CESIS Electronic Working Paper Series

Paper No. 318

The Neighborhood or the Region? Untangling the density-productivity relationship using geocoded data

Johan P. Larsson

June, 2013

The Royal Institute of technology

Centre of Excellence for Science and Innovation Studies (CESIS)

http://www.cesis.se

The Neighborhood or the Region?
Untangling the density-productivity relationship using geocoded data

Johan P. Larsson[∅]

Abstract

I analyze the effects of sub-city level density of economic activity on worker productivity. Using a geocoded dataset on employment and wages in the city areas of Sweden, the analysis is based on squares representing "neighborhoods" (0.0625 km^2), "districts" (1 km^2), and "agglomerations" (10 km^2). The wage-density elasticity depends crucially on spatial resolution, with the elasticity being highest in neighborhood squares. The results are consistent with i) the existence of a localized density spillover effect and ii) quite sharp attenuation of human capital spillovers. An implication of the findings is that if the data source is not sufficiently disaggregated, analyses of the density-productivity link risk understating the benefits of working in dense parts of regions, such as the central business districts.

JEL: J24, J31, R12

Keywords: density, productivity, spatial dependence, geo-coded data, neighborhood effects, human capital, agglomeration economies

[∅]Centre for Entrepreneurship and Spatial Economics (CEnSE), Jönköping International Business School. E-mail: johan.p.larsson@jibs.hj.se

Acknowledgements: This research has been supported by a financial grant from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS).

1. INTRODUCTION

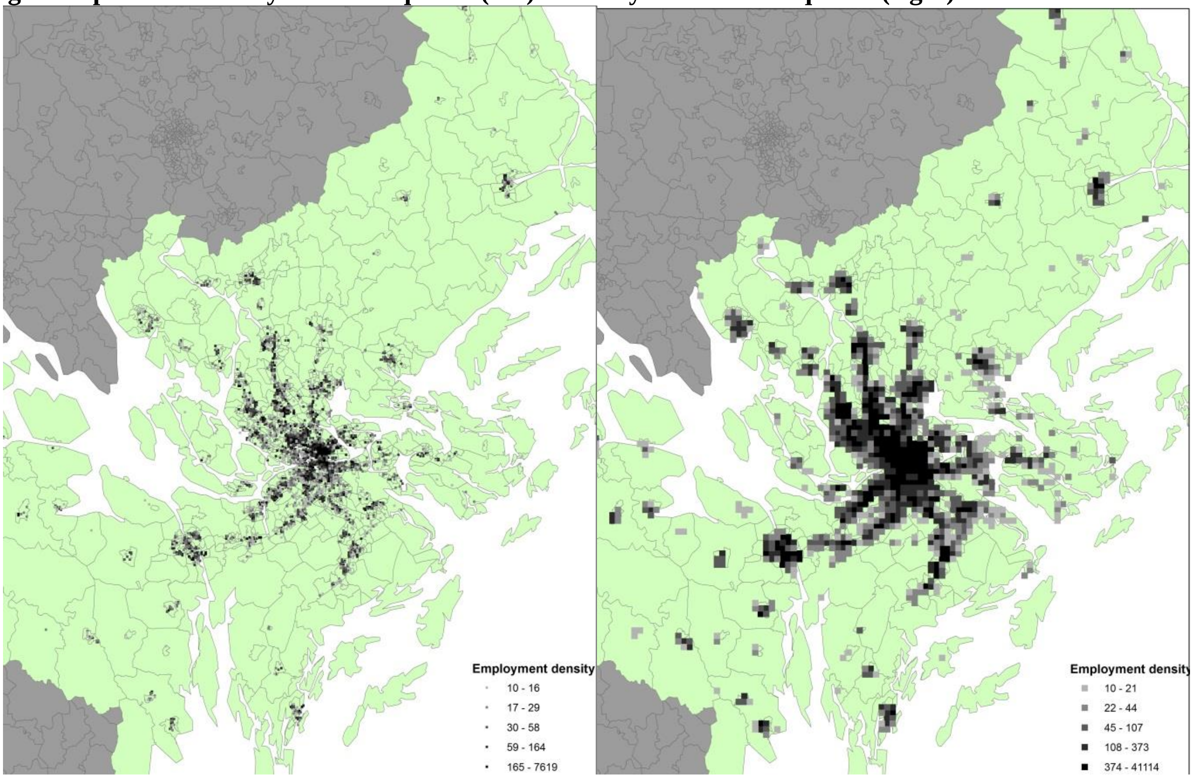
A large literature on the role of agglomeration economies emphasize the role of cities in stimulating local social interactions, or interactions not mediated by the price mechanism, as a source of productivity gains (e.g. Glaeser, 2000). In this paper, I evaluate the magnitude of the neighborhood effect as a source of density externalities, by drawing on the implications of a feature of one of the micro foundations of agglomeration economies, namely the observed tendency for human capital spillovers to work in small, confined environments. It is often postulated in the spatial economics literature that the diffusion of knowledge is driven by social interaction, and that its transfer is subject to distance decay (Rosenthal & Strange, 2001, 2008; Storper & Venables, 2004). Since human capital externalities are driven by interaction, the distance decay is sharp since frequency of contacts depreciate with distance.

Human capital and its externalities are essential in explaining cross-regional variation in productivity (Gennaioli, La Porta, Lopez-de-Silanes, & Shleifer, 2013). To date, most approaches to quantifying agglomeration economies implicitly assume that economic activity is uniformly spread out over rather large areas of land, and that there is zero attenuation of the spillovers within these areas. The effect, however, is likely to be very localized. In the literature on social interactions, the neighborhood approach is in fact quite established. E.g. Glaeser (2000) suggests that local social interactions are driven by what can be heard, seen, and felt, and hence that the potential of them taking place should depreciate sharply with space. An adjacent sociological literature on interpersonal networks confirm that frequency of contacts depreciate rapidly with distance. E.g., close to a majority of frequent contacts in the Toronto area have been shown to live within 1 mile from each other (e.g. Wellman, 1996). Following a large body of research in the urban economics literature, the main contribution of this paper is in estimating the relationship between wages and economic density, but with the addition of a ‘local agglomeration’ effect, looking at variation in economic density at various spatial square-grid resolutions - the smallest of which is 0.25-by-0.25 km - at the sub-city level.

In the literature on agglomeration externalities, the ‘agglomeration’ is generally defined as a labor market region, or a metropolitan statistical area. For most micro foundations, this approach makes perfect sense. When analyzing e.g. labor market matching (e.g. Helsley & Strange, 1990; Strange, 2009), the relevant area of observation is the labor market region, which by definition is highly integrated in terms of commuting flows. But when drawing on the wage-density relationship in analyzing human capital spillovers, it is not intuitively obvious what the correct level of aggregation should be, other than the fact that it most certainly is smaller than the average region (cf. Rosenthal & Strange, 2008). A good measure should proxy the potential for productive interaction in the local arena, but not be so small as to exclude relevant activity.

The need for highly disaggregated data when analyzing an effect that dissipates quickly with distance is illustrated by considering the substantial heterogeneity of sub-city economic activity, as is illustrated in figure 1. The figure shows square-level employment density in the Stockholm metropolitan labor market area for two different spatial resolutions: the 0.25-by-0.25 km squares (left) alluded to above, and for 1-by-1 km squares (right).

Figure 1. Private workforce employment density in the Stockholm metropolitan labor market area. The figure depicts 0.25 km-by-0.25 km squares (left) and 1-by-1 kilometer squares (right).



Note: only squares with more than 10 workers are included in the figure. The shade of each square indicates the quintile to which it belongs in terms of employment density. The dark shaded grey areas delineate adjacent labor market regions.

Economic activity is clustered toward the city center and through corridors following the major highways leading to the city. The uneven distributions fit well with, e.g. with the model in Glaeser (1999), where learning occurs (at some cost) close to the central business district of a city, surrounded by unpopulated hinterland. Even though Stockholm is a metropolitan area, it is evident that a vast number of employees in the region actually work in quite sparsely populated neighborhoods, and that within-region heterogeneity of economic activity is substantial.

The structure of Stockholm fits well with predictions from bid-rent theory, where rents near the city center are bid up by productive firms (e.g. Alonso, 1960). Why would they want to bear these costs? Koster et al. (2013) even document a substantial premium for building height, which the authors explain in part by *within-building* agglomeration gains. In a case study of Manhattan advertising firms, Arzaghi and Henderson (2008) document large externalities, but also that the effects dissipate incredibly fast: already after 750 meters, they are gone. Evidently, there is plenty of cheap, and rather unexploited, land area within the local labor market. And, as evidenced by the rents paid, the potential for productive interaction is unevenly spread out across space within cities which, I argue, should be reflected in the aggregation level of the density measure used.

1.1 Identification strategy

If social interactions are a quantifiable source of human capital externalities, and hence of agglomeration gains, then their effects may be identified by looking at variation in economic mass at the sub-city level. Empirically, this is performed by modeling an outcome of agglomeration economies (i.e. higher wages) as an effect of variation in within-neighborhood economic mass (i.e. density). I use three different square sizes to test the propositions above: 0.25-by-0.25 km (tentatively referred to as "neighborhoods"), 1-by-1 km ("districts"), and 10-by-10 km ("agglomerations")[1].

Empirically, the main idea is straightforward: following Rosenthal and Strange (2008), I analyze the relationship between wages and economic density around the individual's place of work. More specifically, I analyze the relationship between individuals' wages and square-grid economic density, while controlling for region-wide density (representing e.g. labor market pooling), firm-level characteristics, as well individual observed characteristics and individual unobserved heterogeneity. The procedure is repeated for all three square sizes in order to study the sensitivity of the estimate with respect to spatial resolution. The intuition behind using three measures is two-fold: first, it provides a general idea about the spatial scale at which productive non-market interactions occur, and second, substantially different estimates have implications for propositions about the attenuation of human capital externalities. To do this, I exploit a unique, geocoded, full-population dataset, tracking all Swedish workers in the private sector between 1991 and 2008. The data have previously been used in aggregated form by Andersson et al. (2012) to look specifically at attenuation using first and second-order neighboring squares.

Various confounding factors need to be tackled; these can broadly be sorted into two categories. First, sorting of productive individuals and firms to the same places will bias the estimates in the absence of proper controls. Second, there is an observed tendency for similar industries to co-locate within larger agglomerations (see e.g. Capello, 2002), such as financial districts located at the core of many large metropolitan areas. Such effects are indeed part of why workers in cities are more productive, but they are not density externalities.

[1] The fact that all squares are of the same size within regressions means that no normalization is needed to obtain an exact measure of density. This feature makes interpretation of the coefficients particularly straightforward.

I deal with the first issue in a number of ways. First, the data informs on a rich set of basic observable control variables on the individual, and firm levels, as outlined in table 1. Second, the longitudinal structure of the data allows within-transformation of the variables and the resulting fixed effects will absorb any time-invariant heterogeneity[2]. I tackle the second issue by controlling for local clustering of similar firms, by including a variable indicating the relative local presence of related industries, in addition to including industry fixed effects.

In order to study the sensitivity of the estimate, all square-level coefficients are first estimated without controls for region density. This exercise allows an assessment of the potential for conventional region density measures to pick up highly disaggregated phenomena in space, such as knowledge spillovers. If the square-level estimate is positive at first, but then pushed to zero, then the empirical framework supports measures of region density as sufficient proxies for non-market interactions. If the square-level estimate is positive and resilient to region controls, then the empirical context is consistent with a sub-city density externality operating at the very local level.

The empirical part of this paper shows that there are productivity spillovers within neighborhoods, districts, and agglomerations, even when controlling for economic activity at the level of regions. I provide further evidence of the attenuation of human capital spillovers, by observing persistently higher estimates when the unit of analysis is increasingly disaggregated. Further, the movements in the estimates indicate that the neighborhood level picks up only a small fraction of the region-wide effect and is arguably driven quite exclusively by effects that dissipate sharply with space (such as human capital externalities).

The remainder of this paper is structured as follows: section 2 introduces the data source, and the variables. Section 3 presents the empirical estimations, as well as various considerations regarding robustness. Section 4 concludes.

2. DATA AND VARIABLES

The smallest observational units in the United States that to my knowledge have been used for a similar purpose are the American Community Survey's PUMS areas (PUMAs). Even though the smallest of the PUMAs are approaching "neighborhood" size - the New York City PUMAs are typically 7-12 square kilometers - their mean size remains quite large. I exploit a full-population, geo-coded, publicly audited, longitudinal dataset, pinpointing each individual worker in the urban share of

[2] The importance of accounting for sorting on unobservables in studies of agglomeration economies is illustrated in Combes et al. (2008), and in Andersson et al. (2013).

Sweden's private workforce to a square, representing the sub-city level of analysis. I use squares of three sizes: neighborhoods (0.25-by-0.25 km), districts (1-by-1 kilometer), and agglomerations (10-by-10 kilometers). The neighborhood squares are the size of a few blocks, while the agglomeration squares approximately resemble the size of a small urban region in terms of land area. The division of Sweden into neighborhood squares in this fashion renders about a hundred times more areas than there are PUMAs in the United States (for a country that is less than one thirtieth of the United States in terms of population and about one twentieth in terms of surface area).

The data source is a publicly audited, matched employer-employee dataset, maintained by Statistics Sweden for the period 1991-2008. The landscape is divided into 0.25-by-0.25 km squares[3] using geo-codes on the establishment level. By matching each individual worker to a square via the employer, I collect aggregated wage sums for each square. These measures of wage density are then matched back to each individual worker. Since all squares are of equal size, no further normalization is needed to get an exact measure of density. This feature makes interpretation straightforward and eliminates the problem that some geographically large regions tend to have their economic activity concentrated to only a few areas.

The data include all wage workers 20-64 years old active in the private work force, with the exception of workers in the agriculture and mining industries. The remaining population is about 1.5 million individuals per year, although increasing in later periods. The original data source does contain individual observations with wages that are very high, as well as some wages well below subsistence levels. To alleviate problems with outliers and misreported observations, the individuals with wages in the 5 percent upper and lower tails are excluded from the estimations. Cutting the tails decreases the square elasticities somewhat, but the estimates presented below all exhibit strong robustness. The 5 percent cutoff has the benefit of making the dependent variable normally distributed, but using slightly lower or slightly higher cutoff values hardly affect the estimates at all[4]. Estimates using the full population are presented in the robustness section of the empirical part.

2.1 Density measures

The main variable of interest is density at the level of spatial squares. The density measure should to the furthest extent possible proxy the extent of local economic activity. In this paper, I use the sum of wages (w) per square. The intuition behind using wages, rather than employment, is two-fold. First, wages better proxy the actual local production; all else equal, higher paid, more productive workers

[3] The neighborhood squares are used as a base, from which higher-level resolutions are obtained by spatial aggregation.
[4] Incidentally, Combes et al. (2008) truncate their data for the same reason, using a 3 percent cutoff. Adopting that cutoff would not change anything substantial in practice, and nothing in terms of conclusions.

should be more valuable in terms of spillovers, and hence in terms of social multipliers. Second, even though the individuals included are wage workers, the exact number of hours worked is a missing variable, since the data source is publicly audited register data[5]. The density measure facing individual *i*, working in square *j* at time *t* is defined as the sum of wages for all *n* individuals working in the same square:

$$Dens_{i,j,t} = \left(\sum_{i=1}^{n} w_{i,j,t}\right) - w_{i,j,t} \tag{1}$$

Hence, density is measured as the sum of wages for all *n* individuals working in the same square, less the wages generated by individual *i*. The density measure is not a quotient, since all squares are uniform in terms of size. Estimates using employment density are presented in the robustness section of the empirical part.

Region density (accessibility)

Certainly, a control for region density is warranted. After all, perhaps a dense square is in reality a proxy for a dense region. The Swedish labor market is made up of 290 urban regions (municipalities). These are aggregated into 81 local labor market region (functional regions), each of which is highly integrated in terms of commuting flows. The measure of density used here is a distance-decay weighted accessibility approach, where the accessibility to economic activity in each urban region is the sum of local (M), intra-regional (R), and extra-regional (E) accessibility to wage sums. This approach is part of a line of thought dating back to the market potential measure in the seminal article by Harris (1954). Formally, for urban region *r*:

$$Dens_r^{Tot} = Dens_r^M + Dens_r^R + Dens_r^E \tag{2}$$

$Dens_r^M = W_r \exp\{-\lambda_M s_{rr}\}$, local accessibility to total wage earnings of urban region *i*,

$Dens_r^R = \sum_{k \in R_r} W_k \exp\{-\lambda_R s_{rk}\}$, intra-regional accessibility to total wage earnings of urban region *i*,

$Dens_r^E = \sum_{l \notin R_r} W_l \exp\{-\lambda_E s_{rl}\}$, extra-regional accessibility to total wage earnings of urban region *i*,

where internal accessibility is simply the wages accumulated in each urban region, weighted by average commuting time by car (s_{rr}). Intra-regional accessibility is accessibility to wages generated in

[5] Technically, using yearly wages may be a source of bias in an OLS setting under the assumption that workers in dense areas work longer hours than workers in sparsely populated areas, where a wage-differential unmatched by productivity differences would be observed. In a fixed effects setting, this is a smaller problem, since bias would only be introduced in the parameters to the extent that workers in dense areas work increasingly longer hours, relative to workers in sparse areas during the reporting period, and to the extent that such a phenomenon is not picked up by any of the fixed effects.

the same labor market region. The extra-regional measure picks up accessibility to all other urban regions in the country. The distance-decay parameter λ takes on three different values for local, intra-regional and inter-regional accessibility, respectively. These parameter values are based on observed commuting behavior, and are estimated for Swedish municipalities by Johansson et al. (2003). The summed up accessibility measure represents a continuous view of geography, where all activities are related in space, but where the magnitude of the effect diminishes with distance - only marginally within the own local urban region, but outside of it quite sharply, consistent with commuters' non-linear responses to differences in time-travel distances.

2.2 Control variables

All observable characteristics used in the empirical analysis is summarized in table 1.

Table 1. Description of right hand side variables, separated by level of aggregation.

Square level variables	
Density	Wage sums pertaining to the square where an individual is working. Formally defined in (1).
Industry concentration	Share of wage sums generated in the industry where the individual is working[6]. Formally defined in (3).
Regional level variables	
Density (accessibility)	Distance decay weighted accessibility to local, inter-regional, and intra-regional economic activity. Formally defined in (2).
Individual level variables	
Schooling	Theoretical number of years of the completed level of education.
Experience	Number of years working after graduation. Technically age less years of schooling less 6.
Experience squared	The square of the experience variable.
Tenure	Number of years that the worker has been employed by the same company. Measured since 1991.
Firm level variables	
Firm size	Number of workers employed by the firm by which the worker is employed.
Average years of schooling	Average length of education at the firm level.
Industry	Dummy variable on the 2 digit SIC level, indicating industry belonging of the worker's employer.

The industry concentration variable is included to control for possible localization economies within squares, i.e. benefits within clusters of industries. This variable reflects the degree of square-specific localization, representing the potential for sub-city areas to be specialized within the confinements of

[6] The industries are manufacturing, knowledge intensive business services, and other services.

larger economic agglomerations. A branch of the literature has dealt with the urbanization versus localization issue, but there is an observed tendency for similar industries to cluster to each other *within* a larger agglomeration of firms (see e.g.Capello, 2002), such as financial districts located at the core of many large metropolitan areas. Each such localization may be thought of as a specialized part of a diversified city or region, and may sensibly raise the productivity of the localized firms. In theory an agglomeration could even be the aggregation of various localized concentration effects. The industry concentration (or localization) variable is simply defined as the share of wages per square generated in the same broad industry in which the worker is employed. The industries are defined as knowledge intensive business services, other services, and manufacturing. Formally, industry concentration in industry *b* (for square *j* at time *t*) takes the form:

$$IndC_{b,j,t} = \frac{\sum_{b=1}^{3} w_{b,j,t}}{\sum_{i=1}^{n} w_{i,j,t}} \tag{3}$$

This variable pushes the density variable closer to being an indicator of localized urbanization economies, and ideally also closer to capturing human capital spillovers only, once region density is controlled for.

The schooling and experience variables are standard in the literature and are both expected to be positive since they indicate accumulation of human capital. The positive effect of experience is expected to attenuate with time, as is represented by the experience squared term. The tenure variable is measuring the number of consecutive years that each individual has spent with the same employer since 1991, and is an indication of the quality of the employer-employee match.

Firm size is included to control for higher productivity via increased division of labor and specialization in larger firms. This variable appears particularly important in smaller squares, since some of them can tend to be dominated by a few large firms. Average years of schooling at the firm level is added to assess peer effects and learning-by-interacting within firms. There is some empirical evidence to support the notion that an inflow of well-educated individuals at the firm-level may boost the productivity of other employees (e.g. Mas & Moretti, 2009).

Finally, all regressions also include time fixed effects, region fixed effects, and industry fixed effects, where the latter two are identified strictly by workers moving between regions and industries, respectively.

All variables are summarized in table 2.

Table 2. Descriptive statistics (all figures are on the individual level) 1991-2008.

Variable	Mean	St.dev. (overall)	St.dev. (between)	St.dev. (within)	Min	Max
Yearly wage (SEK)	229,123	93,281	80,906	61,557	60,100	495,700
Density (neighborhood , log)	13.07	1.82	1.72	.91	6.40	17.92
Density (district, log)	14.52	1.78	1.68	.90	6.40	19.34
Density (agglomeration, log)	16.94	2.02	1.93	.82	6.41	20.55
Density (region accessibility, log)	19.16	1.27	1.23	.40	14.48	21.32
Localization (neighborhood)	.75	.27	.23	.17	.00	1
Localization (district)	.65	.27	.23	.15	.00	1
Localization (agglomeration)	.47	.20	.18	.10	.00	1
Schooling	11.79	1.94	1.99	.34	9	22
Experience	21.07	11.59	12.04	4.02	0	49
Tenure	4.19	3.64	2.12	2.74	1	18
Firm size (log)	3.98	1.86	1.70	.89	0	9.41
Schooling (firm average)	11.84	1.24	1.17	.52	9	22

Total population size: 25 538 091. Neighborhoods, districts, and agglomerations refer to spatial squares with bases of 0.25 km, 1 km, and 10 km, respectively.

The average district associated with an individual in the dataset is about 4.5 times larger in terms of economic activity compared to the average neighborhood. In turn, the average agglomeration is home to about 10 times as much economic production as the average district. These are not differences in 'density', since the denominator differs between square sizes. The fact that the difference is not proportional to the difference in size is easiest thought of as further evidence that economic activity is not uniform across space. The larger squares contain larger proportions of 'idle lands': areas where there is no economic activity. Since economic production is clustered, smaller squares are on average made denser by a self-determining mechanism in the data generating process. Hence, smaller squares host more production per area unit of land compared to large squares, i.e., smaller squares are 'denser' on average. It is noteworthy that within to total variation is quite high for the square density variables, while it is lower for region density.

Mean industry concentration (localization) is decreasing quite sharply with square size, indicating that there is a higher localization present at the neighborhood level. This is further evidence of the concern briefly mentioned in the introduction: localization should not be thought of as exclusively a regional phenomenon (cf. Capello, 2002).

The mean of the tenure variable is not equal to average number of years that an individual in the dataset has stayed with the same employer since it is only measured since the beginning of the reporting period. It is also pushed downward by the age cutoff values. It should simply be thought of as a control variable, indicating the quality of the employer-employee match. It may also be argued that job switching constitute some form of optimization of the quality of the employee-employer match, making the expected sign of the variable ambiguous.

2.3 Model

The applied model is an augmented version of the wage equation in Mincer (1974), where logarithmic earnings are modeled as a function of schooling and experience. Indexing workers by *i*, time by *t*, firms by *f*, squares by *j*, broader industries by *b*, 2-digit SIC industries by *h*, and labor market regions by *R*, the equation to be estimated is:

$$\ln w_{i,t} = \alpha + \delta \ln Dens_{i,j,t} + \gamma \ln Dens_{r,t} + \phi\, IndC_{b,j,t} + \ldots$$
$$\ldots + Z'_{i,t}\rho + F'_{f,t}\theta + D_t + D_R + D_h + \eta_i + \varepsilon_{i,t} \quad (4)$$

where $w_{i,j}$ is the wage of individual *i* (situated in square *j*, region *k*, and working for industry *b*) in year *t*, $Dens_{i,j,t}$ represents square *j* economic density (adjusted for the contribution for individual *i*), $Dens_{r,t}$ is urban region *r* economic density, and $IndC_{b,j,t}$ is the industry concentration in the relevant square. The density and industry measures are defined in the previous section. Z_{it} is a matrix of individual *i* control variables, and $F_{f,t}$ is a matrix of firm *f* characteristics. The main empirical challenge is to provide credible estimates of δ for the three different square sizes, i.e. the elasticity of wages with respect to density, adjusted for confounding factors.

The individual-specific error term (η_i) contains unobserved, time-invariant, worker characteristics (e.g. ability) from the within transformation and is seen alongside the error term ($\varepsilon_{i,t}$). The underlying idea is to capture sorting on unobservable characteristics, which are known to play a substantial role in the literature on agglomeration gains (Combes et al., 2008). This within-transformation allows a clean estimation of returns to density, identified by changes in the variables over time, i.e. either by workers moving between squares, or by changes in square density.

3. RESULTS

The estimated parameters from (4) are displayed in table 3. Since the parameters are estimated with a fixed effect estimator, they are identified from variation in the squares over time. The elasticities thus reflect the percentage wage change associated with a 1 percent increase in the wage sums of the square where an individual is working, or alternatively, the percentage wage change associated with moving to a square where density is 1 percent higher, holding other factors constant.

Table 3. Estimated effects of density on wages separated by level of agglomeration.

	Aggregation level (square w*h)					
	Neighborhood (0.25 km)		District (1 km)		Agglomeration (10 km)	
Square density (log)	.0128 [.00019]	.0122 [.00020]	.0083 [.00025]	.0072 [.00026]	.0071 [.00023]	.0048 [.00026]
Region density (log)		.014 [.00097]		.015 [.00100]		.014 [.00109]
Industry concentration	.028 [.00112]	.028 [.00113]	.020 [.00135]	.020 [.00135]	.033 [.00202]	.032 [.00203]
Years of schooling	.051 [.00579]	.051 [.00579]	.052 [.00581]	.052 [.00581]	.052 [.00581]	.052 [.00581]
Experience	.029 [.00561]	.029 [.00561]	.029 [.00563]	.029 [.00563]	.029 [.00563]	.029 [.00563]
Experience squared	-.00063 [.00002]	-.00063 [.00002]	-.00063 [.00002]	-.00063 [.00002]	-.00063 [.00002]	-.00063 [.00002]
Tenure	.0054 [.00010]	.0054 [.00010]	.0054 [.00010]	.0054 [.00010]	.0054 [.00010]	.0054 [.00010]
Firm size (log)	.014 [.00045]	.014 [.00045]	.019 [.00045]	.019 [.00045]	.021 [.00047]	.021 [.00047]
Establishment schooling	.0170 [.00046]	.0166 [.00047]	.0177 [.00047]	.0174 [.00048]	.0183 [.00047]	.0181 [.00047]
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Within R^2	.39	.39	.38	.38	.38	.38
# Observations	25 538 091	25 538 091	25 538 091	25 538 091	25,538,091	25,538,091
# Individuals	3 478 681	3 478 681	3, 78 681	3 478 681	3 478 681	3 478 681
# Squares	67 779	67 779	12 714	12 714	1 454	1 454
# Urban regions		290		290		290
# Labor market regions		81		81		81

The robust standard errors are clustered at the square level and are presented in brackets (all presented variables are significant at the 0.1 percent level). The structure of the data is a panel from 1991-2008, and the parameters are estimated using a fixed effects estimator. Industry fixed effects are estimated on the 2 digit SIC level. All variables are defined in section 2. One mile is equal to 1.61 km. Dependent variable: natural logarithm of wage of individual *i*.

In the absence of controls for region density, the elasticity of wage with respect to square-level density is about .013 for neighborhood squares, .008 for district squares, and .007 for agglomeration squares, i.e. when observable and unobservable characteristics are controlled for, a doubling of the density of economic activity in a square is associated with wages increasing by about 0.7-1.3 percent, holding observable worker characteristics, firm attributes, and neighborhood characteristics constant. However, the magnitude of the effect depends on the level of aggregation, with the neighborhoods representing the upper bound.

One interpretation is that an increase of economic production in an area is more beneficial to an individual the closer to him it is; i.e. the image that emerges is certainly consistent with a role for non-market interactions in learning. Such a story would fit e.g. Jovanovic and Rob (1989), Duranton and Puga (2004), or other approaches wherein knowledge diffusion is at least in part the product of learning through human interaction.

The square-level estimates all decrease somewhat as the regional density control is added, but the decrease is higher, both in absolute and relative terms, the higher is the level of aggregation. This result is consistent with the higher-order aggregations picking up some of the micro foundations of agglomeration economies (e.g. effects from the regional labor market). The neighborhood estimate, however, decreases only marginally, indicating that the sub-city density externalities are quite robust to region effects at this level of aggregation. In terms of micro foundations, a plausible driving force for the neighborhood level estimate is, as argued above, human capital spillovers. Consistent with this hypothesis is the tendency for the estimate to diminish as the area is enlarged; a result consistent with quite sharp attenuation of the externality.

In any event, failure to model this effect will result in estimations where either too much of the external effects are attributed to the region or, worse, that the effect is omitted in its entirety, depending on the extent to which dense neighborhoods are correlated with dense regions. In this empirical context, the neighborhood effect would essentially be left out in its entirety. This observation means that analyses of the density-wage relationship are at risk of substantially understating the effect of density on wages.

The industry concentration estimate indicates that positive localization economies are present across square sizes, i.e. increases in concentration of an industry will tend to increase wages for workers in that industry; overall, a 10 percentage point increase in industry concentration increases wages in that industry by about 0.2-0.3 percent. The coefficient is increasing slightly with square size, although the lowest coefficient is registered for districts. Further, the tenure coefficient is weakly positive.

The schooling and experience coefficients have the expected signs and the magnitudes of the effects are broadly in concordance with previous findings (see e.g. Angrist & Krueger, 1991; Rauch, 1993), but it should be noted that fixed effects regressions can be problematic when analyzing these variables. However, since the schooling variable is theoretical years of schooling, it can tend to jump several years from one year to the next, making it easier to distinguish from the general trend.

3.1 Alternative interpretations and robustness

In this section I perform various robustness checks and discuss alternative interpretations of the coefficients in table 3.

Endogeneity

A plausible objection to the results in table 3 deals with what Combes et al. (2007) refer to as *endogenous quantity of labor*, i.e. endogeneity determined by the tendency for some locations in space to be 'inherently' more productive, and thereby made denser in the data generating process. At the level of regions, Combes et al. (2008) find that this form of endogeneity is far less important than selection, but ideally the problem should be addressed with an instrumental variable approach which, given the empirical framework, does present some challenges. Instrumental variables commonly used since Ciccone and Hall (1996) include historical populations at the city level. Such variables correlate rather poorly with square-level density, particularity in the smaller squares. Additionally, such variables are determinants of today's region density, which is another determinant of the dependent variable in this empirical framework. This anomaly is an obvious weakness of this approach and essentially means that most available approaches are invalidated by construction.

A potential way forward is noted in Combes et al. (2007), where geological instruments are used in conjunction with a micro-level dataset of French workers. The argument requires that the geological indicators proxy for historical population distributions, without driving today's productivity. One plausible channel is e.g. the tendency for fertile lands to have determined historical settlement patterns, or natural constraints on construction, without being determinants of modern-day productivity. However, the coefficients thus far have been estimated using changes in the fixed effects regressions, and the theoretical motivations of the instruments rather deal with levels. I use the instrument with a likely upward-biased 2SLS estimator and compare them to the simple pooled OLS estimator in table 5.

By matching data from the European Soil Database (ESDB) to the square grid using GIS software, several square-level indicators of geology are matched to the dataset. The indicators include water capacity, parent material and other properties of the soil. There are at least two additional problems with the approach in this empirical context. First, within-Sweden variation is low for all available variables. Second, the ESDB is not intended for use at the current level of disaggregation. Hence, it should be noted that the IV estimations, using a 2SLS estimator in table 5 are intended as further robustness checks only, and simply reflect the best available information, given the context.

Table 4. Neighborhood density (ln) coefficients from geological instrumental variable regressions.

	Aggregation level (square w*h)					
	Neighborhood (0.25 km)		District (1 km)		Agglomeration (10 km)	
	2SLS	OLS	2SLS	OLS	2SLS	OLS
Square density (log)	.0375 [.001]	.0204 [.001]	.017 [.000]	.0103 [.002]	.011 [.001]	.005 [.001]
F (excluded instruments)	14299		40318		160000	
R^2	.33	.33	.33	.33	.33	.33

The robust standard errors are clustered at the square level and are presented in brackets (all variables are significant at the 0.1 percent level). The control variables are identical to those reported in table 4. Geological instruments used: parent material (3 dummies), water capacity (3 dummies), depth to rock (4 dummies), top soil type (3 dummies), and sub soil type (2 dummies). One mile is equal to 1.61 km. Dependent variable: natural logarithm of wage of individual *i*.

Note that unobserved ability is likely to put upward pressure on the estimates in table 5; the size of the coefficient is therefore of lesser importance. What is more interesting is the direction of the effect: across the board, the 2SLS estimates are *higher* than those obtained by OLS. Even though the coefficients should be interpreted as the upper bounds, they do no indicate that reverse-causality is a likely driving force behind the estimates in table 3.

Alternative specifications

Numerous methods of estimating density externalities have been evaluated in the literature. Under this section, I evaluate a series of plausible alternative specifications of (4). First, in table 5, the coefficients are estimated using all spatial lags in one regression. The only difference is that no density is "double-counted", i.e. the neighborhood density is subtracted from the district squares, and the district density is subtracted from the agglomeration squares. Third, the elasticity of wages with respect to region density is estimated as a reference, without the disaggregated variables.

Table 5. **Estimated effects of spatially lagged density components on wages.**

	(1)	(2)	(3)
Neighborhood density (0.063 km^2, ln)	.0131 [.00021]	.0128 [.00021]	
District density (1 km^2, ln)	-.0012 [.00023]	-.0012 [.00024]	
Agglomeration density (10 km^2, ln)	.0012 [.00021]	-.0004 [.00020]	
Region density (ln)		.017 [.00106]	.024 [.00213]

Robust standard errors are presented in brackets. The standard errors in columns 1-2 are clustered at the square level, and those of column 3 on the level of local labor market regions. The control variables are identical to those reported in table 3 (the industry concentration measure is included for all levels). The structure of the data is a panel from 1991-2008, and the parameters are estimated using a fixed effects estimator. One mile is equal to 1.61 km. Dependent variable: natural logarithm of wage of individual i.

The results are consistent with sharply attenuating density externalities operating at the sub-city level. When controlling for region density, the first and second order spatially lagged terms even have slightly negative coefficients, although they are both fairly close to zero. Certainly, the estimates are consistent with the “near neighborhood” (Glaeser, 2000) as the relevant arena for non-market interactions. The region effect remains robust and in line with reasonable expectations, but other than that, the returns to density coming from outside of the neighborhood is essentially zero (cf. Andersson, Klaesson, & Larsson, 2012). Do note that the coefficient of the agglomeration squares is weakly positive in the left regression, consistent with this measure picking up some region-wide micro foundations, such as labor market pooling.

When observing the third column in table 5 it is obvious that the estimated returns to region density is indeed higher when the square-level variables are excluded. The estimate indicates that the conventional region-wide estimate does pick up some of the disaggregated effects, and may in fact be a good estimate of the total, black-box returns for the average worker. This empirical framework, then, indicates that estimating returns to density in this way will understate the value of density for workers in dense sub-city areas (such as the central business district), while overstating it for workers outside of those areas.

Other possible sources of misspecification of the model are addressed in table 6, highlighting two specific, additional causes for concern. The first set of regressions changes the variable of interest from wage density to employment density, which is a more commonplace measure of economic density. The second set estimates the parameters using the entire population, including potential outliers and misreported observations.

Table 6. Robustness checks of results from table 3. The presented estimate is the elasticity of wage with respect to square-level density.

Change in specification	Aggregation level (square w*h)		
	Neighborhood (0.25 km)	District (1 km)	Agglomeration (10 km)
(1) Employment density (log)	.0110 [.00030]	.0059 [.00028]	.0042 [.00027]
(2) Full population	.0175 [.00032]	.0104 [.00044]	.0060 [.00042]

The robust standard errors are clustered at the square level and are presented in brackets (all variables are significant at the 0.1 percent level). The control variables are identical to those reported in table 4. The structure of the data is a panel from 1991-2008, and the parameters are estimated using a fixed effects estimator. Industry fixed effects are estimated on the 2 digit SIC level. All variables are defined in section 2. One mile is equal to 1.61 km.

As can be seen from the first set, changing from wage density to employment density produces very similar, albeit slightly lower, estimates for each level of aggregation. The full population estimate (using wage density, as in table 3) yields slightly higher results compared to the previous estimates. Certainly, both of these results serve to strengthen the case for robustness in terms of potential misspecification.

4. CONCLUSION

Nonmarket interactions are often put forth as a source of agglomeration gains. The implied mechanism is increased human capital accumulation through interaction with others, implying that nonmarket interactions provide one of the mechanisms explaining the link between productivity and spatial economic density. In addition, the fact that interaction depreciates sharply with distance, provides a mechanism explaining the tendency for human capital externalities to attenuate with space, as documented e.g. in Rosenthal and Strange (2008).

I argue that the classic micro foundations of density externalities imply that a density-productivity analysis should be conducted for different spatial resolutions, and specifically include a sub-city effect representing human capital spillovers, in addition to the regional effect from the classic agglomerating factors. These arguments are far from novel, but the empirical analyses carried out thus far have commonly been constrained by availability of data in analyzing this effect. I estimate the density-wage relationship, using squares with bases of 0.25 km, 1 km and 10 km, tentatively referred to as neighborhoods, districts, and agglomerations, respectively.

The elasticity of wage with respect to neighborhood density is about 1.3 percent, and the estimate is decreasing with square size, suggesting that the results are sensitive to choice of spatial scale. When looking at lagged density variables in the same regressions, almost all of the effects come from the neighborhood density in the disaggregate and from region density in the aggregate, consistent with different micro foundations of agglomeration gains operating at different spatial resolutions. Broadly, the analysis is in concordance with propositions about the tendency for human capital externalities to operate in small, confined environments, with a sharp attenuation effect.

Further, the analysis bears on the practice of estimating returns to density. If the data are not sufficiently disaggregated, the neighborhood effect could result in an omitted variable bias, but the severity of the problem depends on the objective of the study. The results give some support for region level density measures picking up at least some of the effects when used on their own, but two caveats need to be pointed out. First, disaggregated data are needed if it is part of the objective to untangle the micro foundations of agglomeration economies. Second, the implication is that analyses of agglomeration gains risk understating the *learning effect* (Duranton & Puga, 2004) of working in the denser parts of cities, by assuming that this effect is constant across space within regions.

REFERENCES

Alonso, W. (1960). A theory of the urban land market. *Papers in Regional Science, 6*(1), 149-157.

Andersson, M., Klaesson, J., & Larsson, J. P. (2012). How Local are Spatial Density Externalities? Evidence from square grid data. *CIRCLE Working Paper 2012/10*.

Andersson, M., Klaesson, J., & Larsson, J. P. (2013). The Sources of the Urban Wage Premium by Worker Skills - Spatial Sorting or Agglomeration Economies? *Papers in Regional Science (forthcoming)*.

Angrist, J. D., & Krueger, A. B. (1991). Does Compulsory Schooling Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics, 106*, 976-1014.

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

Capello, R. (2002). Entrepreneurship and spatial externalities: Theory and measurement. *The Annals of Regional Science, 36*(3), 387-402. doi: 10.1007/s001680200106

Ciccone, A., & Hall, R. E. (1996). Productivity and the Density of Economic Activity. *The American Economic Review, 86*(1), 54-70.

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. (2007). Estimating Agglomeration Economies with History, Geology, and Worker Effects In E. L. Glaeser (Ed.), *Agglomeration Economics*: The University of Chicago Press.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2063-2017). Amsterdam: North Holland.

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). Human Capital and Regional Development. *The Quarterly Journal of Economics, 128*(1), 105-164.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics, 46*(2), 254-277. doi: 10.1006/juec.1998.2121

Glaeser, E. L. (2000). The Future of Urban Research: Nonmarket Interactions [with Comments]. *Brookings-Wharton Papers on Urban Affairs*, 101-149. doi: 10.2307/25067375

Harris, C. D. (1954). The Market as a Factor in the Localization of Industry in the U.S. *Annals of the Association of American Geographers, 44*, 315-348.

Helsley, R. W., & Strange, W. C. (1990). Matching and agglomeration economies in a system of cities. *Regional Science and Urban Economics, 20*(2), 189-212. doi: 10.1016/0166-0462(90)90004-m

Johansson, B., Klaesson, J., & Olsson, M. (2003). Commuters' non-linear response to time distances. *Journal of Geographical Systems, 5*(3), 315-329. doi: 10.1007/s10109-003-0111-2

Jovanovic, B., & Rob, R. (1989). The Growth and Diffusion of Knowledge. *The Review of Economic Studies, 56*(4), 569-582.

Koster, H. R. A., van Ommeren, J., & Rietveld, P. (2013). Is the sky the limit? High-rise buildings and office rents. *Journal of Economic Geography*.

Mas, A., & Moretti, E. (2009). Peers at Work. *American Economic Review, 99*(1), 112-145.

Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: National Bureau of Economics Research.

Rauch, J. E. (1993). Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. *Journal of Urban Economics, 34*(3), 380-400. doi: 10.1006/juec.1993.1042

Rosenthal, S. S., & Strange, W. C. (2001). The Determinants of Agglomeration. *Journal of Urban Economics, 50*(2), 191-229. doi: 10.1006/juec.2001.2230

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics, 64*(2), 373-389. doi: 10.1016/j.jue.2008.02.006

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography, 4*(4), 351-370.

Strange, W. C. (2009). Viewpoint: Agglomeration Research in the Age of Disaggregation. *The Canadian Journal of Economics / Revue canadienne d'Economique, 42*(1), 1-27.

Wellman, B. (1996). Are personal communities local? A Dumptarian reconsideration. *Social networks, 18*(4), 347-354.