The Delimitation and Coherence of Functional and Administrative Regions

ROA-RM-2006/1E

F. Cörvers, M. Hensen and D. Bongaerts

Research Centre for Education and the Labour Market

Faculty of Economics and Business Administration
Maastricht University

Maastricht, January 2006

ISBN-10: 90-5321-427-5
ISBN-13: 978-90-5321-427-5
Sec05.046.doc

Contents

Abstract

The aim of this paper is to examine the coherence between and within functional as well as administrative regions in a labour market context. The larger the coherence of the local labour markets within the delimited regions, the larger the heterogeneity between the delimited regions is expected to be for particular measures related to the economy and the labour market. Contrary to previous delimitation studies we test for labour market coherence. The functionally defined regions are compared with the administratively defined regions with respect to four economic indicators: (i) income level; (ii) housing prices; (iii) employment rate, and (iv) unemployment rate.

It turns out that the administrative delimitation of the Netherlands performs, on average, equally well as the functional delimitation. The hypothesis that the municipalities within the administratively defined regions show less coherence than the municipalities within the functionally delimited regions, cannot be rejected. We find some minor evidence that the coherence is greater for the average income level of municipalities within functional regions than within administrative regions. It can be concluded that there is not much to be gained in labour market policies by using functional instead of administrative divisions of regional labour markets. Therefore we doubt the usefulness of other studies on functional delimitations of labour market regions. Finally, our results imply that it may be better for regional labour market policies not to use a highly differentiated division of regions for such a small country as the Netherlands.

Acknowledgement

The authors would like to thank Jacco Hakfoort, Marcel van Wijk and Christoph Meng for their valuable comments. The research has been granted by the Dutch Ministry of Economic Affairs.

1 Introduction

The definition of a regional labour market is very important in the light of regional labour market policies (Ball, 1980). In particular for research and policy-making purposes, the delimited areas should exhibit functional similarities. The economic diversity within an administratively defined region may be so large that comparison between regions is not justified. The decisions made concerning the planning, distribution and allocation of resources among the various regions derived, are not likely to be the most effective and meaningful relative to the decision that would be made if the underlying regional patterns were known (Amedo, 1968). Analyses in which regional (labour) markets are examined using administratively defined regions, do generally not inform about the use of other delineations.

The issue of regionalizing countries into functional regions can be regarded as a problem of how to aggregate areas into regions that exhibit some functional similarities. The dominant concept in defining functional regions is that of labour markets, as is illustrated by the substantial literature in this field by e.g. Andersen (2002), Coombes et al. (1986), Casado-Diaz (2000), Fox and Kumar (1965), Killian and Tolbert (1993). For the delimitation of functional labour market regions commuting conditions are used in most OECD countries (OECD, 2002). Commuting conditions like distance, closeness, commuting thresholds, travel times determine the magnitude of the commuting flows between areas. On the basis of commuting flows, a functional region can then be defined as a region in which a large proportion of the workers both live and work. If the administrative boundaries of regions do not follow the functionally linked labour market areas, labour market policies targeting administratively defined regions may be less effective.[1]

The well-known Travel-To-Work-Areas (TTWA) of the United Kingdom (see Coombes et al., 1986, and OECD, 2002) are the result of a delimitation procedure using the direct and indirect relationships between municipalities by analyzing the behaviour of individual commuters. These areas have been defined to analyze labour market phenomena, calculate unemployment rates, identify assisted areas for industrial policies and reorganize local government. However, the use of delimitations of functionally defined regions varies between countries (OECD, 2002). In Germany, the delimitation of local labour markets is used for structural analyses of labour markets, regional economic competitiveness, job opportunities and territorial disparities. The areas serve as functional territorial units in which the administrative bodies or firms can benefit from subsidies. Other countries use the delimitation of functional regions to carry out socioeconomic territorial analyses (Canada, France, Italy, Sweden and the United States). Norway also uses delimitations of functional regions for regional planning and forecasting.

The delimitation procedure of TTWAs was developed to generate the maximum possible number of areas with a self-containment level of at least 75% and a minimum size of the area of 3,500 resident workers. Within the area, at least 75% of the jobs should be fulfilled by the residents of that area (demand-side self-containment) and at least 75% of the

1. However, it may be very difficult for local planning authorities to set policy goals with regard to a different division of areas than the administrative division.

residents should work in the area (supply-side self-containment).[2] In addition, a 70% threshold was accepted if the size of the area exceeded 20,000 residents. The municipalities with the highest self-containment levels are selected as the starting point for the delimitation procedure. However, the determination of the threshold values determines to a great extent the number of local labour market areas defined. Lower threshold values would yield more local labour market regions, as a result of which the usefulness of the delimitation for policy-making may be reduced. Other absolute threshold values to select employment centres are used, for example, by Giuliano and Small (1991), who defined contiguous employment areas in the Los Angeles region of the U.S. as areas with at least ten workers per acre or more than 10,000 workers.

Van der Laan and Schalke (2001) argued that the use of situation-dependent absolute figures is responsible for different classifications that depend on the country and the period of analysis. To avoid the problems related to the use of absolute figures (see also Casado-Diaz, 2000) when defining TTWAs, they used relative instead of absolute criteria to delimit local labour market areas in the Netherlands. In the first step all municipalities with the largest outgoing commuting flow to the same employment centre are clustered, provided that these municipalities meet two criteria. The first criterion is that the largest outgoing commuting flow of the municipalities that are clustered to a particular employment centre should be more than 25% of all outgoing commuters. According to the second criterion, the difference between the two largest outgoing flows of two municipalities to the same employment centre should be 0.6 times the smaller flow for the particular municipality. By applying these criteria, it is intended to only cluster the municipalities where commuting is clearly concentrated. In the subsequent steps the municipalities clustered together are linked to their employment centres. Finally the remaining municipalities are linked to the existing clusters. This ultimately resulted in 31 functional regions for the Netherlands.

As is illustrated by the examples given above, the common methods to define functional regional labour markets usually apply absolute or relative criteria to select employment centres as a first step of the clustering procedure, such as the minimum absolute number of residents or workers in an area, or a minimum percentage of residents of an area working in the same area. However, the selection of absolute and relative criteria seems to be rather arbitrary and not satisfactory from a theoretical point of view. In the delimitation procedure of this paper, we do not require the use of these criteria. Moreover, we try to underpin our delimitation procedure by presenting a simple theoretical model of commuting on the labour market. Our aggregation approach includes a set of the smallest identified areas (e.g. municipalities), a set of criteria, an aggregation procedure and some insight into the evaluation (testing) of the results of the aggregation procedure by testing for the coherence of functional regions.

The aim of this paper is to examine whether the coherence of the functional labour market regions, which are carefully delimited without using arbitrary criteria, is larger than the coherence of the administrative defined labour market regions. The larger the coherence of the local labour markets within the delimited regions, the larger the heterogeneity between the delimited regions is expected to be for particular measures related to the economy and the labour market. The comparison of different regionalizations using economic indicators

2. Travel-To-Work-Areas (TTWA) in Spain were developed using the same thresholds. See Casado-Diaz (2000) for a more detailed description of the definition of TTWA.

has hardly been examined in the field of labour economics.[3] Baumann, Fischer and Schubert (1996) estimated the parameters of a multiregional labour supply model for Austria for different regionalizations. They concluded that the performance of the various models (e.g. for commuting and employment) they use are often critically dependent on the choice of a particular regionalization and of the relative size of the labour market regions (see also Openshaw, 1977). Therefore, we will attempt to make a contribution in this field by comparing different regionalizations for the Netherlands. This will be achieved by testing for labour market coherence, comparing the functionally defined regions with the administratively defined regions with respect to four economic indicators: (i) income level; (ii) housing prices; (iii) employment rate, and (iv) unemployment rate. For these indicators, we will test whether functionally defined regions show more coherence between the municipalities included in it than the administratively defined regions.

Following Brown and Holmes (1971) and Baumann, Fischer and Schubert (1996) we transform the standard interaction matrix of commuting flows between municipalities into a mean first passage time (MFPT) matrix. This implies that we cluster municipalities that have more interaction with each other than with municipalities outside the cluster. This aggregation method maximizes within-region commuting flows by merging the two adjacent municipalities (or clusters) with the smallest distances expressed by the mean first passage time indicators, that is the greatest mutual interaction in commuting flows. Setting the maximum number of clusters preferred, this aggregation method leads to optimal delimitation of functional regions. According to our theoretical model, low commuting flows between these functional regions should go hand in hand with large interregional differences in wages, housing prices, employment and unemployment rates.

This paper is structured as follows. The next section provides a commuting model of the labour market that underpins the empirical testing of regions on their coherence with regard to the economic indicators mentioned above. In Section 3, the delimitation method using travel-to-work data is derived. Furthermore, an application of this method is presented for different delimitations of regions in the Netherlands. Section 4 provides the testing of the functional and administrative regions using the four economic indicators. Lastly, in Section 5, we present a number of conclusions on the different functional and administrative delimitations.

2 A commuting model of the labour market

In this section we will analyse how commuting reduces wage inequality between regions. This is important to justify the delimitation of regions by commuting flows in the next section. It is hypothesized that the larger commuting flows are, the smaller wage inequality between regions, and the larger coherence between regions with respect to the four economic indicators mentioned before. The key question that can be raised is: How does commuting act as a labour market adjustment mechanism to equilibrate regional labour markets?

---

3. Van der Laan and Schalke (2001) compared different regionalizations using ratios of incoming and outgoing commuting flows. They did not really test on the performance of these regionalizations, since they did not use criteria that are independent of the commuting flows that were used for the delimitations.

Our starting point is a classical supply-demand representation of the labour market. In this model the labour demand function can be characterized by:

$$L_{D,t}^{R} = \beta_1^R - \beta_2^R * W_t^R \quad (1)$$

where W_t^R is the real average wage earnings per worker and R denotes regions A, B. Furthermore, t refers to the period before and after commuting is allowed, where in period 1 ($t = 1$) no commuting is allowed and in period 2 ($t = 2$) it is. The labour supply function can be characterized by:

$$L_{S,t}^{R} = \alpha_1^R + \alpha_2^R * W_t^R \quad (2)$$

When regional labour supply and demand in period 1 are equal, the labour market in region R is in equilibrium. That is, $L_{D,1}^{R} = L_{S,1}^{R}$. Setting (1) equal to (2) yields the equilibrium wage level for both regions (A and B) in period 1:

$$W_1^{R*} = \frac{\alpha_1^R - \beta_1^R}{-\beta_2^R - \alpha_2^R} \quad (3)$$

where $\alpha_2^R > 0$ and $\beta_2^R > 0$ are the wage elasticities of supply and demand, respectively. It follows that for W_1^{R*} to be positive, $\alpha_1^R < \beta_1^R$.

In period 2 commuting is allowed. Assume that in period 1 $W_1^{A*} > W_1^{B*}$. This will stimulate workers to commute from B to A. However, workers usually have to make costs to travel or to cross natural or psychological barriers. These costs are related to the so-called commuting conditions mentioned in Section 1. Suppose that individuals are confronted with fixed costs F if they commute from region B to region A. The equilibrium wage level (if commuting between A and B takes place), becomes:

$$W_2^{A*} = W_2^{B*} + F \quad (4)$$

where $F \geq 0$ and $W_2^{A*} < W_1^{A*}$ and $W_2^{B*} > W_1^{B*}$. The higher wage level in region A in period 1 leads to a commuting flow from region B to A, which in turn decreases the wage level in region A and increases the wage level in B. From equations (1), (2) and (4) it follows that the difference in the supply function for region A between periods 1 and 2 can be characterized by:

$$L_{S,2}^{A} - L_{S,1}^{A} = \alpha_2^A \ (W_2^{A*} - W_1^{A*}) \leq 0 \quad (5)$$

This follows directly from $W_2^{A*} < W_1^{A*}$.The same holds for the demand function for region A:

$$L_{D,2}^{A} - L_{D,1}^{A} = -\beta_2^A \ (W_2^{A*} - W_1^{A*}) \geq 0 \quad (6)$$

The commuting flow L_C^A from region B to region A is equal to the difference between labour demand and supply in region A *(or B)*. In period 2, labour demand is larger than labour supply, due to the decrease in the wage level in region A, and vice versa for region B. For region A this is illustrated by equation (7).

$$\begin{aligned} L_C^A &= (L_{D,2}^A - L_{D,1}^A) - (L_{S,2}^A - L_{S,1}^A) \\ &= -(\beta_2^A + \alpha_2^A)(W_2^{A*} - W_1^{A*}) \\ &= -\gamma^A (W_2^{A*} - W_1^{A*}) > 0 \end{aligned} \tag{7}$$

L_C^A is larger than 0 since γ^A>0 and $W_2^{A*} < W_1^{A*}$. The parameters γ^A and γ^B can be interpreted as the sensitivity of commuting flows to wage adjustments within a region. A large γ refers to high wage elasticities of supply and demand. By definition it holds that:

$$L_C^A + L_C^B = 0 \tag{8}$$

Therefore:

$$\begin{aligned} -\gamma^A (W_2^{A*} - W_1^{A*}) - \gamma^B (W_2^{B*} - W_1^{B*}) = 0 \\ \frac{W_1^{A*} - W_2^{A*}}{W_2^{B*} - W_1^{B*}} = \frac{\gamma^B}{\gamma^A} \end{aligned} \tag{9}$$

Equation (9) implies that regions with relatively low wage elasticities are confronted with relatively large changes in the regional equilibrium wage due to commuting. The equilibrium wage levels in period 2 for region A and B are identical if there are no fixed costs. To commute from region B to region A, workers have to incorporate the fixed costs F. Therefore, the difference between the equilibrium wage levels in period 2 consists of the fixed costs F (see equation (4)). Combining (9) with (4) results in:

$$W_2^{A*} = \frac{\gamma^B}{\gamma^B + \gamma^A}(W_1^{B*} + F) + \frac{\gamma^A}{\gamma^B + \gamma^A} W_1^{A*} \tag{10}$$

Equation (10) shows that the new equilibrium wage level in region A is the weighted average of the old equilibrium wage levels in regions A and B corrected for fixed costs. The region with the largest wage elasticities has the largest weight. From equation (7) it follows that:

$$W_2^{A*} = \frac{-L_C^A}{\gamma^A} + W_1^{A*} \tag{11}$$

Combining the result above with equation (10) leads to the equation for the commuting flow from B to A.

$$L_C^A = \frac{\gamma^B \gamma^A}{\gamma^B + \gamma^A}(W_1^{A*} - W_1^{B*} - F) > 0 \tag{12}$$

If fixed costs F decrease, the commuting flow from region B to region A increases. The commuting flow is maximized if F = 0. Starting in period 1, it follows that there will be no commuting if the fixed costs F are too large, i.e. $W_1^{A*} \leq W_1^{B*} + F$. Thus, commuting between regions A and B starts only if $W_1^{A*} - W_1^{B*} > F$ and stops in period 2 if $W_2^{A*} - W_2^{B*} = F$ (see equation (4)). This implies that:

$$(W_1^{A*} - W_1^{B*}) - (W_2^{A*} - W_2^{B*}) > 0 \tag{13}$$

It follows that the difference in equilibrium wage levels between regions A and B is larger before than after commuting. The better the infrastructure and the smaller the natural and psychological barriers between regions, the smaller the difference in wage levels between regions become due to commuting. The wage variance between regions is expected to be smaller if larger commuting flows are observed. Moreover, from (12) it follows that for given fixed costs and equilibrium wages in period 1, higher wage elasticities of demand and supply in regions A and B results into a larger commuting flow from B to A. This implies that if regional labour supply and demand are responsive to wage changes, commuting flows between regions are large.

In Section 4, we want to test the relationship between the wage variance (in period 2) and the commuting flows. This relationship is predicted by the next equation, which can be derived by combining equations (8) and (11) for regions A and B:

$$\begin{gathered} W_2^{A*} - W_2^{B*} = \\ \left(\frac{-L_C^A}{\gamma^A} + W_1^{A*}\right) - \left(\frac{-L_C^B}{\gamma^B} + W_1^{B*}\right) = \\ -\frac{\gamma^B + \gamma^A}{\gamma^B \gamma^A}(L_C^A) + (W_1^{A*} - W_1^{B*}) \end{gathered} \tag{14}$$

From equation (14) it can be readily understood that the interregional wage difference in period 2, which is equal to the fixed costs of commuting as follows from equation (4), is dependent on the interregional wage difference in period 1 and the wage elasticities, which are both exogenously determined. It follows that for given wage elasticities and interregional wage differences when regions are closed, the magnitude of the commuting flows is negatively related to the interregional wage differences when regions are open. The larger the commuting flows between regions, the lower the interregional wage differences. Both commuting flows and interregional wage differences reflect the commuting conditions mentioned before.

In the next sections we will analyse commuting flows between more than 400 municipalities instead of the two regions in the theoretical model. However, this does not alter the conclusions of this section. According to the clustering procedure of the next section, municipalities have a large chance on being clustered when the commuting flows between these municipalities are large. It follows that the wage differences between municipalities belonging to the same region should be low relative to the wage differences between municipalities of different regions. Since higher wages will raise housing prices and will pull more individuals to the labour market, higher costs of commuting may also be reflected in larger interregional differences in housing prices and labour participation. Finally, since job searchers face relatively high costs of commuting between municipalities of different regions, low commuting flows between regions may be related to large interregional differences in unemployment rates.

3 Regionalization based on travel to work flows

3.1 Commuting data

For the delimitation analyses, we have used the travel-to-work[4] data (2001, 1991 and 1992) from Statistics Netherlands, which observed the travelling behaviour of a sample from the Dutch population. This travelling behaviour can be classified according to the motivation of the mobility decision. Apart from the decision to travel to work, other motives to travel are also observed, such as shopping or sports. To delimit the Netherlands, only the home-to-work journey is used as a motive for the mobility decision. Table 1 presents some general descriptives of the background characteristics of the workers and some indicators of mobility for 2001 and 1991.

Gender, educational level, age and working hours are dummy variables for the different classes. It appears that about 60% of the sample in 2001 was, while in 1991 68% was male. Furthermore, the educational level of the Netherlands has been classified into low, middle and high. The low educational level consists of the workers who only completed primary school (Basisonderwijs) and general and vocational lower secondary education (VMBO). The middle class of educational level includes general and vocational higher secondary education (HAVO/VWO and MBO). The higher educational level consists of higher vocational education (HBO) and university education (WO). It appears that the educational level of the workers has increased over the last decennium. In 1991, about 20% of the workers had a higher degree; by 2001, this percentage has increased to more than 30%. The age structure of the sample working people shows a large number of workers between the ages of 25 and 55. More than 80% of the workers worked more than 30 hours per week in 2001, whereas in 1991 more than 85% of the workers worked more than 30 hours per week. The number of part-time workers therefore increased over the last ten years. The average income that workers earned is measured by their personal (net) income. The

4. OVG (Onderzoek VerplaatsingsGedrag). Since the number of observations was substantially lower for 1991, we also used the OVG data for 1992 as if the data sets are from one year. In the remainder of the paper we will refer to 1991 when we use the data of both years. See Corpelijn and Heerschop (2002) for more details on the commuting flows in the Netherlands.

average travel distance the workers travelled to reach their work location, was about 16 kilometres in 2001. In 1991 workers travelled on average 13 kilometres. The commuting behaviour of workers has therefore changed over time. Hence the delimitation of regions may have changed over time.

Table 1
Descriptives of the workers in 2001 and 1991

Descriptives		2001		1991	
		Mean	Std. Dev.	Mean	Std. Dev.
Gender	Male %	61.58	48.64	68.00	47.00
	Female %	38.42	48.64	32.00	47.00
Educational level	Low %	32.09	46.68	43.06	49.52
	Middle %	35.98	47.99	37.43	48.40
	High %	30.37	45.99	19.51	39.63
Age	18-24 years	9.47	29.28	12.90	33.53
	25-29 years	12.95	33.58	16.87	37.45
	30-39 years	28.70	45.24	30.91	46.22
	40-49 years	27.93	44.87	26.35	44.06
	50-65 years	20.94	40.69	12.97	33.60
Working hours	Less than 30 hours	18.82	39.09	13.79	34.48
	More than 30 hours	81.81	39.09	86.21	34.48
Income	Euros (€)	19,621.46	7,991.07	11,090.39	6,420.99
Travel distance	Kilometres	16.29	23.55	13.56	19.60
Travel time	Minutes	25.73	24.65	21.30	20.18
Observations	Number	39,298.00		5,875.00	

Source: OVG, Statistics Netherlands

3.2 Methodology

The smallest geographical units of analysis in the data set are municipalities (489).To delimit regions, a commuting matrix was used of incoming and outgoing commuting flows between all the municipalities. Furthermore, an algorithm was used to cluster the municipalities.[5] The commuting matrix in which all outgoing and incoming commuting flows are observed between all possible combinations of municipalities, was transformed into an MFPT matrix (mean first passage time). The MFPT method models relative travel to work flows - so the size of a community does not play a role - as a stochastic process. The benefit of this method is twofold. First, it can be used as a basis for a clustering algorithm. Second, it gives a hierarchical measure of clusters with respect to the amount of total interaction they have.

All municipalities are origins and destinations of commuting flows. The (weighted) numbers of movements between the municipalities were transformed to probabilities by dividing the number of flows by the columns or row totals. This MFPT matrix was used to derive a 'distance' matrix, where 'distance' is interpreted as follows: If both the direct and indirect distance between municipalities is low, these municipalities will have more interaction with each other than with municipalities with a higher value for distance. Distance is thus transformed into an indication for the willingness of workers to commute to a certain

5. See Baumann, Fischer and Schubert (1996) for an overview of algorithms.

municipality. Together with the restriction of contiguity of municipalities, this leads to a delimitation of regions for the Netherlands. Appendix A discusses the details of the MFPT matrix and the clustering procedure.

3.3 Results of the delimitations with commuting flows

The method described above allows us to produce any number of functional regions. To test for the coherence of the administrative and functional regions, the number of functional regions to be generated in the delimitation procedure has been set equal to the number of administratively defined regions in the Netherlands.

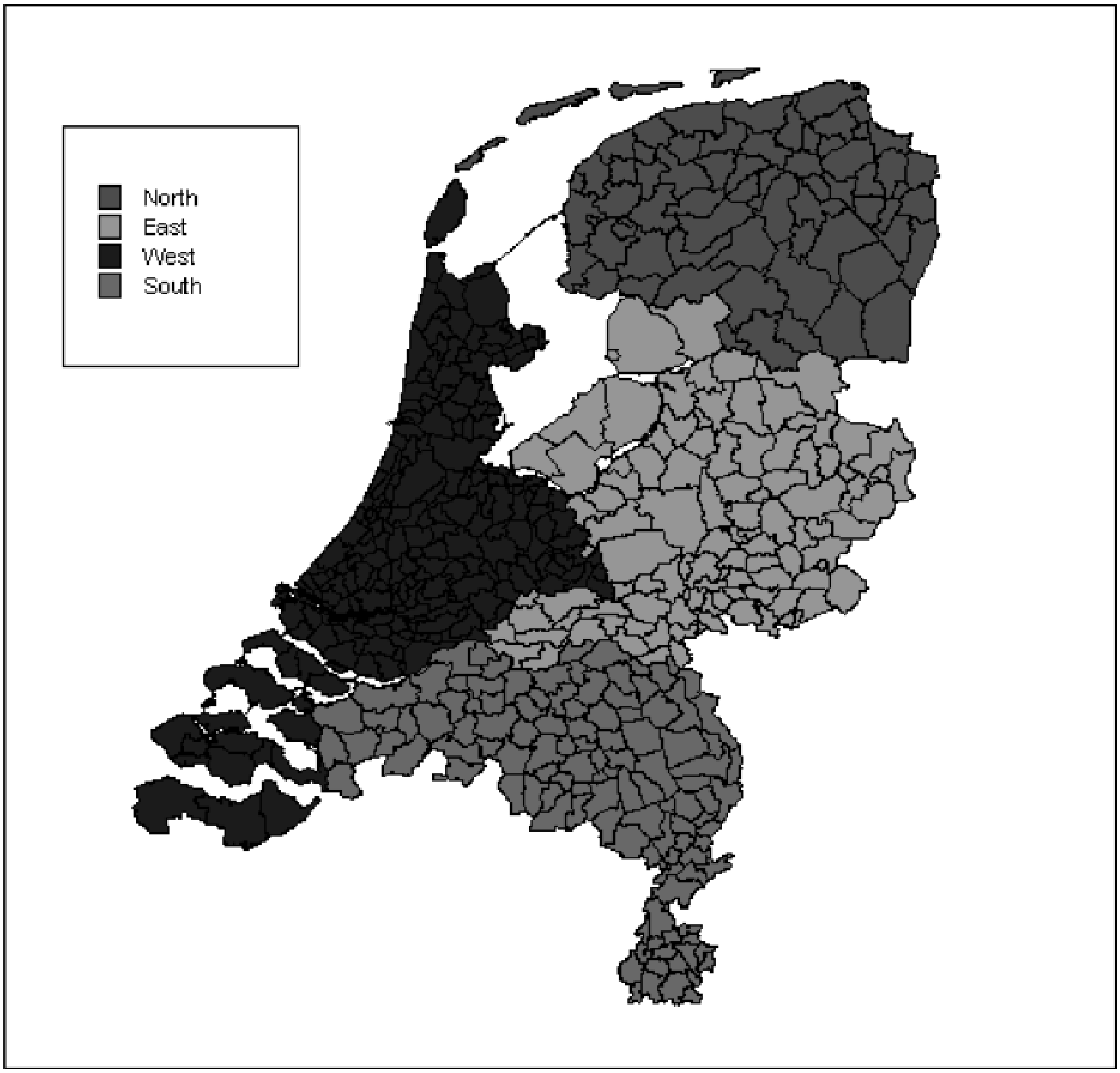
Eurostat uses an administrative division called the Nomenclature of Territorial Units for Statistics (NUTS). According to NUTS1 the Netherlands is divided into 4 administrative regions: north, south, east and west. Figure 1 shows this administrative delineation of the Netherlands. Since the Netherlands is divided into 4 administrative regions in NUTS1, we generated 4 functional regions in accordance with the delimitation procedure from the preceding section.

Figure 2 presents the derived delineation for 4 functional regions in 2001. The functional division into 4 regions of the Netherlands is evidently different from the administrative 4-region division. It appears that the Utrecht region is a separate regional labour market according to the delineation into functional regions. Furthermore, the Zeeuws-Vlaanderen region, consisting of 3 municipalities, can also be seen as a separate – more homogeneous – region. This can be easily explained by the absence of a bridge or a tunnel across the Westerschelde estuary to connect Zeeuws-Vlaanderen and Zuid-Beveland.[6] Moreover, the functional delineation based on commuting flows suggests that the northern part of the Netherlands interacts more with the western and the middle part of the Netherlands than is suggested by the administrative division. The same is true for the southern part. Apparently there is more north-south than east-west distinction between regions. This can be partly explained by the river Rhine flowing from east to west into the North Sea. Probably related to the course of the Rhine, the border between the functional regions Utrecht and South is almost the same as in the administrative 4-division or the administrative 12-division of provinces (see below). However, in the west the border between the functional regions North and South follows the administrative border between the provinces North Holland and South Holland instead of the Rhine.

The regional division of the Netherlands at the NUTS2 level refers to the 12 provinces of the Netherlands, which fall within the boundaries of the NUTS1 regions. These provinces represent important governmental bodies between the national government and the municipalities. A large share of the regional budgets for policy planning is distributed over the provinces. The division of the Netherlands into 12 functional regions have been compared to the Dutch division in 12 provinces.

6. In 1991 and 2001 two car ferries were running across the Westerschelde estuary. In 2003 the Westerschelde tunnel was put into use and the car ferry services were stopped.

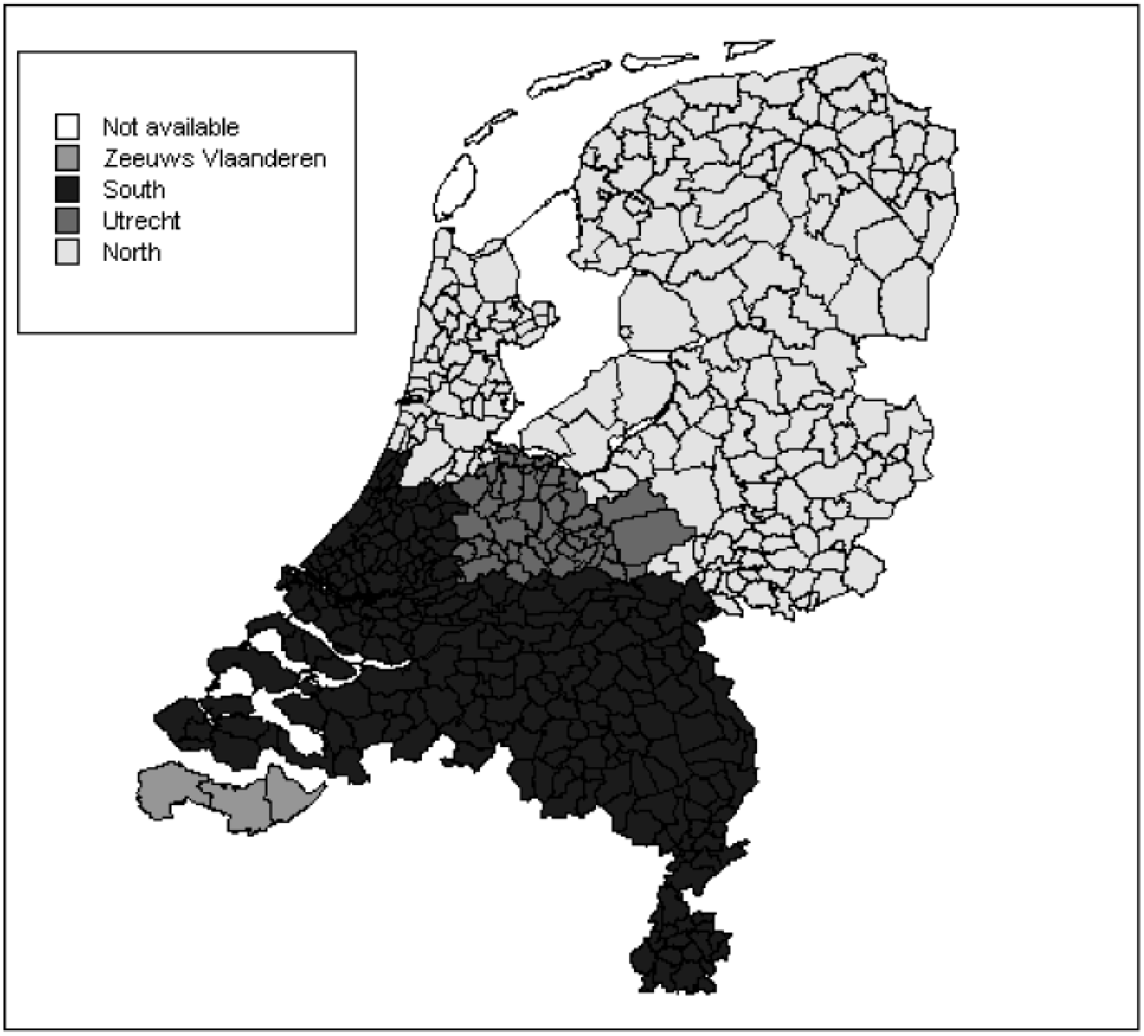
Figure 1
The 4 administrative regions (NUTS1) of the Netherlands



Other administrative delimitations that are compared to the functional delimitations include the RBA division of 28 regions and the COROP division of 40 regions (NUTS3). The RBA division refers to a delimitation of labour market areas, formerly used by the national employment agency. The COROP regions were delimited according to the nodal division principle, which means that every region contains a central municipality. Although the COROP regions can be considered more or less as functional regions, an additional requirement for this delimitation was that the COROP regions were situated within the boundaries of the provinces. Both the RBA and the COROP divisions have been widely used in structural analyses of labour markets, for analyzing territorial disparities, but also by specific administrative bodies to plan their policies. In this section we only show the 4-region administrative and functional divisions of 2001.[7]

7. See for the figures of the 12-, 24- and 40-divisions Bongaerts, Cörvers and Hensen (2004).

Figure 2
The 4 functional regions of the Netherlands, 2001



In general, the delimitations using 1991 commuting flows lead to more small regions than in 2001. An explanation for this could be the shorter travel distances. In 1991 workers may have been less able – due to a less favourable infrastructure or their not having a car – or less willing to commute to reach their work location than in 2001. If commuting distances further increase during the next decade in the Netherlands, then there will hardly be left small regions that represent more or less closed labour markets.

4 Testing for the coherence of regions

4.1 Descriptive statistics of economic indicators

The analyses performed in the previous section have led to a delimitation of the Netherlands into functional regional labour markets. The question remains to what extent municipalities reveal more coherence within functional regions than within administrative regions. In order to test the coherence of functional relative to administrative regions, four economic indicators will be used. The larger the coherence of municipalities within the same (functional or administrative) regions, the larger the differences in economic and labour market performance between municipalities of different regions. From Section 2 it follows that the larger the commuting flows between municipalities, the smaller the wage differences between these municipalities.

In Section 3 the municipalities with the largest direct and indirect commuting flows to each other have been clustered into functional regions. It can be expected that municipalities that have large commuting flows between each other reveal a low wage (or income) variance, and consequently, as has been argued in Section 2, also reveal a low variance in housing prices, labour participation and unemployment. Below we will first describe the mean and the standard deviation of these indicators for the 4 administrative and functional regions distinguished in the previous section.

Table 2

Overview of statistics of the 4 administrative regions of the Netherlands, 2001

Administrative regions (number of municipalities)	Income level €	Std. Dev.	Housing prices €	Std. Dev.
North (65)	17,688	3,154	47,774	9,992
East (102)	17,362	4,170	84,961	22,598
West (207)	18,247	4,170	84,961	22,598
South (115)	17,747	4,170	84,961	22,598
Total (489)	17,874	4,170	84,961	22,598

Administrative regions (number of municipalities)	Employment rate %	Std. Dev.	Unemployment rate %	Std. Dev.
North (65)	61.20	3.54	5.01	2.16
East (102)	64.49	3.87	3.31	1.33
West (207)	66.19	4.29	3.07	1.37
South (115)	64.16	3.52	3.17	1.44
Total (489)	64.67	4.22	3.39	1.62

Source: OVG (2001) and Statline (Statistics Netherlands)

Table 2 gives an overview of the average values and standard deviations with regard to the four economic indicators for the municipalities in the 4 administrative regions of the Netherlands. The North region traditionally has the lowest labour participation, as is indicated by the employment and unemployment rate. The North region also has relatively much open space, which explains the low housing prices. In the West region the income level, the housing prices and the labour participation rates are the highest. However, the standard deviations are also relatively high (except for the unemployment rate) in the West region.

The same overview is presented in Table 3 for the 4 functionally delimited regions. In the functional division the regions of Utrecht and Zeeuws Vlaanderen have the most extreme values for the four indicators. In Utrecht the income level, housing prices and labour participation are the highest, in Zeeuws Vlaanderen the lowest (except for the income level).

Table 3

Overview of statistics of the functional delimitation into 4 regions of the Netherlands, 2001[8]

Functional regions (number of municipalities)	Income level €	Std. Dev.	Housing prices €	Std. Dev.
Zeeuws Vlaanderen (3)	17,893	2,909	38,721	3,780
South (258)	17,911	4,166	69,043	13,645
Utrecht (47)	17,974	4,170	84,961	22,598
North (176)	17,775	4,107	60,984	16,899
Total (484)	17,861	4,116	66,896	17,620

Functional regions (number of municipalities)	Employment rate (%)	Std. Dev.	Unemployment rate (%)	Std. Dev.
Zeeuws Vlaanderen (3)	62.53	4.20	4.90	2.00
South (258)	65.07	4.13	3.20	1.41
Utrecht (47)	66.77	4.17	2.74	1.58
North (176)	63.74	4.18	3.76	1.75
Total (484)	64.66	4.24	3.40	1.61

Source: OVG (2001) and Statline (Statistics Netherlands)

4.2 Specification of the test

The differences in income levels, housing prices, employment and unemployment rates between municipalities of different regions are used as indicators for the coherence of both functional and administrative regions. According to the theoretical model of Section 2 and due to the clustering procedure of Section 3, the functional regions should reveal more significant differences with regard to these economic indicators than the administrative regions. If the regions in the functional delimitation show more coherence than in the administrative delimitation, regional labour market policies can be directed more successfully to municipalities of the same functional regions than to municipalities of the same administrative regions.

To test for the coherence of the regions, we regress each of the economic indicators of the municipalities in the functional or administrative clusters of municipalities. Dummy

8. As a result of missing data, the 5 islands in the north ('Waddeneilanden') have not been included. This means that we have 484 municipalities instead of 489. Moreover, data for the economic indicators of employment and unemployment rates were only available for municipalities with more than 10,000 inhabitants. For these indicators, data for 300 municipalities were used in the analysis.

variables were included as explanatory variables to account for the differences in average income levels, housing prices, employment and unemployment rates between regions. The question in this context is whether there are any significant differences between the different clusters of municipalities with regard to the four economic indicators mentioned before. From the analyses in Sections 2 and 3, we expected the differences between the administrative regions to be less significant than the functional regions. The following equation has been estimated to reveal the average income differences between the functional and administrative regions:

$$Income_m = \beta_0 + \beta_{1,..k} * delimitation \qquad (15)$$

where m stands for the municipalities, k is the number of regions minus 1, β_0 represents the average income level of the reference region, and $\beta_{1,..k}$ represent the differences between the average income level of the other regions and the reference region. Similar regression equations are estimated for housing prices, employment and unemployment rates. To estimate the average income levels of the four administrative regions of the Netherlands (north, south, east and west) the equation is specified as:

$$Income_m = \beta_0 + \beta_1 * South + \beta_2 * East + \beta_3 * West \qquad (16)$$

The regressions are repeated for all combinations of regions (i.e. taking different reference regions). Comparing the significance of the coefficients of the regression equation above with the significance of the coefficients of the four functional regions presented in equation (19) indicates the extent to which the two delimitations can explain differences in income level between municipalities.

$$Income_m = \beta_0 + \beta_1 * Region\ 1 + \beta_2 * Region\ 2 + \beta_3 * Region\ 3 \qquad (17)$$

Significant differences in economic indicators between regions indicate that the delimitation is based on coherent regions. Therefore we have counted the number of significant differences between the average levels of the economic indicators of the municipalities in the functional and the administrative regions.

4.3 Results

The higher the number of significant differences between the regions in the estimated equations, the lower the interaction of workers between these regions, and the higher the coherence of the municipalities within the regions. Tables 4 and 5 show the mean differences of the four economic indicators of all possible combinations of the four administrative and functional regions, respectively. Remarkably, the differences in income levels are not significant between the administrative and functional regions. On the other hand, housing prices are significantly different for almost all of the 6 pairs of regions. The employment and unemployment rates are significantly different for about half of the 6 pairs.

Table 4
The mean differences of the four economic indicators for the four administrative regions (NUTS1) of the Netherlands, 2001

(I) Region	(J) Region	Income level Mean dif. (I-J)	Income level Sig. p-value	Housing price Mean dif. (I-J)	Housing price Sig. p-value
1) North	2) East	325	0.622	-18,625*	0.000
	3) West	-560	0.343	-23,535*	0.000
	4) South	-60	0.926	-23,011*	0.000
2) East	3) West	-884	0.079	-4,909*	0.011
	4) South	-385	0.497	-4,385*	0.044
3) West	4) South	499	0.303	524	0.778
Number of sig. dif.	Out of 6	0		5	

* = Significantly different at the 5%-level

(I) Region	(J) Region	Employment rate Mean dif. (I-J)	Employment rate Sig. p-value	Unemployment rate Mean dif. (I-J)	Unemployment rate Sig. p-value
1) North	2) East	-3.292*	0.000	1.698*	0.000
	3) West	-4.996*	0.000	1.939*	0.000
	4) South	-2.967*	0.000	1.835*	0.000
2) East	3) West	-1.704*	0.003	0.241	0.270
	4) South	0.325	0.613	0.136	0.578
3) West	4) South	2.029*	0.000	-0.105	0.218
Number of sig. dif.	Out of 6	5		3	

* = Significantly different at the 5%-level

The positive and negative signs of the differences across the four economic indicators are generally in accordance with the predictions made at the end of Section 2. In most cases a region with a lower average income level than another region, also has a lower average housing price, a lower average employment rate and a higher average unemployment rate relative to the other region. The North region, for example, has a lower income level than the West region - although not significantly so -, a significantly lower housing price and employment rate, and a significantly higher unemployment rate than the West region.

Table 5
The mean differences of the four economic indicators for the four functional regions of the Netherlands, 2001

(I) Region	(J) Region	Income level Mean dif. (I-J)	Income level Sig. p-value	Housing price Mean dif. (I-J)	Housing price Sig. p-value
1) Zeeuws Vlaanderen	2) South	-17	0.992	-30,319*	0.000
	3) Utrecht	-82	0.964	-46,237*	0.000
	4) North	117	0.946	-22,262*	0.001
2) South	3) Utrecht	-65	0.924	-15,918*	0.000
	4) North	134	0.741	8,058*	0.000
3) Utrecht	4) North	199	0.773	23,976*	0.000
Number of sig.dif.	Out of 6	0		6	

* = Significantly different at the 5%-level

(I) Region	(J) Region	Employment rate Mean dif. (I-J)	Employment rate Sig. p-value	Unemployment rate Mean dif. (I-J)	Unemployment rate Sig. p-value
1) Zeeuws Vlaanderen	2) South	-2.532	0.297	1.697	0.066
	3) Utrecht	-4.235	0.094	2.164*	0.025
	4) North	-1.203	0.620	1.143	0.216
2) South	3) Utrecht	-1.702*	0.047	0.467	0.151
	4) North	1.329*	0.008	-0.555*	0.004
3) Utrecht	4) North	3.031*	0.001	-1.021*	0.002
Number of sig. dif.	Out of 6	3		3	

* = Significantly different at the 5%-level

From the theoretical model of Section 2 it follows that the differences in the income level between municipalities of different regions are larger if the commuting flows between these municipalities are smaller. In other words, the differences in income level should be larger between the functional regions than between the administrative regions. The same reasoning holds for the other three indicators related to the income level and the commuting flows.

Tables 6 and 7 below give a complete overview of the percentages of significant differences for the four economic indicators with respect to the 4, 12, 28 and 40 functional and administrative delimitations in 2001 and 1991, respectively. For the income level in the 12, 28 and 40 division of regions, the functional delimitation performs slightly better than the administrative delimitation. It appears that, in terms of average income level, the functionally defined regions have slightly more coherence than the administrative regions of the Netherlands. The functional division of 12 regions has the best score, since 27% out of the 66 pairs of regions have a significantly different income level. For the other three economic indicators, the performance is generally much higher for both the administrative and the functional delimitation. However, for these economic indicators the functional delimitation is not better than the administrative delimitation.

With respect to the four economic indicators it follows that the number of regions that are significantly different from each other is more or less equal for the administrative and the functional 2001 delimitations. Only in the case of the 28 division of regions the performance is slightly better for the functional delimitation. The differences in performance between the administrative and functional divisions of 4, 12, 28 and 40 regions in 2001 are, however, small.

Table 6

Percentages of significant differences (at the 5%-level) between the means of the economic indicators, delimitations of 2001

Indicator	4 region division		12 region division	
	Administrative %	Functional %	Administrative %	Functional %
Income level (2001)	0	0	0	27
Housing price (2001)	83	100	74	67
Employment rate (2001)	83	50	56	33
Unemployment rate (2001)	50	50	50	36
Total	54	50	45	41

Indicator	28 region division		40 region division	
	Administrative %	Functional %	Administrative %	Functional %
Income level (2001)	3	5	2	8
Housing price (2001)	58	62	55	49
Employment rate (2001)	38	36	24	20
Unemployment rate (2001)	28	40	29	24
Total	32	36	28	25

Table 7 shows the percentages of significant differences based on the 1991 delimitation of the Netherlands. For the interregional income differences in both 1991 and 2001, the functional 1991 delimitation performs better than the administrative delimitation (except for the 28 division with the 1991 average income level). As in 2001, the performance of the income level as an economic indicator of interregional differences is low. For the three other economic indicators we again find relatively large percentages of significant differences between administrative and functional regions. The functional division performs slightly worse for these indicators with respect to the 12- and 24-division.

A possible reason for the low percentage of significant differences in the income levels between regions, is that income is in fact an approximation for the wages of individual workers. The average wage levels of individual workers per municipality, measured by for example gross earnings paid by the employer, are not available for the Netherlands. Therefore we have used the net personal income of workers. This definition of income incorporates not only the gross wages earned, but also income taxes, tax allowances and fiscal deductions.

It is generally true that the more differentiated the delimitation is, the worse the relative performance. This holds for both the administrative and the functional delimitations. Although the absolute number of significant differences usually increases when the delimitation is more differentiated, we conclude that there is no overwhelming evidence to differentiate between relatively small regions. One explanation for this may be that the Netherlands is already a small country, and could be regarded as a relatively homogeneous region in itself.

Table 7

Percentages of significant differences (at the 5%-level) between the means of the economic indicators[9], delimitations of 1991

Indicator	4 region division		12 region division	
	Administrative %	Functional %	Administrative %	Functional %
Income level (1991)	0	0	0	8
Income level (2001)	0	17	0	18
Housing price (2001)	83	83	74	51
Employment rate (2001)	83	67	56	27
Unemployment rate (2001)	50	50	50	36
Total	43	54	36	28

Indicator	28 region division		40 region division	
	Administrative %	Functional %	Administrative %	Functional %
Income level (1991)	6	3	2	10
Income level (2001)	3	8	2	6
Housing price (2001)	58	50	55	56
Employment rate (2001)	38	28	24	29
Unemployment rate (2001)	28	21	29	29
Total	27	22	22	26

5 Conclusion

Functional regions can be characterized as regions that have a relatively closed labour and housing market. A closed market in this context refers to a market where the majority of interactions occurs within the boundaries of the functionally defined regions. Commuting flows provide an indication of the extent to which interaction takes place within the boundaries of regions. In this paper, we have provided a delimitation procedure for defining functional labour markets on the basis of the commuting flows between 489 Dutch municipalities for both 1991 and 2001. To delimit regions, a commuting matrix is used of incoming

9. For housing prices, employment rate and unemployment rate, no data was available for 1991 or 1992.

and outgoing commuting flows between all the municipalities. The method applied transforms the standard interaction matrix of commuting flows between municipalities into a mean first passage time (MFPT) matrix.

Furthermore, to cluster municipalities that have more interaction with each other than with municipalities outside the cluster, an aggregation procedure was applied. This aggregation method maximizes within-region commuting flows by merging the two adjacent municipalities (or clusters) with the greatest mutual interaction. The delimitation procedure used in this paper generates those functional regions between which commuting flows are minimized. To compare the functionally defined regions with the administrative ones, the number of functional regions was tuned to the number of regions of the administrative delineation.

To determine the coherence between the delimited regions, a measure of economic relatedness was used. Four economic indicators were used to determine coherence in 1991 and 2001: (i) income level; (ii) housing prices; (iii) employment rate, and (iv) unemployment rate. For these indicators, we tested whether the municipalities within the functionally defined regions show more coherence than the municipalities within the administratively defined regions. For both 1991and 2001 it appears that, in terms of income level, the functional regions have slightly more coherence than the administrative regions. The performance of income level as an economic indicator of differences between regions, however, was much worse than for the other economic indicators. For the other three economic indicators, the functional and the administrative regions showed, on average, the same coherence for both 1991 and 2001.

It can be concluded that the administrative delimitation of the Netherlands performs, on average, equally well as the functional delimitation. The hypothesis that the municipalities within the administratively defined regions show less coherence than the municipalities within the functionally delimited regions, cannot be rejected. There is some minor evidence that the coherence is greater for the average income level of municipalities within functional regions than within administrative regions. The performance of housing prices, employment rate and unemployment rate, is not very different for the functional delimitation on the one hand and the administrative on the other.

Our results imply that there is not much to be gained in labour market policies by using functional instead of administrative divisions of regional labour markets. Other reasons for particular delimitations of regions, such as the existence of regional administrative and governmental bodies and the managerial control over regions, may be more important. Therefore we doubt the relevance of the functional delimitations by Andersen (2002), Ball (1980), Casado-Diaz (2000) and Van der Laan and Schalke (2001) for Denmark, Great Britain, Spain and The Netherlands, respectively. Moreover, it may be better for regional labour market policies not to use a highly differentiated division of regions for small countries like the Netherlands. In general, the regionalization of the Netherlands into four regions seems to be sufficient.

If data is available it is useful to test for coherence by using other indicators, such as real wage instead of personal income, or to consider another measure of interaction between municipalities. For example, migration flows instead of commuting flows could be used in the aggregation procedure. Finally, the coherence within functional regions may be different for particular labour market segments: for low skilled workers and workers in unskilled occupations the functional regions may be smaller than for highly-skilled workers.

References

Amedo, D. (1968), An Optimization Approach to the Identification of a System of Planning Regions, *Regional Science Association*, vol. 23, pp. 25-44.

Andersen, A.K. (2002), Are Commuting Areas Relevant for the Delimitation of Administrative Regions in Denmark?, *Regional Studies*, vol. 36, pp. 833-44.

Ball, R.M. (1980), The Use and Definition of Travel-To-Work Areas in Great Britain: Some Problems, *Regional Studies*, vol. 14, pp. 125-39.

Baumann, J.H., M.M. Fischer and U. Schubert (1996), A Multiregional Labour Supply Model for Austria: The Effects of Different Regionalisations in Multiregional Labour Market Modelling; in M.M. Fischer, K.J. Button and P. Nijkamp (eds.), *Regional Housing and Labour Markets*, Elgar Reference Collection, vol. 5, Cheltenham, U.K. and Lyme, N.H., pp. 366-96.

Bongaerts, D., Cörvers, F. and M. Hensen (2004), *The Delimitation and Coherence of Functional and Administrative Regions*, Research Series 04O19, Ministry of Economic Affairs, The Netherlands, The Hague.

Brown, L.A. and J. Holmes (1971), The Delimitation of Functional regions, Nodal Regions, and Hierarchies by Functional Distance Approaches, *Journal of Regional Science*, vol. 11, pp. 57-72.

Casado-Diaz, J.M. (2000), Local Labour Market Areas in Spain: A Case Study, *Regional Studies*, vol. 34, pp. 843-56.

Coombes, M.G., A.E. Green and S. Openshaw (1986), An Efficient Algorithm to Generate Official Statistical Reporting Areas: The Case of the 1984 Travel-To-Work-Areas Revision in Britain, *Journal of Operational Research Society*, vol. 37, pp. 943-53.

Corpelijn, A.W.F. and M.J. Heerschop (2002), Inkomende en Uitgaande Pendel, 1998, *Sociaal-Economische Maandstatistiek*, no. 2002/12, Statistics Netherlands, Voorburg/ Heerlen, pp. 53-57.

Fox, K.A. and T.K. Kumar (1965), The Functional Economic Area: Delineation and Implications for Economic Analysis and Policy, *Regional Science Association Papers*, vol. 15, pp. 57-85.

Giulliano, G. and K.A. Small (1991), Subcenters in the Los Angeles Region, *Regional Science and Urban Economics*, vol. 21, pp. 162-82.

Killian, M.S. and C.M. Tolbert (1993), Mapping Social and Economic Space: The Delineation of Local Labor Markets in the United States; in J. Singelmann and F.A. Deseran (eds.), *Inequalities in Labor Market Areas*, Boulder and Oxford Westview Press, pp. 69-79.

Laan, L. van der and R. Schalke (2001), Reality versus Policy: The Delineation and Testing of Local Labour Market and Spatial Policy Areas, *European Planning Studies*, vol. 9, pp. 201-21.

Masser, I. and J. Scheurwater (1980), Functional Regionalisation of Spatial Interaction Data: An Evaluation of Some Suggested Strategies, *Environment and Planning A*, vol. 12, pp. 1357-82.

OECD (2002), *Redefining Territories, The Functional Regions*, Paris.

Openshaw, S. (1977), Optimal Zoning Systems for Spatial Interaction Models, *Environment and Planning A*, vol. 9, pp. 169-84.

Ward, J.H. (1963), Hierarchical Grouping to Optimise an Objective Function, *Journal of the American Statistical Association*, vol. 58, pp. 236-44.

Web-pages:
http://tecfa.unige.ch/~lemay/thesis/THX-Doctorat
www.quantlet.com/mdstat
www.clustan.com/general_distances.html

Appendix A: MFPT method and clustering procedure

From travel flows to a Markov Chain

In the analysis presented here, an MFPT matrix has been computed. In order to understand the concept of an MFPT matrix, we first have to regard our daily travel-to-work commuting flows of workers as a Markov chain. A Markov chain is a stochastic process, which describes the transition from one state to another over time using probabilities. It has a number of states and a discrete time line. At each point in time t, we have a square transition matrix P, which represents probabilities to go from state i at time t to state j at time $t + 1$. One can multiply matrix P with itself in order to obtain transition probabilities over more time periods. For example, P^3 gives the transition probabilities from t to $t + 3$.

In our case, we have 492 Dutch municipalities as states (origins and destinations of travel flows) and the (weighted) number of travel-to-work movements from i to j, from each $\{i,j\}$ made during a whole year. To transform these numbers to probabilities, we have to divide these numbers by the column or row totals, which correspond with the total number of people heading for municipality j and the number of people originating from municipality i, respectively. In turn, we can derive a destination-based or an origin-based probability matrix. Masser and Scheurwater (1980) argued in favour of using the destination based probability matrix when handling travel-to-work flows. Therefore, the destination-based approach will be used throughout this paper.

From Markov Chain to MFPT

As stated above, we computed an MFPT matrix for the current analysis. The idea behind this concept is that if you have a Markov chain in which it is possible to re-enter each state at any point in time (i.e. $\sum_j P_{i,j} = 0$ for some i where $p_{i,j}$ is element (i, j) of matrix $lim_{t\to\infty}(P^t)$), one can compute the average number of transitions needed to arrive from origin i in destination j for the first time. Note that because the probabilities to go from i to j are in general not equal to the probabilities to go from j to i by construction, our MFPT matrix is typically asymmetric.

Computation of the MFPT

Suppose we have a Markov chain with single-period transition matrix P. As mentioned above, the j-period transition matrix is defined by P^t. If we let this process run for an indefinite time span, we end up in an equilibrium state. The proportions of time spent in each state are then $lim_{t\to\infty}P^t = A$. A is called the limit matrix. Having these two matrices, we can compute the so-called fundamental matrix Z of the process. Z can be computed by $Z = (I - (P - A))^{-1}$, where I is the identity matrix. From the limit matrix, matrix D is defined by $1/a_i$ on its diagonal and zeros for all other elements. The MFPT matrix can then be computed by $M = (I - Z + EZ_{diag})D$ where E is a matrix containing ones everywhere and Z_{diag} is the matrix containing the diagonal elements of Z and zeros for all other elements.

Properties and interpretation of the MFPT matrix

One item that immediately draws attention is the structure of the MFPT matrix. First of all, the diagonal elements are very small; this indicates that there are many travel-to-work flows within a region, something quite intuitive. Furthermore, all other values in the columns are relatively close to the column average, that is to say, they are of the same order. These column averages are indicators for how much attraction a region has to work in. The lower the column average, the more attractive the region is.

From MFPT to distances

The asymmetry that we observed for the MFPT matrix is particularly inconvenient if we want to cluster regions, as clustering procedures often implicitly assume symmetric distances. Another problem arising from the MFPT is that the order of the column averages differs considerably among columns, which may result in the clustering of all larger regions together, even though the distance in kilometres between these regions is very large. In fact, we want to cluster the regions in such a way that the variation within clusters is minimal. Therefore we need appropriate measures for variation. Let us first solve the problem of differences in the order of column averages. This can be done by taking the z-values, which are defined by $z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$. Note that although the diagonal values of this z-matrix can be computed, they make no sense and should be equal to zero or even nonexistent. From these z-values we want to obtain a measure for how close regions i and j are to each other. This is done with a so-called squared distance matrix. For each column k, we will compute the difference between z_{ik} and z_{jk} (where $i,j \neq k$) and square it. This is the marginal contribution from k to the squared distance. In formula this can be written as:

$$d_{i,j} = \sqrt{\left(\sum_{k:i,j \neq k} (z_{ik} - z_{jk})^2 \right)}$$

By construction, d_{ij} is equal to d_{ji}, so this transformation also handles all other problems. Having established this as a starting distance matrix, we can look at the clustering procedure.

The clustering procedure: Ward's Method

Now that we have an initial distance matrix, we can start the clustering procedure. Two municipalities/clusters can only be clustered together if they are adjacent. In general, the clustering procedure works as follows (in standard pseudo code):

- Start
- While (*NrOfClusters* > *TargetNrOfClusters*)
  - Search for i,j such that $d_{ij} = \min_{i,j} d_{ij}$
  - Cluster i and j in a new cluster a
  - Compute new distances from a to all the other regions/clusters
- Return clusters

• Stop

The only part that is lacking now, is the way to compute the new distances from a to all other clusters. Ward (1963) proposed a method that minimizes variance within clusters for any type of clustering. This is the method we have used as well. It computes the new distance $d(a,r)$ between a new cluster a consisting of p and q to another region/cluster r according to the formula:

$$d(r,a) = d(r, p \cup q) = \frac{(n_p + n_r) \cdot d(p,r) + (n_q + n_r) \cdot d(q,r) - n_r \cdot d(p,q)}{n_p + n_q + n_r}$$

where n_x is the number of elements in cluster x.