate a model for share of full time employment on those individuals that have the relevant information and subsequently use this model to impute the information for the remaining individuals in our sample. The imputation model for full time employment use the following information: sector and industry of the individual's employment, age, education, number of children, a dummy variable for living close to a metropolitan area, distance to work, sum of income derived from early retirement and sickness benefits. We estimate separate imputation models for males and females. Finally, we divide the information on gross annual earnings with the information on the share of full time employment to arrive at our variable for the gross annual wage earnings at full time employment.

For each individual we have geocoded information for place of residence and workplace. The geocoded information is defined on a grid which is 1000 by 1000 meters in rural areas and 250 by 250 meters in urban areas. To construct our key explanatory variable employment per square kilometer land, we first count the number of individuals per workplace (jobs per workplace) and then we sum for each individual's place of residence the number of jobs within each of the following distance bands: 0-5 kilometers; 5-25 kilometers; 25-50 kilometers and 50-100 kilometers, respectively. We then use geocoded information on land and water (sea, lakes or rivers) defined on a grid of 1000 by 1000 meters to summarize the amount of land within each individual's distance bands, previously defined. To get the employment density we then simply take the ratio between number of jobs and land area within each distance band.

The control variables on individual specific characteristics includes, as previously, mentioned: a dummy variable for being male, age, dummy variables for educational attainment (45 different educational categories defined in terms of highest attained level and field of education), a dummy variable for marital status and separate variables for the number of children aged 0-3 years, the number of children aged 4-6 years, the number of children aged 7-10 years, the number of children aged 11-15 years, the number of children aged 16-17 years, the number of children aged 18-19 years. The variables concerning number of children all pertain to children still living at home with the individual. The workplace characteristics that we control for in the analysis pertains to the workplace at which the individual worked in November the relevant year. These characteristics include dummy variables for 11 specific industries and dummy variables for 10 specific sectors. The analyses below are based on a 10% random sample of the entire data set to save computation time.

Table 1 presents descriptive statistics of key variables used in the analysis. We see that the average gross annual earnings at full time employment is some 317 600 SEK (100 SEK in 2007 was approximately equal to 15 USD and 11 Euros). The sample weighted average employment density for the first circle is some 710 jobs per square kilometer. It is clear from the table that the sample weighted average employment densities fall the further away from the place of living. This is not surprising since we have more individuals living in circles that have a relatively high density. We also see that there is a substantial variation in the employment density between different individuals in the sample. Furthermore, the average land area of the first circle is some 70 square kilometers. Note also that the maximum value for this variable is 80 square kilometers which is an effect that we work with geographic information defined on a grid (the maximum value would be some 78.5 square kilometers for a true circle). The remaining values for the amount of land within each distance band indicates that bands grow larger size the further away it is located from the individual's place of living. The remaining part of Table 1 suggest that 50% of the individuals are male in the sample. The average individual is 42 years of age and works slightly more than 90% of full time.

Table 1. Descriptive statistics

Variable	*N*	*Mean*	*Std Dev*	*Minimum*	*Maximum*
Annual full time earnings	376426	3176.36	3044.53	1.00	609555.00
Density 5 km's	376426	709.57	1452.12	0.01	8273.65
Density 25 km's	376426	122.19	173.19	0.00	601.74
Density 50 km's	376426	27.92	37.95	0.00	444.11
Dens 100 km's	376426	20.48	16.25	0.00	137.81
Land area 5 km's	376426	69.71	11.59	2.00	80.00
Land area 25 km's	376426	1520.77	301.03	167.00	1896.00
Land area 50 km's	376426	4196.49	1134.61	574.00	5902.00
Land area 100 km's	376426	14511.14	4716.08	308.00	23534.00
Male	376426	0.50	0.50	0.00	1.00
Age	376426	41.99	12.19	20.00	64.00
Percent of full time	376426	90.57	16.21	5.00	100.00

5. Results

Table 2 contains the results of the empirical analysis. The column headed 'Equation 3' contains the estimates from the most simple model with no control variables, no municipality fixed effects and no instrumental variables involved. The results indicate positive effects of employment density in all but the distance band located furthest away from the individual where it seems to be negative if anything. The estimated coefficients indicate that the elasticity of wages with respect to employment density is some 0.007 in the first circle (radius 5 kilometers); 0.022 within the next-to-closest distance band (radius between 5 kilometers and 25 kilometers); 0.002 within the next-to-furthest away distance band (radius between 25 and 50 kilometers); and -0.002 within the distance band located furthest away (radius between 50 and 100 kilometers). Note that the two latter estimates are insignificantly different from zero, however. According to the theoretical framework outlined in section 2, these results thus indicate that the effect of agglomeration economies dominates the effect of congestion up to

50 kilometers away from the individual. If anything, the reverse seems true beyond 50 kilometers. This would suggest that the effect of agglomeration economies, that may result from externalities, decline more rapidly than the effect of congestion. In sum, these results seem quite in line with much of previous research in this field; i.e. there are some positive effects of employment densities in a relatively large area around an individual and these effects decline with distance.

To address the issue of spatial sorting on observables we include control variables for: gender, age, a set of dummy variables for highest educational attainment, a set of dummy variables for the industry and a set of dummy variables for the sector in which the individual is employed, and a set of variables for number of children in different age intervals in the estimation equation. The results are presented in column 2 (‘Equation 4’). The effects closest to the individual seem somewhat smaller than in the previous specification of the model. The effect next-to-furthest away seem somewhat larger while the effect furthest away is again insignificantly different from zero at conventional levels of statistical significance. The main impression is, however, that the estimated elasticities seem rather similar to those obtained with the simple model.

Next we control for municipality fixed effects beside the controls for individual specific characteristics. Here we find that the elasticity of wages with respect to wages is 0.007 in the area closest to the individual and 0.010 within the distance band next-to-closest to the individual. Furthermore, there is no statistically significant effect of employment density beyond 25 kilometers on the individual’s wage. So far the positive effect of employment density in the area closest to the individual seem rather robust across the three different specifications whereas the effects further away depend more on how the model is specified. The wage elasticity with respect to employment density in the distance band 50-100 kilometers is, however, insignificantly different from zero at conventional levels of significance across all three specifications.

Table 2. Estimates of the effect of employment density on wages

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.007*** (0.001)	0.005*** (0.001)	0.007*** (0.001)	0.114*** (0.050)
Ln density 25 km's	0.022*** (0.001)	0.016*** (0.001)	0.010*** (0.002)	-0.292*** (0.096)
Ln density 50 km's	0.002 (0.002)	0.005*** (0.001)	0.005 (0.004)	0.153 (0.116)
Ln density 100 km's	-0.002 (0.002)	-0.002 (0.002)	-0.001 (0.009)	0.002 (0.299)
R2-adjusted	0.006	0.178	-	-
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	376 426			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively.

In the last column ('Equation 5 by IV') of Table 2 we present the estimates obtained with the instrumental variable approach outlined in section 3 where the first stage regression involves regressing land area of each circle/distance band on the employment density of the corresponding circle/distance band while controlling for observable individual specific characteristics and municipality fixed effects. The results only indicate statistically significant effects of employment density up to 25 kilometers away from the individual. Here the estimated elasticity of employment density in the circle defined by a radius of 5 kilometers distance is found to be 0.114 which is thus substantially larger than in the other specifications. Furthermore, the estimated elasticity of wages and employment density in the distance band 5 kilometers to 25 kilometers is now strongly negative and equal to some -0.292. These results would thus indicate that in an area close to the individual, the effect of agglomeration economies, which may be related to externalities, tend to dominate negative congestion effects while the opposite is true in the distance band 5-25 kilometers away from the individual. Note also that the latter effect is consistent with a downward sloping demand curve for labor while the former effect suggest that the demand curve slope upwards in the area closest to the individual.

Sensitivity tests

The large differences found between the OLS estimates of equation 5 and the corresponding IV estimates might suggest that former suffer from a large omitted variables bias. There are of course alternative explanations for these differences. First, the effects of employment density may well be very heterogeneous. This heterogeneity may be the result of spatial factors, individual specific characteristics like education, or be a result of differences in the production function across different industries. It is well established in the evaluation literature and elsewhere (see e.g. Imbens and Angrist, 1994) that an instrumental variables estimator in such a situation identifies a so-called "local average treatment effect" (LATE).

Secondly, it is possible to question whether the exclusion restrictions hold for the instrumental variables in our model. For example, if there is spatial selection on unobservable characteristics then it may be the case that individuals with a value for the residual in the wage equation (high wage at a specific set of values for the control variables in equation 5) prefer to live close to water. Although we use the same type of instrument as Ciccone (2002), we note that this instrument might work better with his type of aggregated data for NUTS 3 regions in a set of European countries.

To investigate the scope of these potential problems in the previous analysis, we first estimate the previous models separately for the counties containing the three metropolitan areas in Sweden and the rest of the country to investigate spatial heterogeneity. We subsequently estimate the models separately for 10 different industries to see whether industry differences are relevant. The industries are classified in 10 broad categories: (1) Agriculture, forestry and fishing, (2) Manufacturing and mining, (3) Power production and waste disposal, (4) Construction, (5) Retailing and transport, (6) Financial and business services, (7) Education and research, (8) Health care, (9) Cultural services, (10) Public administration. To investigate whether the effect of employment density on wages vary by educational attainment of the individual we estimate the previous models separately for three broad educational categories: primary school, secondary school, and university.

To investigate how sensitive the previous results are to the choice of instrument we estimated the model with an additional set of instrument beside land area of each circle/distance band. This new set of instruments is motivated by the zero effect of employment density in the two distance bands furthest away from the individual. These two variables are, however, quite strongly correlated with the employment densities closer to the individual, which seems to suggest that we can use spatial lags of employment density as instruments for employment density itself. More specifically, we use employment density in the distance band 25-50 kilometers as an instrument for employment density in the circle closest to the individual's place of living and employment density in the distance band 50-100 kilometers as an instrument for employment density in the distance band 5-25 kilometers.

The results obtained from dividing the sample in different subsamples defined by county, industry and educational attainment, respectively, and re-estimate equation 5 by IV all indicate that the effects of employment density are indeed heterogeneous. In most cases the effects are statistically insignificantly different from zero. This is, of course, not so surprising since each subsample is smaller than the full sample. However, the effect in the area closest to the individual is positive rather than negative whenever the effect is statistically significant whereas the sign of the effects in the distance band further away from the individual varies and are rarely significantly different from zero. Thus, the basic result pertaining to 'Equation 5 by IV' in Table 2 with a positive effect in the area closest to the individual and a negative effect in the distance band 5-25 kilometers seem not completely at odds with the results presented in this part of our sensitivity analysis although the magnitudes of the two effects vary substantially between the different subsamples. Thus, the large difference between the OLS estimates and the IV estimates of model 5 in Table 5 may indeed be due to heterogeneity in the effect of employment density on wages. This corroborates the results reported by Graham and van Dender (2010) who empirically investigate the stability of agglomeration benefits in a productivity framework.

The results obtained from using spatial lags of employment density as additional instruments tell a similar story; i.e. the effect of employment density seems to be positive and statistically different from zero at conventional levels of significance in the area closest to the individual and negative and statistically different from zero at conventional levels in the distance band 5-25 kilometers away from the individual. The magnitudes are, however, substantially smaller in absolute terms than those reported in Table 2. The elasticity of wages with respect to employment density in the area closest to the individual (up to 5 kilometers away) is found to be some 0.036 (standard error 0.007) while it is -0.039 (standard error 0.014) in the distance band 5-25 kilometers away from the individual.

6. Concluding discussion

The aim of this paper is to estimate the effect of employment density on wages in Sweden using geocoded micro data on individuals and workplaces. We basically follow Ciccone (2002) to study agglomeration effects in Europe. However we modify his approach in a couple of ways. Firstly, we use a wage equation framework rather than estimating the effects of employment density on productivity. Secondly, we use micro data on individuals rather than aggregate data on regions. Finally, our geocoded data allows us to define each individual's relevant labor market area our self by a circle with four distance bands centered on the individual's place of living.

One result in our study is that there are positive and potentially large effects of employment density on individuals' wages. This result is coherent with many other studies (including Ciccone, 2002) on the topic. However, what is also obvious from our results is that the observed positive effect of employment density on individuals' wages is limited within a relatively small area around the individual. Beyond this closest surrounding (here 5 km) of the individual, we find a relatively large negative effect of employment density on wages. If we interpret these two results using the theoretical framework of the paper, this suggest that the effect of agglomeration economies, e.g. because of externalities, is large enough to dominate congestion effects of economic activity in the area just nearby of the individual but that the reverse is true thereafter. Furthermore, beyond 50 kilometers away from the individual there appears to be no statistically significant effects of employment density on wages, which indicate that agglomeration economies quite quickly becomes unimportant for individuals' wages.

We have conducted different sensitivity analyses to stress test these result. These analyzes suggest that the employment effect is indeed heterogeneous both spatially and individually. We obtain qualitatively the same results when using an additional set of instrumental variables (spatial lags for employment density); that is, a positive and statistically significant effect of employment density on the individual's wage in the area closest to him/her and a negative effect in the distance band 5-25 kilometers.

Acknowledgements

We would like to thank participants at the 2nd Symposium of the European Association for Research in Transportation Conference in Stockholm, Sweden, for valuable comments. Financial support from VINNOVA is gratefully acknowledged.

References

Angrist, J.D., Pischke, J.-S., (2009) *Mostly Harmless Econometrics – An Empiricist's Companion*, Princeton University Press.

Ciccone, A., Hall, R.E., (1996) Productivity and the density of economic activity, *American Economic Review*, 86: 54-70.

Ciccone, A., (2002) Agglomeration effects in Europe, *European Economic Review*, 46: 213-227.

Combes, P-P., Duranton, G., Gobillon, L., Roux, L., (2010) Estimating Agglomeration Economies with history, Geology and Workers Effects, in: Glaeser, E.L., (ed.) *Agglomeration Economies*, The University of Chicago Press, pp. 15-66.

Combes, P-P., Duranton, G., Gobillon, L., Roux, L., (2011) The identification of agglomeration economies, *Journal of Economic Geography*, 11: 253-266.

Dicken, P., Lloyd, P.E., (1990) Location in Space. Theoretical Perspectives in Economic Geography. 3rd edition, Harper Collins Publisher, New York.

Duranton, G., Puga, D., (2004) Micro-foundations of urban agglomeration economies., In Handbook of regional and urban economics, vol 4, (ed.) Henderson, V., Thisse, J-F., pp. 2063-2117, Amsterdam: North-Holland.

Eliasson, K., Lindgren, U., Westerlund, O., (2003) Geographical Labour Mobility: Migration or Commuting?, *Regional Studies*, 37(8): 827-837.

Eriksson, R., Lindgren, U., Malmberg, G., (2008) Agglomeration mobility: effects of localisation, urbanisation and scale on job changes, Environment and Planning A, 40: 2419–2434.

Fallah (2011), XX

Glaeser, E.L., (2000) The new economics of Urban and Regional Growth, in: Clark, G.L., Feldman, M.P., Gertler, M.S., (eds.) The Oxford Handbook of Economic Geography, Oxford University Press, Oxford, pp 83-98.

Glaeser, E.L., Maré, D.C., (2001) Cities and skills, Journal of Labour Economics, 19(2): 316-42

Graham, D.J., van Dender, K. (2010), "Estimating the agglomeration benefits of transport investments: Some tests for stability", *OECD Joint Transport Research Center*, Discussion paper No. 2009-32 July 2010 (Revised version).

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: 315-348.

Imbens, G., Angrist, J., (1994) Identification and Estimation of Local Average Treatment Effects, *Econometrica* 62, pp. 467-476.

Fujita, M., Krugman, P., Venables, A.J., (1999) *The Spatial Economy - Cities, Regions and International Trade*, Cambridge MA: The MIT Press.

Malmberg, A., Malmberg, B., Lundeqvist, B., (2000) Agglomeration and firm performance: economies of scale, localisation, and urbanisation among Swedish export firms., Environment and Planning A, 32:305-321.

Marshall, A., (1890) Principles of economics: An Introductory Volume, MacMillan, London.

Openshaw, S. Taylor, P.J., (1981) The modifiable areal unit problem, In: *Quantitative Geography: A British View*. (eds.) Wrigley, N., Bennett, R., pp. 60–69. Routledge and Kegan Paul, London.

Ottaviano, G.I.P., (2011) 'New' new economic geography: firm heterogeneity and agglomeration economies, *Journal of Economic Geography*, 11: 231-240.

Potter, A., Watts, H.D., (2011) Evolutionary agglomeration theory: increasing returns, diminishing returns and the industry life cycle, *Journal of Economic geography*, 11: 417-455.

Rosenthal, S.S., Strange, W.C., (2008) Evidence on the nature and sources of agglomeration economies. In: Henderson, V., Thisse, J-F., (eds.) *Handbook of regional and urban economics*, Vol 4, Amsterdam: North-Holland, pp. 2119-71

SCB, (2009) *Longitudinell Integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990-2007*, Arbetsmarknads- och utbildningsstatistik 2009:1

SCB, (2012): *Lönestrukturstatistik hela ekonomin 2011*, Beskrivning av statistiken 2012-06-20, AM0110.

Venables, A., (2011) Productivity in cities: self-selection and sorting, *Journal of economic geography*, 11: 241-251.

Åslund, O., Östh, J., Zenou, Y., (2010) How important is acess to jobs? Old question-improved answer, Journal of Economic Geography, 10: 389-442.

Appendix

Spatial heterogeneity in the estimated effects

In tables 3-6 we present results for models intended to investigate whether the results are different across the three counties containing the three major cities in Sweden and the rest of the country. According to the results obtained with equation 5 estimated by instrumental variables we see that the effect of employment density in the circle defined by a radius of 5 kilometers has a positive effect on wages for all counties containing one of the three largest cities in Sweden whereas the corresponding effect in the rest of the country is not significantly different from zero. Although the results for the “city-counties” all suggest a positive relationship between employment density in the area 0-5 kilometers away from the individual, the magnitudes appear quite heterogeneous. Regarding the effect of employment densities we see that they are quite heterogeneous across the different spatial divisions of the sample with some positive, some negative and some insignificant results. (The results obtained with equations 3, 4 and 5 estimated by OLS are presented in the tables for expository purposes. But we don’t comment them further in the text). These heterogeneous results may reflect a similar kind of heterogeneity reported by Graham and van Dender (2010) who use a productivity framework to assess agglomeration benefits of transport investments.

Table 3. Estimates of the effect of employment density on wages – *Stockholm county*

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.011*** (0.003)	0.004* (0.002)	0.011*** (0.003)	0.027*** (0.010)
Ln density 25 km's	-0.021*** (0.008)	NS	-0.031*** (0.010)	NS
Ln density 50 km's	-0.046*** (0.009)	-0.015** (0.007)	-0.024** (0.012)	-0.365** (0.186)
Ln density 100 km's	-0.131*** (0.024)	-0.067*** (0.017)	-0.071** (0.032)	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	83 453			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 4. Estimates of the effect of employment density on wages – *Skane county*

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	-0.011*** (0.002)	-0.007*** (0.002)	NS	0.075*** (0.018)
Ln density 25 km's	0.042*** (0.005)	0.020*** (0.004)	0.048*** (0.012)	NS
Ln density 50 km's	NS	0.011* (0.006)	0.059*** (0.016)	0.185*** (0.032)
Ln density 100 km's	NS	0.026* (0.015)	0.069** (0.033)	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	46 466			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively.

Table 5. Estimates of the effect of employment density on wages – "*Gothenburg" county*

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.005*** (0.002)	0.004*** (0.001)	0.007*** (0.002)	0.114** (0.056)
Ln density 25 km's	0.015*** (0.003)	0.009*** (0.002)	0.014* (0.008)	NS
Ln density 50 km's	0.027*** (0.006)	0.019*** (0.004)	NS	0.278** (0.125)
Ln density 100 km's	NS	NS	NS	0.815* (0.427)
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	64 133			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively.

Table 6. Estimates of the effect of employment density on wages – The rest of the country (all counties but the counties of Stockholm, Skane and "*Gothenburg"*)

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.006*** (0.001)	0.005*** (0.001)	0.007*** (0.001)	NS
Ln density 25 km's	0.016*** (0.002)	0.009*** (0.001)	0.014*** (0.002)	0.170** (0.068)
Ln density 50 km's	NS	NS	0.008* (0.004)	0.446** (0.180)
Ln density 100 km's	0.005** (0.002)	0.007*** (0.002)	NS	-0.512** (0.239)
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	182 374			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively.

Individual heterogeneity in the estimated effects

To investigate heterogeneity in the estimated effects of employment densities we divided the sample by industry (tables 7-17) and educational attainment (tables 18-20), respectively. Focusing on the results obtained with equation 5 estimated by IV we find virtually no significant effects of employment densities on wages. (The results obtained with equations 3, 4 and 5 estimated by OLS are presented in the tables for expository purposes. But we don't comment them further in the text). This may reflect the smaller number of observations used to obtain these results compared to using the full sample. This might underscore the importance of having a large enough sample to detect statistically significant effects of employment density in micro data.[4]

Table 7. Estimates of the effect of employment density on wages, industry: Non-specified industry

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.025*** (0.009)	0.021*** (0.009)	0.024** (0.012)	NS
Ln density 25 km's	0.051*** (0.014)	0.037*** (0.014)	NS	NS
Ln density 50 km's	-0.056*** (0.023)	-0.059*** (0.022)	-0.099* (0.055)	NS
Ln density 100 km's	NS	0.050*** (0.023)	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	2 822			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

[4] The standard errors reported in tables 7-17 have not been adjusted for clustering at the coordinate level, since the average number of individuals per coordinate in many of these regressions were considered to be too low for obtaining reliable estimates of the intra class correlation in the residuals.

Table 8. Estimates of the effect of employment density on wages, industry: Agriculture, forestry and fishing

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.022*** (0.006)	0.018*** (0.005)	0.027*** (0.006)	NS
Ln density 25 km's	NS	NS	NS	NS
Ln density 50 km's	NS	NS	0.049* (0.026)	NS
Ln density 100 km's	NS	NS	-0.073* (0.043)	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	2 952			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 9. Estimates of the effect of employment density on wages, industry: Manufacturing and mining

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.010*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	NS
Ln density 25 km's	0.038*** (0.002)	0.020*** (0.002)	0.018*** (0.004)	NS
Ln density 50 km's	NS	NS	NS	0.501** (0.231)
Ln density 100 km's	NS	NS	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	66 054			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 10. Estimates of the effect of employment density on wages, industry: Power production and waste disposal

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.008* (0.004)	NS	NS	NS
Ln density 25 km's	0.012* (0.006)	NS	NS	NS
Ln density 50 km's	NS	NS	NS	NS
Ln density 100 km's	NS	NS	-0.076* (0.042)	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	3 914			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 11. Estimates of the effect of employment density on wages, industry: Construction

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	NS	NS	0.003* (0.002)	NS
Ln density 25 km's	0.012*** (0.002)	0.008*** (0.002)	0.016*** (0.006)	NS
Ln density 50 km's	NS	NS	0.023*** (0.009)	NS
Ln density 100 km's	NS	NS	0.038** (0.016)	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	21 617			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 12. Estimates of the effect of employment density on wages, industry: Retailing and transport

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	NS	0.005*** (0.001)	0.005*** (0.002)	NS
Ln density 25 km's	0.029*** (0.002)	0.019*** (0.002)	NS	NS
Ln density 50 km's	0.007** (0.003)	0.008** (0.003)	-0.015* (0.008)	NS
Ln density 100 km's	NS	NS	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	66 796			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 13. Estimates of the effect of employment density on wages, industry: Financial and business services

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.011*** (0.002)	0.015*** (0.002)	0.011*** (0..003)	0.424*** (0.109)
Ln density 25 km's	0.058*** (0.003)	0.033*** (0.002)	NS	NS
Ln density 50 km's	NS	0.012*** (0.004)	NS	1.048*** (0.269)
Ln density 100 km's	-0.012** (0.005)	NS	NS	-3.058*** (0.953)
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	52 585			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 14. Estimates of the effect of employment density on wages, industry: Education and research

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.012*** (0.002)	NS	NS	NS
Ln density 25 km's	0.004* (0.002)	0.006*** (0.002)	NS	NS
Ln density 50 km's	NS	NS	NS	NS
Ln density 100 km's	NS	NS	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	45 707			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 15. Estimates of the effect of employment density on wages, industry: Health care

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.013*** (0.001)	NS	0.004** (0.002)	NS
Ln density 25 km's	0.004* (0.002)	NS	NS	NS
Ln density 50 km's	-0.007** (0.003)	NS	NS	NS
Ln density 100 km's	NS	NS	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	66 612			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 16. Estimates of the effect of employment density on wages, industry: Cultural services

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.011*** (0.003)	0.017*** (0.003)	0.023*** (0.004)	NS
Ln density 25 km's	0.018*** (0.004)	0.012*** (0.004)	0.018* (0.011)	NS
Ln density 50 km's	NS	0.014** (0.006)	0.030* (0.016)	NS
Ln density 100 km's	-0.020*** (0.007)	-0.020*** (0.007)	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	23 280			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 17. Estimates of the effect of employment density on wages, industry: Public administration

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.011*** (0.002)	0.008*** (0.002)	0.010*** (0.003)	NS
Ln density 25 km's	0.015*** (0.003)	0.013*** (0.003)	NS	-1.886** (0.866)
Ln density 50 km's	NS	NS	NS	NS
Ln density 100 km's	NS	NS	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	24 087			

Notes: Standard errors in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 18. Estimates of the effect of employment density on wages, educational attainment: Primary school

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.008*** (0.002)	0.004** (0.002)	0.006*** (0.002)	NS
Ln density 25 km's	NS	0.008*** (0.002)	NS	NS
Ln density 50 km's	NS	0.006* (0.004)	NS	NS
Ln density 100 km's	NS	NS	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	44 750			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 19. Estimates of the effect of employment density on wages, educational attainment: Secondary school

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	NS	0.004*** (0.001)	0.005*** (0.001)	NS
Ln density 25 km's	0.011*** (0.001)	0.013*** (0.001)	NS	NS
Ln density 50 km's	0.004** (0.002)	0.005*** (0.002)	NS	NS
Ln density 100 km's	NS	-0.004** (0.002)	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	187 318			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 20. Estimates of the effect of employment density on wages, educational attainment: University

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	NS	0.013*** (0.001)	0.015*** (0.002)	NS
Ln density 25 km's	0.033*** (0.002)	0.025*** (0.001)	0.019*** (0.005)	NS
Ln density 50 km's	0.006** (0.003)	0.005** (0.003)	NS	NS
Ln density 100 km's	NS	NS	NS	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	142 796			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Sensitivity to choice of instruments

To investigate how sensitive the results in table 2 are to the choice of instruments we estimated equations 3-5 only including employment densities in the two zones closest to the individual and used the employment densities in the two zones furthest away from the individual as additional instruments. The results of this sensitivity test are displayed in tables 21 and 22. Here we see that using only spatial lags as instruments in equation 5 produces results are insignificantly different from zero at the 10 percent level. However, when using both land areas and spatial lags as instruments the qualitative results are the same as in table 2; i.e. a positive effect of employment density in the zone closest to the individual and a negative effect in the zone next-to-closest to the individual. The magnitudes are much smaller in absolute terms, however.

Table 21. Estimates of the effect of employment density on wages – IV = spatial lags

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.007*** (0.001)	0.005*** (0.001)	0.007*** (0.001)	NS
Ln density 25 km's	0.022*** (0.001)	0.018*** (0.001)	0.009*** (0.002)	NS
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	376 426			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.

Table 22. Estimates of the effect of employment density on wages – IV = land areas + spatial lags

	Equation 3	*Equation 4*	*Equation 5*	*Equation 5 by IV*
Ln density 5 km's	0.007*** (0.001)	0.005*** (0.001)	0.007*** (0.001)	0.036*** (0.007)
Ln density 25 km's	0.022*** (0.001)	0.018*** (0.001)	0.009*** (0.002)	-0.039*** (0.014)
Control for observables	No	Yes	Yes	Yes
Municipality fixed effects	No	No	Yes	Yes
Instrumental variables	No	No	No	Yes
Number of observations	376 426			

Notes: Standard errors adjusted for clustering at the coordinate level in parentheses. ***, ** and * indicates that the estimate is significantly different from zero at the 1, 5 and 10 percent levels; respectively. NS means not significantly different from zero at the 10 percent level.