CPB Netherlands Bureau for Economic Policy Analysis

CPB Discussion Paper | 184

Regional wage differences in the Netherlands: Micro-evidence on agglomeration externalities

Stefan Groot
Henri de Groot
Martijn Smit

CPB Discussion Paper

No 184

July 2011

Regional wage differences in the Netherlands: Micro-evidence on agglomeration externalities

Stefan P.T. Groot, Henri L.F. de Groot and Martijn J. Smit

CPB Netherlands Bureau for Economic Policy Analysis
Van Stolkweg 14
P.O. Box 80510
2508 GM The Hague, the Netherlands

Telephone +31 70 338 33 80
Telefax +31 70 338 33 50
Internet www.cpb.nl

ISBN 978-90-5833-518-0

Abstract in Dutch

Op basis van microdata over individuele werknemers voor de periode 2000–2005 laat deze studie zien dat er – hoewel relatief klein in internationaal perspectief – substantiële loonverschillen bestaan tussen Nederlandse regio's. Een groot deel van deze verschillen kan worden toegeschreven aan individuele kenmerken van werknemers. Het resterende verschil wordt deels verklaard door variaties in dichtheid van werkgelegenheid – met een elasticiteit van ongeveer 3,8 procent, en door Marshall-Arrow-Romer externaliteiten, waarbij een verdubbeling van het aandeel van een (2-digit NACE) sector in een 2,4 procent hogere productiviteit resulteert. Voor de aanwezigheid van een negatief effect van concurrentie (door Porter externaliteiten) en diversiteit (geassocieerd met Jacobs externaliteiten) worden geen bewijzen gevonden.

Regional wage differences in the Netherlands: Micro-evidence on agglomeration externalities

Stefan P.T. Groot[a,b], Henri L.F. de Groot[a,b,c,d] and Martijn J. Smit[a,1]

[a] Dept. of Spatial Economics, VU University Amsterdam, The Netherlands

[b] CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands

[c] Tinbergen Institute, Amsterdam-Rotterdam, The Netherlands

[d] Ecorys-NEI, Rotterdam, The Netherlands

Abstract

Based on micro-data on individual workers for the period 2000–2005, we show that regional wage differentials in the Netherlands are small but present. A large part of these differentials can be attributed to individual characteristics of workers. Remaining effects are partially explained by variations in employment density, with an elasticity of about 3.8 percent and by Marshall-Arrow-Romer externalities, where doubling the share of a (2-digit NACE) industry results in a 2.4 percent higher productivity. We find evidence for a negative effect of competition (associated with Porter externalities) and diversity (associated with Jacobs externalities).

Keywords: regional labour markets, wage differentials, agglomeration externalities

JEL codes: J24, O12, R11, R23

---

1 Corresponding author: Henri L.F. de Groot, Dept. of Spatial Economics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, email: hgroot@feweb.vu.nl. Useful comments by Pierre-Philippe Combes, Harry Garretsen, Gerard Marlet, Piet Rietveld, Coen Teulings, Wouter Vermeulen and seminar participants at the SERC research seminar at the London School of Economics are gratefully acknowledged. This research heavily draws on data made available by Statistics Netherlands. Their support has been indispensible. The usual disclaimer applies.

1. Introduction

Regional wage disparities are known to be large in many countries, and they are often a source of public concern. They are the reflection of several forces, including sorting processes of individuals and firms with different characteristics and potentially also agglomeration externalities that affect the productivity of individuals as a function of characteristics of the area in which they work. Most governments have specific policies targeted at regions that structurally lag behind. Properly targeted policies require a thorough understanding of the sources of the productivity differences.

This paper aims to identify the nature and causes of wage differences in the Netherlands. The Netherlands is an interesting case because of its perceived flatness in both geographical as well as economic dimensions. Its institutional setting is known to result in a fairly equal distribution of income (see, for example, De Groot et al., 2006). Even though regional wage differences in the Netherlands are relatively small, we will show that there are still substantial regional differences mainly between the main agglomerations in the Randstad region and the more peripheral regions.

To achieve these goals, we first describe the nature and magnitude of regional (pre-tax individual worker) wage differences in the Netherlands. We subsequently relate the spatial component of observed wage differences to agglomeration effects. Our analysis is based on unique labour market micro-data provided by Statistics Netherlands (CBS). In its setup, we closely follow the analysis on spatial wage disparities in France by Combes et al. (2008a). To our knowledge, the current study is the first that uses this type of micro-data to estimate agglomeration economies in the Netherlands. An important advantage of the use of micro-data is that it provides an opportunity to reduce worker heterogeneity that remains unobserved at a more aggregated level. Previous studies have shown that these effects may be substantial. The meta-analyses of De Groot et al. (2009) and Melo et al. (2009) provide overviews of these fields of research.

In the remainder of this paper, we start by discussing different causes of regional wage differences. Section 3 provides a description of the data and methodology that are used. Section 4 presents stylized facts about regional differences in wages. Section 5 uses the Mincer equation to relate wage differences to observed worker characteristics and to subsequently derive a spatial residual that captures wage variation across space that cannot be attributed to individual characteristics. Section 6 attempts to further explain this spatial residual, and relates it to different agglomeration externalities. Section 7 discusses the robustness of the results along several dimensions and Section 8 concludes.

2. Sources of regional wage differences

Three main sources of regional wage differences are typically distinguished in the literature (see, for example, Combes et al., 2008a). The first is the composition of the labour market, which is related to urbanization externalities. The second set of explanations relies on differences in the presence of local non-human endowments. The third consists of agglomeration economies: spatial proximity of firms to other firms, to producers or to suppliers. We will briefly discuss these three in turn.

Workers with different skills and experience levels, or with different ethnic backgrounds, are not homogeneously distributed across space.[2] As sectors are not spread evenly across regions as well, and different industries require a different mix of worker characteristics, workers tend to spatially sort themselves based on the supply and demand for their specific competences (cf. Combes et al., 2008a). Yet in an equilibrium situation[3] this sorting will occur only if firms in those regions that experience a shortage of workers in specific industries or occupations offer a wage premium over firms in regions where these workers are relatively abundant. One reason for the absence of an isotropic wage landscape is that institutions in higher education as well as industries that require highly skilled labour are usually concentrated in densely populated cities. Students that move from the periphery to a city to take education there, have little incentive to move back to the periphery after completing their education. Composition can be held accountable for a part of the spatial differences in wages. In other words, assuming that wages are equal to the marginal product of labour, average wages will differ across regions, even when there are no regional differences in the productivity of workers with equal characteristics.[4]

Local non-human endowments are a second source of wage and productivity differences. Regions that have good access to waterways, a favourable climate, or valuable

[2] We assume that gender, which is also a common cause for wage differentials (e.g., Altonji and Blank 1999), has a more or less uniform spatial distribution in the Netherlands.

[3] Workers are of course more inclined to migrate to other regions if they cannot find employment in their own region, but we build here on the commonly made assumption that in an equilibrium situation the labour market can be characterized by full employment.

[4] This paper does not take consumers and their preferences into account. A consumer preference for densely populated regions could also result in spatial sorting of different education groups, as workers with a higher income can pay a higher price for housing in their most preferred areas.

natural recourses can have a higher productivity than less endowed areas. An especially interesting type of non-human endowments are those with a non-natural nature, like technology, local institutions and private capital, as they are often endogenous. Railway stations in the nineteenth and twentieth century are a nice example. Stations were built on the most populated locations at that time, but as they strongly reduced distance (measured in time), they further reinforced agglomeration forces.

Third, various authors have pointed at the importance of agglomeration externalities for economic growth. One of the mechanisms through which agglomeration works is physical proximity as well as the scale of both demand and supply, which reduces transaction costs, both on markets for goods and on markets for production inputs. Thus, agglomeration results in cost reductions through enabling groups of firms to enjoy collective economies of scale (Harvey, 1981, p. 105). Yet another source of productivity growth are knowledge spillovers, resulting in more innovation in agglomerations (Jaffe et al., 1993). According to the Marshall-Arrow-Romer (MAR) model the main 'route' of these spillovers is intra-sectoral, which Glaeser et al. (1992) tested in their seminal paper on cities as the centres of growth. In the MAR model, knowledge is industry-specific and regional concentration of certain industries therefore allows knowledge spillovers between firms in the same industry. Yet some types of spillovers occur across sectors; these diversity effects are known as Jacobs externalities (cf. Jacobs, 1969). Where differences meet, innovations are born. A third category of agglomeration economies consists of the so-called Porter externalities after Porter (1990) who pointed at the importance of (local) intra-sectoral competition as a source for productivity gains.

After the work of Glaeser et al. (1992), who found that Jacobs externalities were empirically the most important agglomeration effect, many studies repeated their analysis for different countries, regional definitions, time periods, proxies for the agglomeration externalities, etc. Reviews of this strand of literature are provided, among others, by Rosenthal and Strange (2004), Beaudry and Schiffauerova (2009), De Groot et al. (2009) and Melo et al. (2009). The latter two contributions present meta-analyses of the existing literature and find that agglomeration externalities are generally positive, but with large variation across space, time and research method. The inclusion of control variables (like industry effects), or the use of micro data instead of macro-level are of importance for the outcomes. Another interesting result is that these agglomeration externalities, if anything, tend to become more important over time.

3. Data and methodology

This paper combines data from Dutch tax records with the labour force survey (EBB) and firm data. The level of observation is that of the job, so a worker can have multiple entries in each year. Data are available for the period 2000–2005, with over 10 million observations annually for tax data and 70,000 observations for the labour force survey.

We use pre-tax hourly wages of individual workers and jobs, which provide the closest approximation of the productivity level of workers. By combining employer (ABR) and census data (SSB), the work location is available at the municipality level. For part of the analyses in this paper we aggregate the location-specific data to NUTS-3 level (the so-called COROP; see Appendix A for a map of the Netherlands) or the municipality level.[5] We use the classification of 2005, and have data on 40 NUTS-3 regions and 467 municipalities. The work location is defined as the job site or business unit. For each employee, we have information on his or her age, gender, ethnicity, hourly wage, and workplace location. For each business unit, we have the sectoral classification on the 2-digit NACE level and the number of employees.

To estimate wage regressions, we combine this large dataset with the Dutch employment survey. This dataset includes data on education level and job type. We exclude all employees earning less than 10% or more than 1000% of the average hourly wage, all workers younger than 18 or older than 65, and all workers with a working week of less than 12 hours.[6] After merging this data with census and firm data, and selections as described above, we have on average 34,935 observations for each year.

We construct our agglomeration variables directly from the micro-data. For this purpose we use cross sections of tax data. We use the share of industries in the regional economy to capture MAR externalities (with E for total employment):

$$Specialization_{ir} = \frac{E_{ir}}{E_r}\,. \tag{1}$$

[5] Briant et al. (2008) discuss the importance of regional classifications for the outcomes of economic geographical research, and conclude that it is important that the chosen scale of a regional classification corresponds with the level of aggregation at which the researched phenomenon is expected to operate. Even though COROP regions are not strictly local labour market areas, they provide us with the most reasonable approximation in the Dutch case.

[6] We distinguish between part-time employees – working 12 hours or more per week, but less then 32 hours – and full-time employees working 32 hours per week or more. Having a working week of 12 hours or more is the official definition of Statistics Netherlands of being employed.

When industry dummies are included, this variable captures the same effect as a location quotient, viz. the effect of a smaller or larger share of an industry relative to the share of that same industry in all other regions.

We use Shannon's entropy (after Shannon, 1948) to capture externalities from diversity:

$$Diversity_r = -\sum_i \left(\frac{E_{ir}}{E_r} \times \ln \frac{E_{ir}}{E_r}\right), \tag{2}$$

where we sum over the industries. A high value means that the region is highly diversified in terms of its employment structure, whereas a low value means that the regional economy is rather specialized in only a few large sectors.

Competition is measured using a Hirschman-Herfindahl based index on the distribution of employees across firms:

$$Competition_{ir} = 1 - \sum_f \left(\frac{E_{fir}}{E_{ir}}\right)^2, \tag{3}$$

where we sum over the individual firms in each sector-region combination. Since we calculate the index as one minus the Hirschman-Herfindahl index (HHI), a value close to one indicates fierce competition in a region. When the index is low (e.g., below 0.8), the regional employment is highly concentrated in a relatively small number of firms.

Finally, we use the employment density to capture general urbanization effects in a region (where A stands for the surface of the area):[7]

$$Density = ln\left(\frac{E_r}{A_r}\right) = lnE_r - lnA_r\ . \tag{4}$$

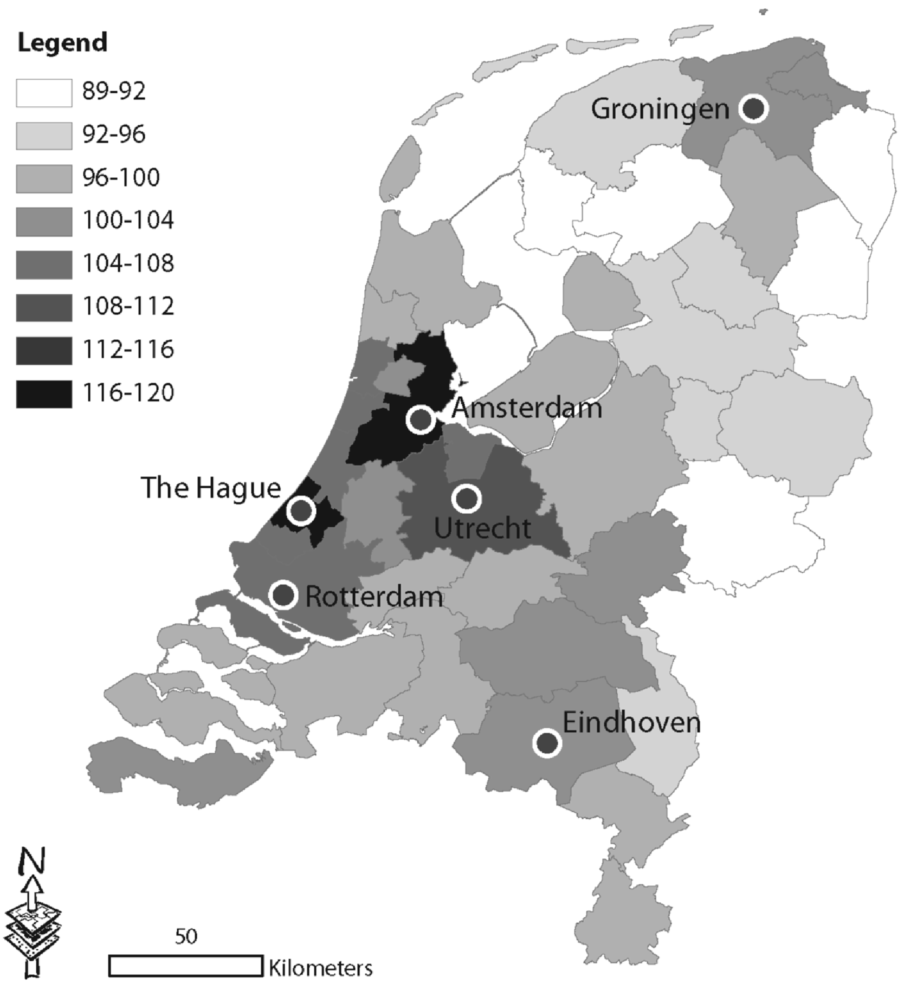
4. Regional wage differences in the Netherlands: stylized facts

Before turning to the econometric analysis, we present some stylized facts about wage differences between Dutch regions. Figure 1 shows average hourly wages per worker in each NUTS-3 region. (The corresponding names of the Dutch NUTS-3 classification are included in Appendix A.) On average, employees working in the Amsterdam agglomeration receive the highest hourly wages, while those in the North-Eastern part (Zuidwest-Friesland) earn the lowest. In general, wage levels are higher in the western

[7] As pointed out in Combes et al. (2008b), when the area of regions has already been included as a separate variable, employment can – with proper reinterpretation of the coefficients – be included directly in the equation without subtracting the log of the area.

provinces of the Netherlands – mainly in the Randstad area – than in the rest of the country. This is partly explained by a relatively strong concentration of highly educated people in the Northern wing of the Randstad area and the agglomeration of The Hague (the residence of Dutch parliament and the political centre of the country). But at the same time, these are also the areas with by far the largest employment density as well as areas with clear natural advantages. The difference would have been even more pronounced if we would have included the highest incomes (over 10 times the overall average), since many (international) headquarters are located in the largest cities in the Randstad.

Figure 1. Indexed average hourly wages by NUTS-3 region, 2000–2005



5. Estimation of the Mincer equation

In Section 2, we discussed three broad explanations for regional wage disparities. The composition of regional labour markets relates to the characteristics of individual workers that live in a region, whereas regional endowments and agglomeration economies result in higher productivity for a given labour market composition. A method that is often used in economics to analyse wage differences is the Mincerian wage regression (cf. Mincer, 1974).[8]

We use dummy variables for the highest qualification that was obtained by workers. This allows for differences in the quality of education. The regression equation is formally denoted by the following equation (where *D* are dummies):

$$\log(w) = \alpha + \sum_{edu=1}^{8} \beta_{1,\mathrm{edu}} D_{\mathrm{edu}} + \beta_2 age + \beta_3 age^2 + \beta_4 D_{\mathrm{gender}} + \beta_5 D_{\mathrm{immigrant}} + \beta_6 D_{\mathrm{parttime}} + \sum_i \beta_{7,i} D_i + \sum_r \beta_{8,r} D_r + \varepsilon \tag{5}$$

The estimated skill premiums do not perfectly reflect the effect of education, as this is also related to unobserved variables like ability. So we use education as a proxy for the 'knowledge' of a worker in an attempt to estimate how the knowledge of workers is rewarded. It is, however, important to bear in mind that to the extent that individuals cluster in space according to the non-observed heterogeneity, our estimated spatial residual in part captures this clustering. We leave it for further research to address this issue by, for example, following people over time.

To give an indication of the relation between key variables, Table 1 presents some simple correlations. Strong correlations exist between average wages, average education attendance, the share of highly educated workers (e.g., those with at least tertiary education) and the share of managers and professionals in the working population.

[8] Although often applied, it should be noted that the causal relationship between the variables in the Mincer regression and the wages earned is actually not very strong. There exists an extensive literature on this subject, often using instrumental variable (IV) estimation methods in natural experiments where an exogenous shock affects the wages at a specific moment. Some of the contributions to this literature are Griliches (1977); Ashenfelter et al. (1999); Heckman et al. (2003) and Webbink (2004). Elaboration on this in the context of this study is left for future research.

Table 1. Simple correlations

	(1)	(2)	(3)	(4)	(5)
(1) Average paid wages	1.000				
(2) Spatial residual	0.847	1.000			
(3) Average years of education	0.762	0.454	1.000		
(4) Share of higher educated workers	0.767	0.463	0.985	1.000	
(5) Share of managers and professionals	0.755	0.501	0.944	0.972	1.000

Note: The spatial residual and its estimation is the topic of Section 6. It represents the part of the wages that is not explained by (available) individual worker characteristics.

Table 2 presents the regression results for the years 2000 and 2005, and for a regression on the combined cross sections from 2000 to 2005. We find that the impact on wages of the different worker characteristics that were evaluated has remained fairly constant during the reference period. Estimated coefficients for the skill and experience premiums are comparable to the values that are generally estimated in the literature.

There is a moderate relationship with average regional wages, as Table 1 indicated, but there are notable exceptions to this. The experience premium is estimated to be somewhat higher in cities than in less densely populated areas, showing a pattern that is comparable to the distribution of average wages across space. A possible explanation is that the observed differences are the result of differences in the type of jobs in the different regions.

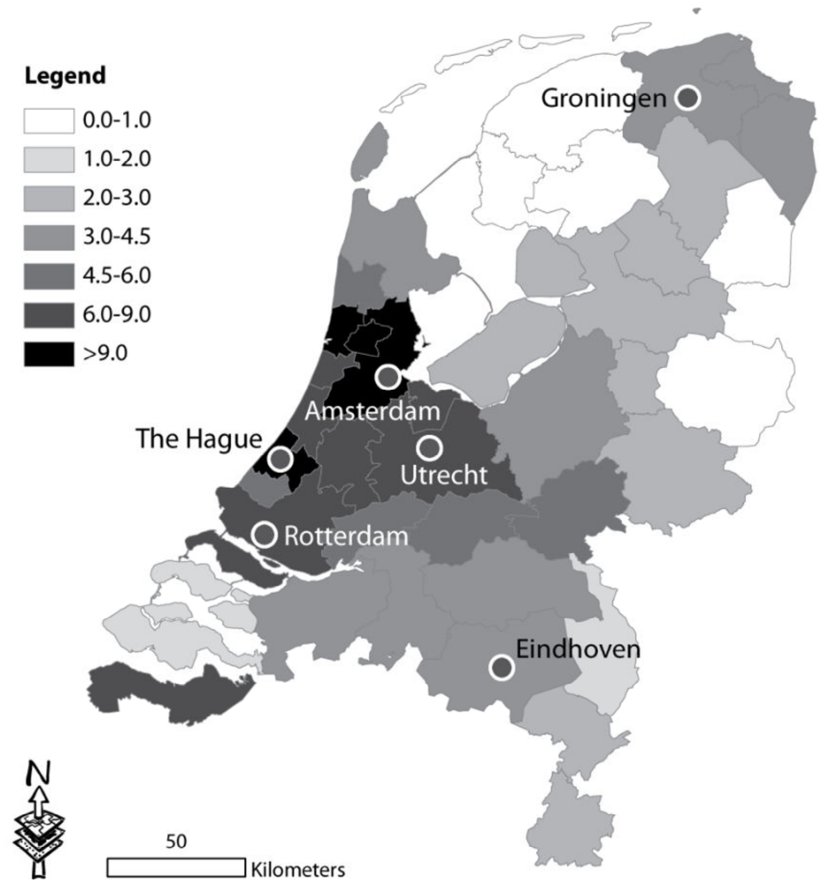
Table 2. Mincer regression (dependent variable: log of individual wage)

	2000	**2005**	**Panel 2000–2005**
Age	0.066	0.062	0.064
	(54.6)	(69.3)	(151.6)
Age-squared	–0.0006	–0.0006	–0.0006
	(42.1)	(54.9)	(119.2)
Female	–0.125	–0.116	–0.120
	(27.5)	(33.5)	(74.2)
Immigrant	–0.097	–0.096	–0.093
	(14.1)	(18.4)	(37.7)
Part-time worker	–0.087	–0.108	–0.100
	(19.0)	(31.4)	(62.5)
Education dummies[*]			
Lower secundary education (vmbo, mbo 1)	0.056	0.060	0.056
	(5.8)	(7.4)	(15.6)
Higher secundary education (havo, vwo)	0.256	0.256	0.257
	(25.2)	(30.4)	(69.1)
Lower tertiary education (mbo 2 + 3)	0.176	0.192	0.182
	(18.9)	(24.2)	(52.5)
Lower tertiary education (mbo 4)	0.269	0.270	0.268
	(28.49)	(35.0)	(78.3)
Higher tertiary education (hbo, BA)	0.452	0.472	0.466
	(48.3)	(61.3)	(136.5)
Higher tertiary education (MA, PhD)	0.662	0.681	0.678
	(64.6)	(83.6)	(184.6)
Industry dummies	yes	yes	yes
Year dummies	no	no	yes
Region dummies	yes	yes	yes
R^2	0.58	0.56	0.58
Number of observations	25,212	48,601	213,940

Note: *t*–statistics (in absolute values) are reported between parentheses. [*] Education dummies denote the highest qualification obtained, with as omitted category those individuals who have only primary education.

The region dummies estimated for each NUTS-3 area represent a spatial residual. Since we left out the dummy for the region with the lowest residual (Zuidwest-Friesland), the estimated spatial residuals can be interpreted as the premium that workers with equal characteristics can earn in a region, relative to the omitted region. Zuidwest-Friesland also happens to be the region were the lowest average wages were paid. The highest premium is paid in the Amsterdam region. When we compare the results presented in Figure 2 with those on the distribution of average wages presented in Figure 1, knowing that according to Table 1 average paid wages are very strongly correlated to the spatial residual, it can be observed that the spatial residual is generally higher in the Randstad.

Figure 2. Average spatial residual by NUTS-3 region (2000–2005)



6. Explanation of the spatial residual

This section aims to explain the spatial residual, following the approach developed in Combes et al. (2008a). The spatial residual is the part of variation in wages that is not explained by employee characteristics. We do this by applying a two-stage regression approach. To exploit time variation and increase the number of observations, we will use a panel of cross sections instead of single years. In the first stage we re-estimate the Mincer equation (4), but instead of region dummies we include a dummy γ for each combination of industry, region and year:

$$\log(w) = \alpha + \sum_{edu=1}^{8} \beta_{1,edu} D_{edu} + \beta_2 age + \beta_3 age^2 + \beta_4 D_{gender} + \beta_5 D_{immigrant} + \beta_6 D_{parttime} + \sum_i \sum_r \sum_t \gamma_{irt} D_i D_r D_t + \varepsilon \quad (6)$$

We repeat the same procedure using municipalities instead of NUTS-3 regions. Estimated coefficients are comparable to those estimated in the previous section.

In the second stage of our analysis we explain the resulting residual (a premium paid to employees that change jobs to a certain industry and region) by a set of geographical variables, to test for the presence of different types of agglomeration externalities. Here we include the total number of employees in each region (urbanization effect), employment in each industry and region (specialization effect), surface area, Shannon's entropy (Jacobs diversity effect), and a Hirschman-Herfindahl based index on the distribution of employment over firms (Porter competition effect). The instruments we use to measure agglomeration forces have been introduced more extensively in Section 2. We estimate the following regression:

$$\begin{aligned}\gamma_{irt} = \alpha + \beta_1 Density_{rt} + \beta_2 Specialization_{irt} + \beta_3 Diversity_{rt} + \\ \beta_4 Competition_{irt} + \beta_5 \log area_{rt} + \textstyle\sum_i \beta_{6,i} D_i + \sum_t \beta_{7,t} D_t + \varepsilon_{irt}.\end{aligned} \tag{7}$$

The interpretation of the results presented in Table 3 is that – according to the Mincer equation residuals on the NUTS-3 level – doubling the density of employees working in a region is associated with a 3.8% higher wage on average. Doubling the share of an industry in a region results in a 2.4% higher wage for the workers in that region. Additionally, we find a statistically significant negative relation between the residual wage component and both competition and diversity, contradicting the presence of Porter and Jacobs externalities (and consistent with insights from the efficiency wage literature; see, for example, Krueger and Summers, 1988, amongst many others).[9]

[9] Note that the *t*-values are low for both diversity and competition, considering the large number of observations.

Table 3. Explaining the spatial residual

	NUTS-3 regions	**Municipalities**
Density	0.038	0.021
	(13.6)	(14.9)
Specialisation (industry share)	0.024	0.023
	(8.2)	(13.1)
Diversity (Shannon's entropy)	–0.078	–0.042
	(2.9)	(5.2)
Competition (1–HHI)	–0.068	–0.012
	(4.5)	(1.5)
Log(area)	0.013	0.011
	(3.7)	(5.9)
Industry dummies	yes	yes
Year dummies	yes	yes
R^2	0.32	0.19
Number of observations	7,747	28,048

Note: *t*–statistics (in absolute values) are reported between parentheses.

To see how the effects change for a more detailed level of regional aggregation, we repeated our analyses using municipalities instead of NUTS-3 regions. Comparing the residuals from the Mincer equation estimated for NUTS-3 regions with those for municipalities, we see that the better an estimation strategy accounts for regional heterogeneity in worker characteristics, the less agglomeration effects are found. For municipalities, the employment density-wage elasticity is 0.021, where it was 0.038 for the NUTS-3 regions. This could be due to the fact that differences between municipalities are much larger than those that exist between local labour market areas, and that we capture a different type of effects than those that are the topic of this paper.[10]

In underlining the importance of using micro-data for the identification of agglomeration externalities, Combes et al. (2008a and 2008b) have pointed out that aggregate regional data (especially average sectoral composition and worker characteristics) do not sufficiently correct for worker and firm characteristics. The remaining unobserved heterogeneity will result in an upward bias when estimating agglomeration economies. Melo et al. (2009) support this observation in their meta-

[10] As differences in density and economic activities can be very large at the municipality level (for example, a small village can neighbour a large city), the agglomeration parameters might be identified on differences between cities and the countryside instead of more moderate differences that exist between local labour market areas. As variety in density is larger than variety of average wages, this is likely to result in a lower elasticity between employment density and average wages.

analyses of 34 studies, finding that the use of macro-data generally results in higher elasticities than the use of micro-data, as does Smit (2010) in a meta-analysis of 73 studies.

In view of this discussion, our data allow us to compare our work with previous research that did not use micro-data by aggregating our data to estimate agglomeration effects at the regional level. Regional averages of all variables used in the micro regressions were calculated directly from the micro-data. Table 4 (left) shows an employment density elasticity of 0.043 for NUTS-3 regions. This implies that doubling the number of workers on a given area results in a 4.3% increase of productivity. This figure is within the range of the 3–8% found in the meta-analysis of Melo et al. (2009), but much lower than the 18% found by Gorter and Kok (2009). Yet Gorter and Kok use production density instead of employment density, and the high elasticity that they find is most likely to be the result of the fact that they use aggregate data and do not correct for worker characteristics at all. If we do not include the average level of education and the average age in our macro specification, we also find a higher elasticity of 0.061. Estimating our equation using macro-data on municipalities results in an employment density elasticity of 0.024%, when correcting for average age and education. This much lower figure could again be due to unobserved heterogeneity when using the COROP classification.

Table 4. Explaining regional productivity differences (log of average regional wage)

	NUTS-3 regions	**Municipalities**
Average age	0.026	0.017
	(38.7)	(68.4)
Average education	0.062	0.076
	(51.7)	(119.5)
Density	0.043	0.024
	(66.6)	(59.3)
Specialization (industry share)	0.000	0.003
	(0.4)	(6.4)
Diversity (Shannon's entropy)	–0.077	–0.032
	(15.6)	(14.2)
Competition (1–HHI)	–0.012	–0.015
	(4.6)	(7.1)
Log(area)	0.011	0.015
	(17.1)	(28.2)
Sector dummies	yes	yes
Year dummies	yes	yes
R^2	0.82	0.61
Number of observations	7,747	28,048

Note: *t*–statistics (in absolute values) are reported between parentheses.

Even if a certain worker characteristic has been identified as a strong determinant of individual wages, its contribution to explaining wage differences can be limited if its distribution over space is more or less uniform. Figure 5 therefore presents some information about the economic implications of the findings presented in Tables 2 and 3. The economic implications are illustrated by multiplying the Mincer estimates (for both NUTS-3 regions and municipalities) with the standard deviation of the regional averages of each of the independent variables in the analysis. The one standard deviation gives us a reasonable proxy for the real variation across space of the explanatory variables.

Table 5. Economic impact on regional wage differentials derived from Mincer estimates

	NUTS-3 regions	**Municipalities**
Age	0.75	2.43
Share of highly educated workers	1.90	3.69
Share of part-time workers	–0.45	–1.12
Share of female workers	–0.40	–1.36
Share of immigrant workers	–0.24	–0.36
Density	3.50	2.65
Diversity (Shannon's entropy)	–0.69	–0.89
Area	0.99	1.03

Note: Impact is measures as percentage change of average wage resulting from a one standard deviation increase of the respective variables. Specialization and competition are not included in this table, as they are sector-specific measures. Detailed results for individual sectors are available upon request.

Even though the male-female wage gap is substantial (12 percent, according to Table 2), it has a relatively small economic impact due to the fact that the distribution of the share of females on the labour market is fairly uniform across regions. A one standard deviation increase in the share of females in a NUTS-3 region is associated with a 0.40 percent decrease in the average regional wage. As there is large regional diversity in the share of high skilled workers, while education is at the same time one of the most important wage determinants, variation in the share of high skilled workers has strong explanatory power. Also the employment density – even between NUTS-3 regions – can be considered as an important determinant of observed wage differences between regions. Due to the fact that variety in worker characteristics and agglomeration variables is larger between municipalities than between NUTS-3 regions, the economic impact of the estimated coefficients is in most of the cases larger for municipalities.

To conclude, Table 6 presents the expected and actual wages differences between urbanized and non-urbanized areas (as a percentage deviation). Urbanized areas are taken as the 22 agglomerations that are defined by Statistics Netherlands. The rest of the country is classified as non-urban. In total, the urbanized areas cover about 50% of the Dutch population. All figures are relative to the (weighted) average of municipalities outside the agglomerations. Expected wage differences were calculated by multiplying the coefficients that were estimated for municipalities with averages of the independent variables within each agglomeration (in deviation from their non-urbanized counterparts). The columns in the right part of Table 6 present the contribution of each component to the expected wage differential. On average, wages are 7 percent higher in agglomerations than in peripheral municipalities. The variables that explain the largest part of the expected wage differential between agglomerations and the periphery are the level of education and density. Other variables do not provide a structural explanation, with the exception of the share of non-native workers, which are relatively overrepresented in the large cities and earn a slightly lower wage *ceteris paribus*.

The explanatory power of the models that were estimated in this paper is relatively high. The correlation between actual wages and expected wages is 0.91 for the 22 agglomerations in Table 6, and 0.79 for all 467 Dutch municipalities. Amsterdam and The Hague have a large difference between the expected wage and the average wage. This suggests that these cities – the capital and the government seat – have something 'extra' that is not captured in any of the variables in our regression model.

Table 6. Economic implications for 22 agglomerations: decomposition of expected average wage differences with non-urbanized areas

Agglomeration	Expected Wage	Actual wage	Decomposition of expected average wage in different components										
			Gender	Non-natives	Part-time	Age	Education	Density	Diversity	Competition	Specialization	Area	Industry
Amsterdam	10.99	19.52	0.76	–0.72	0.81	0.41	4.75	4.41	0.04	–0.27	0.021	0.77	0.06
's-Gravenhage	11.08	19.22	0.47	–0.76	0.98	0.15	4.95	4.70	0.56	–0.25	0.302	–0.03	0.52
Utrecht	9.95	14.10	0.53	–0.27	0.42	–0.09	5.30	4.23	–0.10	–0.14	–0.108	0.17	0.63
Nijmegen	9.39	12.99	–0.11	–0.16	–0.30	0.80	5.25	3.84	0.38	–0.35	0.225	–0.19	–0.22
Amersfoort	8.26	9.88	0.16	–0.11	0.30	0.56	3.91	2.69	0.10	0.65	0.026	–0.02	–0.28
Rotterdam	5.18	9.77	0.30	–0.70	0.63	0.02	1.21	3.63	–0.12	–0.22	–0.167	0.60	0.25
Leiden	5.70	9.04	–0.67	–0.28	–0.54	–0.51	3.94	3.87	0.48	0.56	0.363	–1.51	0.30
Eindhoven	7.39	8.57	0.54	–0.23	0.70	0.12	3.59	3.17	–0.23	–0.21	–0.195	0.11	–0.24
Haarlem	5.50	7.71	–0.37	–0.22	–0.09	0.83	2.15	4.05	0.17	–0.15	0.123	–0.99	0.20
Groningen	8.03	7.02	0.13	0.08	–0.23	–0.42	3.88	3.16	0.24	0.77	0.189	0.24	–0.06
Arnhem	7.47	6.61	0.14	–0.03	–0.04	1.46	2.21	2.66	0.27	0.21	0.098	0.50	0.09
's-Hertogenbosch	5.49	6.61	0.32	–0.02	0.39	0.44	2.05	2.54	–0.10	–0.25	–0.103	0.21	0.22
Apeldoorn	2.00	5.61	0.50	0.11	0.39	–0.03	0.05	–0.05	–0.40	–0.28	–0.215	1.93	1.07
Maastricht	4.73	4.60	–0.11	–0.01	0.02	–0.47	2.45	3.22	–0.15	–0.04	–0.070	–0.12	–0.60
Geleen/Sittard	4.70	4.52	0.94	–0.05	0.55	2.18	–0.54	1.72	–0.59	0.71	–0.288	0.07	0.35
Breda	3.38	2.90	–0.17	–0.13	0.29	–0.10	1.65	2.03	–0.40	–0.37	–0.207	0.80	0.43
Tilburg	2.21	0.81	0.08	–0.10	–0.10	–1.13	1.01	2.29	–0.03	–0.29	–0.147	0.65	–0.77
Zwolle	3.50	0.63	–0.02	0.09	–0.31	0.16	0.95	1.95	0.05	–0.23	0.186	0.67	–0.42
Dordrecht	1.11	0.63	0.14	–0.13	0.18	0.15	–0.78	2.22	–0.18	0.24	–0.161	–0.59	0.57
Heerlen	3.75	0.17	0.48	0.00	0.56	1.24	–0.37	2.42	0.01	0.18	–0.109	–0.65	–0.19
Enschede	2.94	0.15	–0.02	–0.24	0.03	–0.10	1.71	1.28	–0.38	–0.16	–0.113	0.92	–0.05
Leeuwarden	3.55	0.11	–0.04	0.19	–0.35	–0.17	1.79	2.39	–0.20	–0.27	–0.070	0.27	1.40

Notes: Expected wage differences are based on the estimates of the Mincerian wage regressions for municipalities, and measured as percentage deviations from the average municipality outside the 22 Dutch agglomerations as defined by Statistics Netherlands. ‘Industry’ refers to sectoral composition.

7. Robustness

As we discussed in Section 2, there is discussion in the literature regarding the empirical proxies that are to be used to identify the importance of agglomeration externalities. In a meta-analysis, De Groot et al. (2009) found that different proxies can lead to substantially different results, *ceteris paribus*. The use of a location quotient to measure specialisation, for example, makes it more likely that a significantly positive agglomeration effect is found.

The proxies we used in the previous sections are our own preference, and also commonly used in the literature. We will now investigate the robustness of our results (and those found in the agglomeration literature more in general) by varying the specification of the agglomeration variables used in the second stage. Our original estimates (Table 3) included non-sector specific employment density and the area as urbanisation variables, the industry share for specialisation, a Hirschman-Herfindahl based index for competition, and Shannon's entropy as a diversity variable. We will now test three different proxies for specialisation, competition and diversity, which we will use in our estimations once together with urbanisation effects, and once without controlling for urbanisation. This results in 32 estimates for each proxy of one of the agglomeration variables. The variables chosen are presented in Table 7. For ease of comparison, the variables are defined such that a higher value corresponds to more specialisation, competition or diversity. Results of the robustness analysis are presented in the form of box-and-whisker plots in Figures 3 and 4. A similar plot for the urbanization variables can be found in Figure 5.

Table 7. Agglomeration variables and their correlations

type		variable	*S_ellis*	*S_lq*	*S_emp*	*C_hhi*	*C_fpe*	*C_firms*	*D_hhi*	*D_shann*	*D_glae*
SPEC	*S_ellis*	Ellison-Glaeser index	1.000								
	S_lq	Industry share	0.064	1.000							
	S_emp	Total employment in industry	-0.060	-0.007	1.000						
COMP	*C_hhi*	1 – HHI on firm employment shares	-0.057	-0.236	0.235	1.000					
	C_fpe	Firms per employee in sector-region	0.015	-0.200	-0.118	-0.045	1.000				
	C_firms	Number of firms in local industry	-0.065	-0.059	0.817	0.326	-0.091	1.000			
DIV	*D_hhi*	1 – HHI on industry employment shares	0.078	-0.030	-0.075	0.045	0.044	-0.039	1.000		
	D_shann	Shannon's entropy	0.066	-0.028	-0.041	0.053	0.016	-0.015	0.928	1.000	
	D_glae	1 – share of Glaeser's 'largest sectors'	0.113	0.049	-0.495	0.256	-0.072	0.423	0.576	0.581	1.000

Note: SPEC refers to the proxies for MAR-externalities, COMP to Porter externalities and DIV to Jacobs externalities.

We note that some variables are highly robust to the inclusion of other agglomeration variables: for example, the Ellison-Glaeser index *S_ellis* hardly varies across the 2×9 estimations, which is in line with its low correlation with the other variables (see Table 7). The spread of the results found is larger when urbanization variables are *not* included, and significance levels are higher. However, no variable changes sign (see Figure 4), although *S_ellis* and *D_glae* become statistically insignificant when controlling for urbanization.

Yet results differ widely within each group. For both competition and diversity, some proxies render quite consistently positive results (*C_firms*), while others show negative results (*C_cfpe* and *D_hhi*). These results confirm the findings of De Groot et al. (2009), who concluded in their meta-analysis that the specification of variables matters for the effect that will be found. This implies that even where some studies claim to look at the same variable, their results will actually not be comparable but depend on the proxies they included. There are a few proxies that give similar results between them, at least for our data and method: those are for example *S_emp* and *S_lq*, or, in most cases, all three diversity variables. Our results suggest that estimations from studies using these variables can sensibly be compared, *ceteris paribus.*

Figure 3: Box-and-whisker plot of repeated regressions with different specifications of the variables, controlling for urbanisation. Horizontal lines indicate t = –1.96 and t = +1.96; as usual, the small black lines indicate the mean, and dots represent outliers.

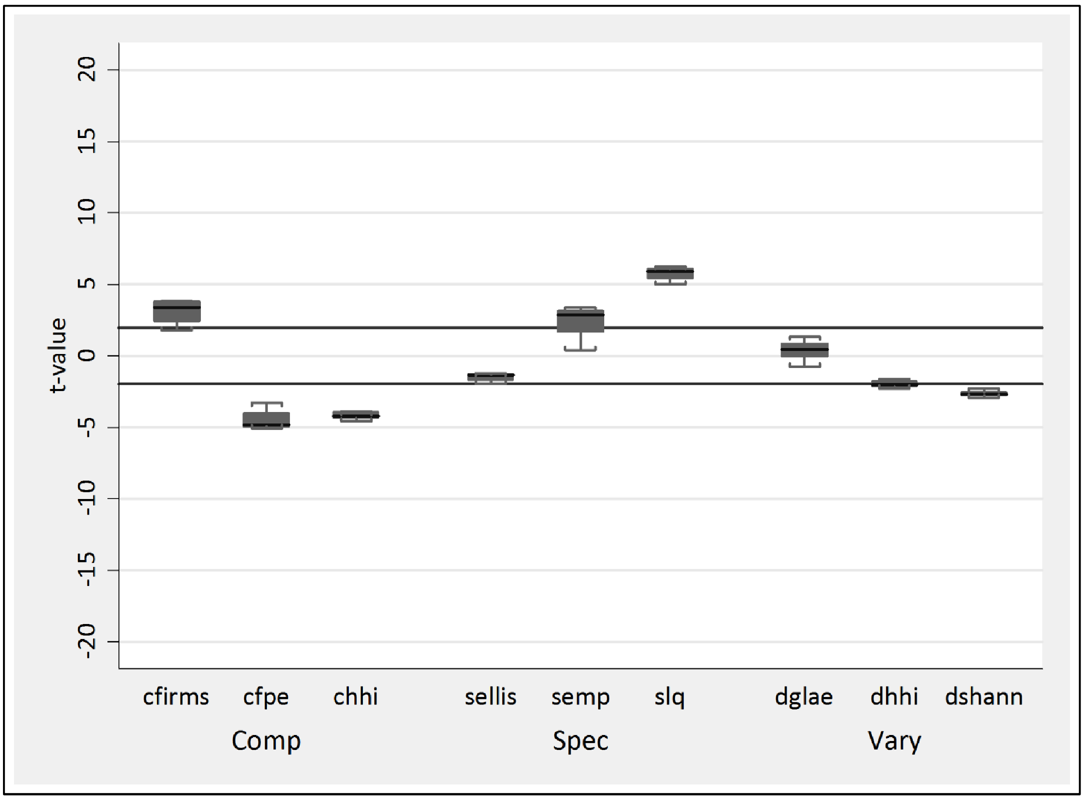


Figure 4: Box-and-whisker plot of repeated regressions with different specifications of the variables, not controlling for urbanisation.

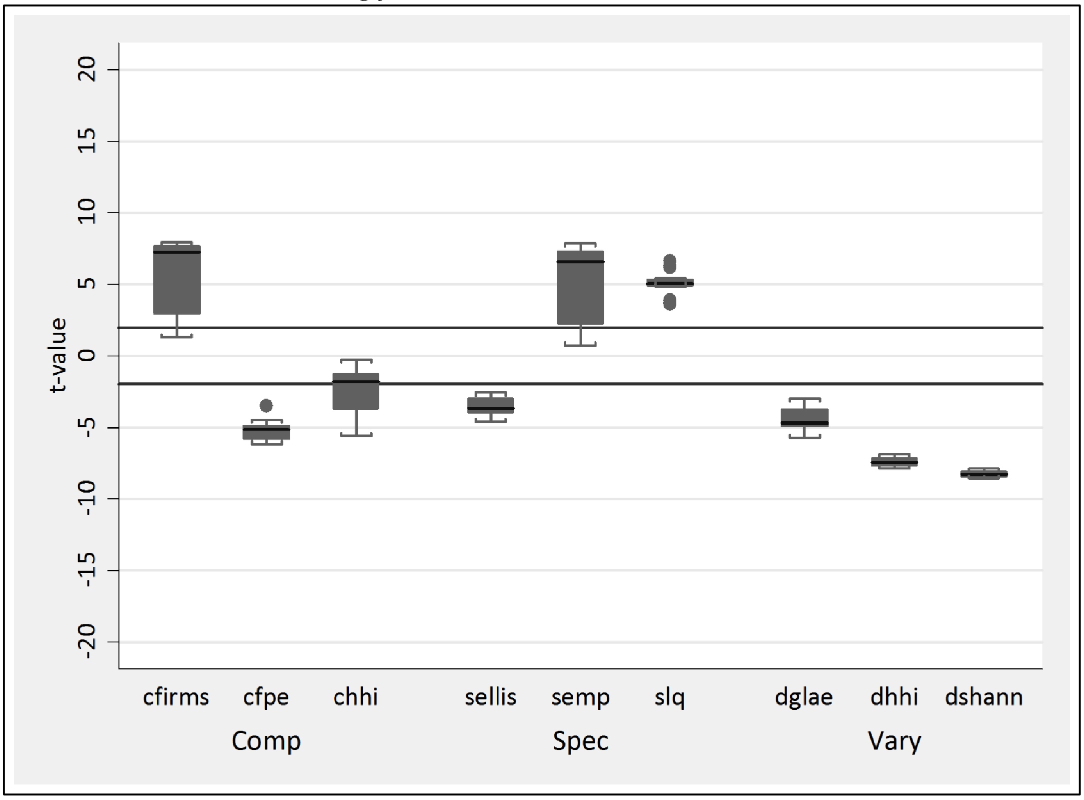
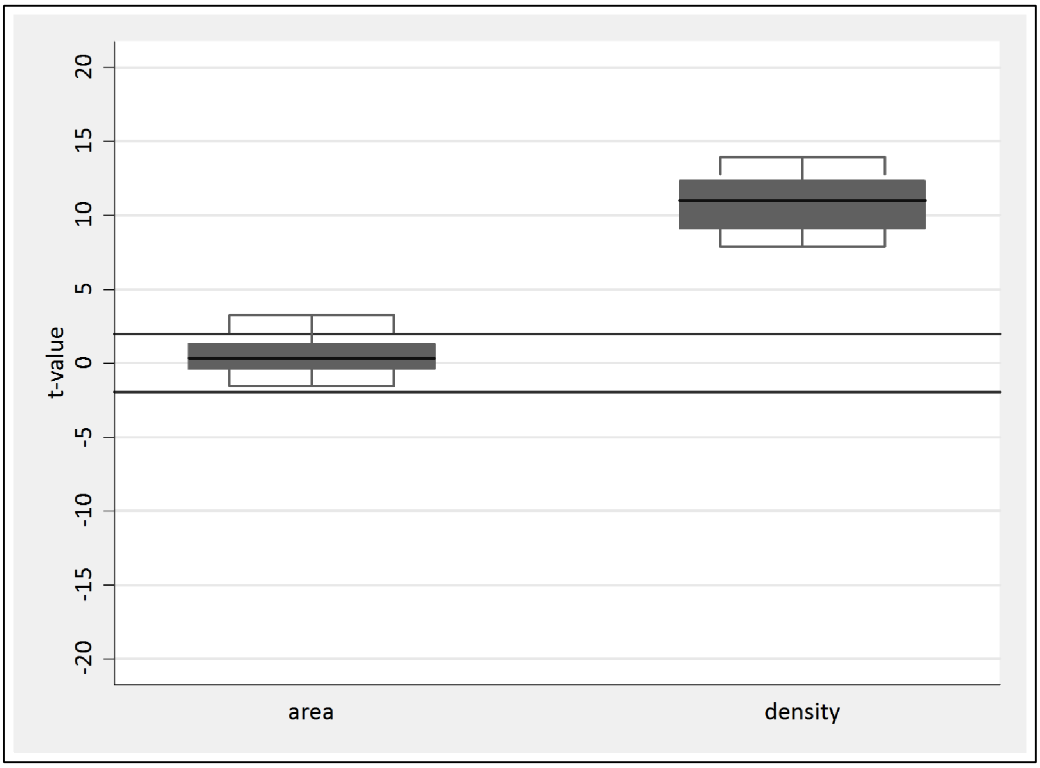


Figure 5: Box-and-whisker plot for urbanisation variables



8. Conclusions

The first part of this paper described differences in wages between NUTS-3 regions in the Netherlands. We confirmed that wages are substantially higher in the urbanized Randstad area than in the rest of the Netherlands. Also, average wages show a clear pattern of positive spatial association among neighbouring NUTS-3 regions. A geographical representation of the spatial residual showed that, after correcting for regional differences in human capital, workers in densely populated areas get paid a premium. The spatial residual, which is the regional average wage corrected for observed worker heterogeneity, is strongly correlated to average regional wage. At the heart of the analysis is the explanation of the spatial residual, i.e. the part of variety in wages that was not explained by employee characteristics, as a function of various agglomeration variables. We found that the total size of the regional labour market has a statistically significant and positive effect on wages, even though this explained a relatively small part of the residual wage component. Using the Mincer residuals on the NUTS-3 level, we found an employment density elasticity of 3.8%, and also clear evidence for the presence of MAR externalities. Doubling the share of an industry results in a 2.4% higher productivity. In our main specification, we find small evidence for *negative* effects of Porter and Jacobs externalities. However, we showed in an extensive robustness test that the specification of these variables matters to a large degree for the effects found.

The estimated agglomeration economies are lower than those estimated in previous work for the Netherlands by Gorter and Kok (2008), but correspond to what is found in the existing international literature by Combes et al. (2008a) and Melo et al. (2009). The current study for the Netherlands supports the finding that the size of estimates of agglomeration externalities are to a large extent determined by the ability of the data and methods that are used to correct for regional and individual heterogeneity. An issue that remains unaddressed in most of the current literature, and this article as well, is the endogeneity problem caused by local endowments. The presence of universities, infrastructure and local institutions all increase local productivity while being highly correlated with density. Due to a lack of good instruments, it has proven difficult to isolate the effect of density. Even though agglomeration externalities that have been estimated in the current literature are insightful, the causal relation between agglomeration and productivity will remain unclear until this endogeneity issue has been solved.

References

Altonji, J.G. and R.M. Blank (1999), ‘Race and gender in the labor market’, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, vol. 3C, Amsterdam: Elsevier, pp. 3143–3259.

Ashenfelter, O., C. Harmon and H. Oosterbeek (1999), ‘A review of estimates of the schooling/earnings relationship, with tests for publication bias’, *Labour Economics*, vol. 6, pp. 453–470.

Beaudry, C. and A. Schiffauerova (2009), ‘Who's right, Marshall or Jacobs? The localization versus urbanization debate’, *Research Policy*, vol. 38, pp. 318–337.

Briant, A., P.P. Combes and M. Lafourcade (2008), ‘Dots to boxes: do the size and shape of spatial units jeopardize economic geography estimations?’, CEPR Discussion Paper, no. 6928, London.

Combes, P.P., G. Duranton and L. Gobillon (2008a), ‘Spatial wage disparities: sorting matters’, *Journal of Urban Economics*, vol. 63, pp. 723–742.

Combes, P.P., T. Mayer and J.F. Thisse (2008b), *Economic geography: The integration of regions and nations*, Princeton University Press.

De Groot, H.L.F., R. Nahuis and P.J.G. Tang (2006), ‘Is the American model miss world? Choosing between the Anglo-Saxon and a European-Style Alternative’, in S. Munschenk, M. Stierle, U. Stierle-von Schütz and I. Traistaru (eds), *Competitiveness and growth in Europe*, Cheltenham: Edward Elgar.

De Groot, H.L.F., J. Poot and M.J. Smit (2009), ‘Agglomeration, innovation and regional development: theoretical perspectives and meta–analysis’, in R. Capello and P. Nijkamp (eds), *Handbook of regional growth and development theories*, Cheltenham: Edward Elgar, pp. 256–281.

Fishlow, A. (1972), ‘Brazilian size distribution of income’, *American Economic Review*, vol. 62, pp. 391–402.

Glaeser, E.L., H.D. Kallal, J.A. Scheinkman and A. Shleifer (1992), ‘Growth in cities’, *Journal of Political Economy*, vol. 100, pp. 1126–1152.

Gorter, J. and S. Kok (2009), ‘Agglomeration economies in the Netherlands’, CPB Discussion Paper, no. 144, The Hague.

Griliches, Z. (1977), ‘Estimating the returns to schooling: Some econometric problems’, *Econometrica*, vol. 45, pp. 1–22.

Harvey, D. (1981), ‘The urban process under capitalism: a framework for analysis’, in M.J. Dear and A.J. Scott (eds), *Urbanization and urban planning in capitalist society*, London: Taylor & Francis, pp. 91–122.

Heckman, J.J., L. Lochner and P. Todd (2003), ‘Fifty years of Mincer earnings regressions’, IZA Discussion Paper, no. 775, Bonn.

Jacobs, J. (1969), *The economy of cities*, New York: Random House.

Jaffe, A.B., M. Trajtenberg and R. Henderson (1993), ‘Geographic localization of knowledge spillovers as evidenced by patent citations’, *Quarterly Journal of Economics*, vol. 108, pp. 577–598.

Krueger, A.B. and L.H. Summers (1988), ‘Efficiency wages and the inter-industry wage structure’, *Econometrica*, vol. 56, pp. 259–293.

Mincer, J. (1974), *Schooling, earnings, and experience*, New York: NBER Press.

Melo, P.C., D.J. Graham and R.B. Noland (2009), ‘A meta–analysis of estimates of urban agglomeration externalities’, *Regional Science and Urban Economics*, vol. 39, pp. 332–342.

Morduch, J. and T. Sicular (2002), ‘Rethinking inequality decomposition, with evidence from rural China’, *Economic Journal*, vol. 112, pp. 93–106.

Oaxaca, R. (1973), ‘Male–female wage differentials in urban labor markets’, *International Economic Review*, vol. 14, pp. 693–709.

Porter, M.E. (1990), *The competitive advantage of nations*, Houndmills: Macmillan.

Rosenthal, S.S. and W.C. Strange (2004), ‘Evidence on the nature and sources of agglomeration economies’, in V. Henderson and J.–F. Thisse (eds), *Handbook of regional and urban economics*, Amsterdam: Elsevier.

Shannon, C.E. (1948), ‘A mathematical theory of communication’, *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656.

Smit, M.J. (2010), *Agglomeration and innovation: Evidence from Dutch microdata*, Amsterdam: Ph.D. Thesis VU University Amsterdam.

Webbink, D. (2004), 'Returns to university education: evidence from an institutional reform', CPB Discussion Paper 34, The Hague.

Appendix A. NUTS-3 (COROP) classification

Source: Wikipedia Commons (File:Coropgebieden plain.png)

Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (070) 3383 380

July 2011 | ISBN 978-90-5833-518-0