Dynamic Information Externalities and Employment Growth in the Province of South-Holland

Paper prepared for the 40th ERSA-congress,
Barcelona august 29-september 1, 2000

Frank G. van Oort[1], Daan P. van Soest[2] and Shelby D. Gerking[3]
[1]National Spatial Planning Agency, Directorate of Spatial Research and Planning, The Hague, and Tinbergen Institute, Erasmus University Rotterdam, The Netherlands. E-mail: Frank.vanOort@rop.rpd.minvrom.nl.
[2]Department of Economics and CentER, Tilburg University, Tilburg, The Netherlands.
[3]Department of Economics and Finance, University of Wyoming, Laramie 82071,

Abstract

This paper brings together two important strands of literature on the relationship between knowledge spillovers and employment growth. The first strand tests for evidence of enogenous growth linked to knowledge and knowledge spillovers between economic agents within cities and the second tests whether knowledge spills over between economic agents in different locations. The link between these two topics is made by extending the work of Glaeser, Kallal, Schienkman, and Schliefer (1992) to develop a spatial lag model that allows employment growth in one location to affect growth in other locations. The empirical work presented focuses on the province of South-Holland, the Netherlands. A key finding reported here is that local industrial diversity and increased local competition tend to promote growth. Additionally results in this study suggest that knowledge spillovers in one location can lead to growth in other locations although the magnitude of this effect appears to be small.

1. Introduction

Beginning with Romer (1986) and Lucas (1988), the theory of endogenous economic growth has emphasised the role in the growth process of both the stock of knowledge and knowledge spillovers between economic agents. Competing theories have been advanced regarding how knowledge spillovers affect growth and subsequent tests have been sharply divided on the relative roles played by, among others, historical industrial concentration and historical industrial diversity. In their study of U.S. cities, for example, Glaeser, Kallal, Scheinkman, and Schliefer (GKSS) (1992) find that spillovers associated with industrial diversity are more important for growth than externalities that arise from information accumulated within a localized industry. Audretsch and Feldman (1996) and Feldman and Audretsch (1999) find similar results in looking at new product introductions and innovations. Other recent papers, however, have reached different conclusions regarding the relative importance of different types of knowledge spillovers (Henderson, Kuncoro, and Turner 1995, Bostic, Gans and Stern 1997, Henderson 1997, Black and Henderson 1999, Beardsell and Henderson 1999). A recurring theme here is that more traditional factors such as cost, demand, and resource endowment differences between locations are also a critical part of the growth process (see also Kim 1999). Ellison and Glaeser (1999) do not dispute this point, but instead present empirical evidence suggesting that interstate differences in resource endowments alone probably do not fully account for the observed geographic pattern of U.S. industrial concentration. Yet, in any case, no clear picture has emerged from this literature as to the relative importance of alternative explanations of urban growth.

Additionally, recent studies of urban growth have focused on spillovers between industries in a city and have not explicitly or systematically examined effects of possible knowledge spillovers between cities. The main reason for this omission is that studies of patents and innovations generally show that knowledge spillovers tend to be geographically bounded within the region where the knowledge was generated (Jaffe 1989, Jaffe, Trajtenberg, and Henderson 1993, Anselin, Varga, and Acs 1997). These studies, however, focus on the geographic limits of knowledge transmission and do not consider that effects of knowledge creation might indirectly spill between locations via the growth process. Also, empirical urban growth studies implicitly treat possible linkages between cities quite unevenly. Henderson, Kuncoro, and Turner (1995) specify their regressions so as to rule out links between cities. In contrast, a possible interpretation of the GKSS analysis is that a particular growth rate in an industry in a city has virtually the same effect on growth in that industry in all other cities, no matter how much distance lies between them. Thus, further attention to knowledge spillovers between locations would be desirable, particularly if it can also help resolve some of the other issues prevailing in the literature.

Determining the relative importance of the various types of spillovers in the growth process has broad implications ranging from the formulation and interpretation of endogenous growth models to practical conclusions that might be drawn by urban specialists regarding which types of cities should be expected to grow fastest. This paper analyses dynamic knowledge spillovers within and between industries as well as within and between locations using a spatial econometric model that arises as a natural extension of the work by GKSS. The analysis focuses on economic growth in the Dutch province of South-Holland, the core economic region of the Netherlands. This province is heavily urbanised with a population density of over 1000 inhabitants per km^2 and is of interest because it is small enough to offer a natural control for location-specific attributes.[1] Within this province, cultural and economic differences between locations are simply less important and more easily controlled than they would be between the major U.S. cities considered in previous studies. Also, detailed longitudinal employment data are available in each year from 1988-97 for virtually *all* establishments present in all economic sectors in each of the 416 4-digit zip (postal) code areas that make up 69 municipalities. With such dense economic activity and establishments located on a fine spatial grid, South-Holland is an ideal area in which to test for the extent of knowledge spillovers between locations. The data also show employment changes in each sector in each year attributable to births of new establishments, in- and out-migrating establishments, and existing establishments that did not move (including establishment deaths). This level of detail permits sources of growth analyses to be performed for the first time that show how knowledge spillovers affect "demographic" components of urban growth.

The remainder of the paper is divided into four sections. Section 2 discusses key aspects of previous studies and presents the conceptual framework for the empirical study. Section 3 describes the South-Holland data. Section 4 presents empirical results that strongly support those of GKSS as well as the notion that knowledge, once generated, does not travel far. Implications and conclusions are drawn out in Section 5.

2. Conceptual Framework

Cities are fertile grounds for testing knowledge-based theories of endogenous growth because economic activity is dense, thus facilitating face-to-face contact as well as other forms of communication (Lucas 1993). Prior studies have tested three (in some respects

[1] The size of South-Holland is about 2350 km^2, and hence it is smaller than, for example, Rhode Island.

competing) hypotheses concerning the conditions under which knowledge spillovers affect growth. The first hypothesis, originally developed by Marshall (1890) and later formalized by Arrow (1962) and Romer (1986) (MAR), emphasises spillovers between firms in a city-industry and contends that these spillovers are most important when there is little prevailing local competition. The second (see Porter 1990) agrees that knowledge spillovers within a localised industry are most important, but argues that their effects on growth are enhanced by local competition. The third (see Jacobs 1969) emphasises spillovers between industries and contends that they promote growth most effectively in a competitive environment.

As discussed in the introduction, prior empirical tests of these hypotheses reach different conclusions, particularly regarding effects of local industrial concentration versus local industrial diversity. One key difference between these studies rests on whether data from all cities in a given industry are analysed (Henderson, Kuncoro, and Turner 1995) or whether only the largest city-industries are included in the sample (GKSS). Consequently, Glaeser (1998, p.148) suggests that "[a] possible reconciliation of results [on this point] is that scale and concentration may have value for smaller firms; however, diversity has more value for long term growth." Beardsell and Henderson (1999) argue that another important difference lies in the treatment of time invariant location attributes. In particular, they state (p.449) that "…rather than the link between the present and the past representing mostly dynamic externalities, an alternative explanation is that there is a location fixed/random effect in estimation that gives rise to the role of history."

This paper extends the methods used by GKSS in that it focuses on growth rates of the largest city (actually zip code)-industries. In principle, a study of growth of individual industries across all zip codes could be carried out, but in practice it would be difficult because most industries are represented in only a relatively small number of zip codes. Consequently, beginning-of-period and end-of-period employment would be zero in a large number of cases. This aspect would not be a problem if the aim of the study was to ask why particular industries chose to locate in particular zip codes. However, the primary focus for this study is on the closely related issue of mechanisms thought to be important to the growth process. This emphasis motivates the decision to look at the largest industries that were present in the zip code area at the beginning of the sample period.[2]

[2] In their study of U.S. cities, Henderson, Kuncoro, and Turner (1995) tried to incorporate all cities. They faced the additional problem that because of federal disclosure rules, industry employment values were censored in as many as 30% of the cities studied; censoring is not a problem in the South-Holland data.

GKSS specified employment growth in a city-industry as a function of local specialisation in that industry, extent of local competition, local industrial diversity, and control variables including the national employment growth rate of the industry outside the city. The national employment growth variable was included to account for national demand shifts (see Blanchard and Katz 1992) and was measured as the log of the ratio of end-of-period employment to beginning-of-period employment. Data were drawn from the six largest industries in U.S. cities for the period 1956-87. In addition to supporting the importance of Jacobs-type externalities as a determinant of urban growth, estimates of the coefficient of the national industry growth rate were significantly larger than unity.[3] This result suggests that the industries studied grew faster in cities than in rural areas.

The outcome for the national industry growth-rate variable, however, has an alternative interpretation involving the way that knowledge is transmitted between cities. This variable weights industry growth rates according to shares of national employment. Thus, had the own city not been excluded from this calculation, the weight assigned to the growth rate of an industry in one city would be the same when estimating the growth rate of that industry in any other city in the data set. The fact that the own city growth rate was excluded probably makes little difference in this regard except in cases where its share of beginning-of-period employment is large. In any case, the implication here is that growth is transmitted from one city-industry to the same industry in another city. Thus, knowledge spills over between cities via this "within-industry" mechanism, which functions independently of distance between cities. Additionally, the national industry growth rate variable does not provide for growth effects to be directly transmitted between different industries in different cities. These ideas are formalised in equation (1),

$$y=\rho Wy^{*}+X\beta+\varepsilon \tag{1}$$

in which y denotes an Nx1 vector of city-industry growth rates (of those sectors belonging to the largest six in each city), W denotes an $N\text{x}N^{*}$ ($N\leq N^{*}$) weighting matrix, y^{*} denotes an N^{*}x1 vector of industry growth rates at *all* locations in the country (which exceeds the number of cities in the sample), X denotes a matrix of observations on a set of city-industry explanatory variables, ε denotes a vector of disturbances, and $[\rho\ \beta]^{T}$ is the coefficient vector where ρ is a scalar. Each row of W would be constructed so as to create an explanatory variable that measures the national growth-rate of the industry outside the city in question.

[3] It may interesting to note that had Glaeser, Kalla, Scheinkman, and Schliefer (1992) constrained coefficients of the national industry growth rate to unity, resulting estimates would be interpreted as an explanation of the differential shift term in shift-share analysis (see Dunn 1960 and Perloff, Dunn, Lampard and Muth 1960). The differential shift term measures the extent to which an industry in a region (or city) grows faster or slower than it does on average in a broader geographic area. In an earlier day, there was considerable debate among

That is, in a given row of *W*, each element either would be a national employment share or zero depending on whether it multiplies an own-industry value in y^*. Of course, equation (1) also can be broadly interpreted (recall that W is not square) in the context of a spatial lag model (Anselin 1988, p.22). In this case, the elements of *W* would reflect distances between locations. Thus, the GKSS framework uses the same information on growth rates of city-industries as does the spatial lag model, but applies a different (non-spatial) weighting scheme. However, the spatial interpretation of equation (1) is appealing in light of interest in the question of how far knowledge travels once it is generated. The patent and innovation literature, previously cited, indicates that spatial spillovers are quite limited, but this proposition has never been tested in a growth context.

In the analysis presented below for South-Holland, N equals N^* (so that *W* is square) and y is identical to y^*. In this case, equation (1) can be rewritten to show how and to what extent knowledge externalities are transmitted between regions.

(2) $y=(I-\rho W)^{-1}X\beta=AX\beta$

In equation (2), the matrix $A=(I-\rho W)^{-1}$ spreads the effect of a change in *X* in one location to employment growth rates in other locations. More specifically, let x_{jk} be value of the k^{th} indicator of knowledge spillovers for the j^{th} city-industry. Then, $\beta_k=\partial y_j/\partial x_{kj}$ denotes the partial effect of a one-unit change in x_{jk} on y_j and the total effect of a one-unit change in x_{jk} on growth rates in all areas would be $\beta_k\Sigma_i a_{ij}=dy_j/dx_{kj}$. Thus, the quantity $A_i=\Sigma_i a_{ij}$, which is simply a column sum of elements in *A*, gives the factor by which the total effect differs from the partial effect. This factor will exceed unity if the column sums of ρW are positive and strictly less than one. In the analysis presented below, there is no mathematical guarantee that this condition on ρW will be met; in practice, however, it is met in all of the equations estimated. Additionally, after removing values pertaining to other industries in the same region, a column sum of *A* can be interpreted as a measure of the extent to which a particular type of knowledge externality spills over from one location to others. Partial and total effects of dynamic knowledge externalities are calculated in a spatial context for South-Holland in Section 4. Data are described in Section 3, which follows.

3. Data

---

regional scientists as to what determines the value of the regional share term (see Houston 1967). The Glaeser, Kallal, Scheinkman, and Schliefer results suggest that Jacobs-type externalities are important in this regard.

Data from the Dutch province of South-Holland are used to estimate the empirical models reported in the next section. South-Holland is located at the south-western edge of the Randstad-region and has a high population density (about 1190 persons/km^2). It includes the second and third largest cities in the Netherlands (Rotterdam and The Hague) as well as numerous medium sized cities such as Leiden, Delft, Schiedam, and Dordrecht. In 1997, the two sectors having the largest number of workers are wholesale and retail trade, which each employ over 100,000 persons. Agriculture employs about 36,000 persons and within manufacturing, the furniture, metal products, machinery, chemicals, and printing and reproduction sectors all employ more than 10,000 persons. Longitudinal data covering the period 1988-97 on employment and related variables for individual establishments are available from the Firm Register South-Holland (see Appendix A for a more detailed description of the data sources).

These data are of interest for several reasons. First, the data are comprehensive in that they include virtually all establishments present in South-Holland in each year of the sample period. The data set contains just over 1 million observations, giving annual information on approximately 100,000 establishments.[4] Establishments are enumerated based on information furnished by the Chamber of Commerce, insurance companies, and industrial sector associations and an annual questionnaire is sent to each. The average annual response rate to the questionnaire is 96%. Second, the data are available at a very fine scale. Questionnaire results identify each firm's 6-digit zipcode (a small area containing about 100 different mailing addresses), and 5-digit activity code. These features are an advantage when testing for spatial knowledge spillovers, but the level of detail may already be too low for meaningful analysis. When the data are organised into a location-by-activity matrix at this level of detail, most of the cells contain no information. Many of the 6-digit zipcode areas have only residences. Consequently, the data were aggregated up to the 4-digit zipcode, 2-digit activity code level. The spatial scale at which the firm dynamics can be studied still is very small, particularly when compared to U.S. counties or cities, which in some cases are defined as two or more contiguous counties. In the entire area of South Holland (2350 km^2), there are 416 4-digit zip code areas, which together make up 69 municipalities. The average size of a 4-digit zip code is about 5.65 km^2, although they tend to be smaller in urban areas where the density of addresses is high and larger in areas that

[4] Each establishment in the area firm is also given a unique identification number, so that it is traceable through time and space. For the analysis of growth pursued here, such level of detail is not needed. Growth is essentially a long-run phenomenon and hence we only look at the total change in employment over the entire 1988-1997 period rather than at annual changes.

have more open space. Also, the 2-digit activity code in the Dutch industrial classification system is roughly comparable to the 2-digit SIC classification used in the U.S.

The small size of South-Holland is used to assist in controlling for location-specific factors that might affect growth. In fact, many of the variables enumerated in prior studies (Henderson, Kuncoro, and Turner 1995, Henderson 1997, Kim 1999, Ellison and Glaeser 1999) as potentially important location-specific factors either are roughly constant between locations in South-Holland, or else can be controlled for to some extent. Cultural differences between locations in South-Holland are negligible. Variations in taxes, environmental amenities (such as climate), and environmental regulations between locations are quite small. Differences in prices of non-land inputs exhibit little variation across the province. Prices charged for energy inputs vary by sector, but within a sector, they are the same throughout the province. Wages also vary by sector, but not much within sectors. The province is small enough that workers can live in one zip code area and commute to work in almost any other (as well as to areas in other provinces) using either public or private transportation modes, and in fact they do. Thus, wage rates within a sector would be uniform and there is no need to control for labour force characteristics such as level of education, percent of workers with particular skills, or percent of workers who are union members. There are, of course, historical factors that have led to the current spatial organisation of economic activity. These factors can be controlled to some extent by including appropriate explanatory variables (see below).

4. Empirical Results

This section, presenting empirical results, is divided into three parts. The first seeks to apply the GKSS analysis to the case of South-Holland to get a grasp of the differences and similarities between the two studies. The second estimates the spatial lag model in equation (2) in order to measure the extent of knowledge spillovers between zip code areas. The third part compares how knowledge externalities affect employment growth in establishments whose location in the zip code area predated the sample period and establishments that were new to the area during the sample period.

a. Preliminaries

This empirical analysis begins by estimating a specification similar to the GKSS model using the South-Holland data. The dependent variable is change in the natural logarithm of sectoral employment in a zip code-industry over the period 1988-1997. Observations on employment in each of the two years are obtained by adding up the number of employees in each 2-digit SIC sector. As indicated previously, many industries are not present in most zip codes, so analysis was initially limited to the six zip code-industries that had the highest employment in 1987. However, even in his case, some of the zip code-industries had very little employment and those with 50 employees or less were excluded from the analysis. This minimum employment cut-off reduced the number of zip-code industries in the data set from 2496 (416x6) to 1797.

Results for two specifications are presented in Table 1 along with means of the explanatory variables (the mean of the dependent variable is –0.264, indicating a net reduction of employment in the sample).[5] The three explanatory variables measuring dynamic knowledge externalities, *CONCENTRATION*, *COMPETITION*, and *SHARE*, are defined in exactly the same way as in GKSS. *CONCENTRATION* is defined as a location quotient showing the percentage of employment accounted for by an industry in a zip code relative to the percentage of employment accounted for by that industry in South-Holland. This variable measures whether an industry is over- or under-represented in a zip code compared with the average representation in the province. *COMPETITION*, measured as firms per worker in a zip code industry divided by firms per worker in that industry in South-Holland, indicates whether firms tend to be larger or smaller in a zip code compared to the province as a whole. *SHARE*, the employment share of the other five largest industries in a zip code, measures the absence of industrial diversity. A positive coefficient of *CONCENTRATION* and a negative coefficient of *COMPETITION* support the MAR hypothesis. A positive coefficient of *CONCENTRATION* and a positive coefficient of *COMPETITION* support the Porter hypothesis. A negative coefficient of *SHARE* and a positive coefficient of *COMPETITION* support the Jacobs hypothesis.

Also, three variables control for industry differences in growth and wages. *GROWTH* is the change in the natural logarithm of employment in an industry in South-Holland outside the zip code area. (Recall that this variable was discussed in connection with equation (1).) *WAGE88* measures the difference in wages between industries at the national level (in the Netherlands) in 1988 and *ΔWAGE* measures the change in the natural logarithm of wages for each industry at the national level over the sample period. No data are available regarding wage payments by establishments, so it is not possible to build up data on wages for each individual zip code industry.

Six variables are used to capture historical factors affecting the spatial organisation of economic activity in South-Holland. *EMPLOYMENT88* measures employment in a zip code-industry in 1988 in number of workers. *ROTTERDAM* measures the distance of a zip code from the seaport at Rotterdam. *SCHIPOL* measures distance of a zip code from the international airport at Schipol and is also indicative of the distance to Amsterdam. *UTRECHT* measures distance from Utrecht, which indicates the proximity to the region's hinterland. *WORKAREA* measures whether a zip code is predominantly a work area rather than a residential area.[6] *SITES* is a dummy variable indicating whether a higher than average number of new industrial sites have been opened in a zip code since the beginning of the sample period. Further details regarding construction of these variables as well as more complete information about data sources are presented in Appendix A.

Variable means reported in Table 1 reflect some differences between the South-Holland data and the data on U.S. cities used by GKSS. For South-Holland, employment exhibited more concentration and less diversity within zip codes than did employment in U.S. cities. The small size of the zip code areas may be partly responsible for this outcome because of the following reason.[7] Given that very often only a few industries are represented in a specific zip code, five of the six largest industries (excluding the own zip code sector's employment) in a zip code will account for a very high percentage of total employment in that zip code. Also, the mean of *COMPETITION* is larger in the South-Holland data, indicating that establishments tend to be smaller than was the case for U.S. cities. *GROWTH* has a lower mean in the South-Holland data than in the U.S. data, however, the length of the sample period used here is correspondingly shorter as well.

Column (1) of Table 1 presents estimates of a specification similar to that used by GKSS, while the column (2) regression that includes more controls for location-specific attributes. Estimation of both equations is by ordinary least squares. The R^2 in the column (1) regression is 0.126, and for column (2) it is 0.143. Thus, the explanatory power of both equations is rather low.[8] Again, the small size of the zip code areas may be partly

---

[5] Therefore, employment in the sample decreased by 4% over the 88-97 period. This is comparable with the 2.4% employment decrease of the largest six zip code sectors (that is, including the sectors with less than 51 employees).

[6] One could also consider to include the distinction between urban and rural areas as an explanatory variable. However, we argue that the main criterion for spillovers is proximity of firms, and hence the zip code's function is more important (although the two variables are undoubtedly correlated).

[7] Notice that mean employment of zip code industries in 1988 was 402, while base year mean employment in the GKSS city-industries was 9,700.

[8] In absolute value, the largest Pearson correlation between any two explanatory variables used in Table 1 is 0.49.

responsible here. Many zip code industries in South-Holland have less than 100 employees, so relatively small absolute employment changes over the sample period can produce relatively large changes in growth rates. Correspondingly, with a comparatively small number of establishments operating in each zip code there is more room for growth rates to be affected by firm-specific factors (such as entrepreneurial effort and luck). Coefficient estimates in Table 1 are similar in many respects to those found by GKSS. Both the column (1) and column (2) regressions show that specialisation in an industry significantly retards growth, while competition and industrial diversity significantly accelerate growth. Thus, the results here support the idea that Jacobs-type externalities foster growth and MAR type externalities tend to slow it down. With respect to control variables, the initial wage level and the rate of wage growth in an industry result in significantly lower rates of growth and initial employment in a zip code industry has no appreciable effect on the rate of growth. Also, the coefficients of *GROWTH* in the two regressions are 0.77 and 0.93, respectively. Both *GROWTH* and the dependent variable are measured in natural logarithms so these coefficients indicate that a 10% increase in industry growth outside the zip code is associated with less than a 10% increase in growth in that industry in the zip code. As indicated previously, GKSS interpret this variable as a measure of demand shifts, but as shown in equation (1), redefining it invokes the spatial lag model. Finally, in column (2), the coefficients of *WORKAREA*, *SITES*, and *ROTTERDAM* are positive and significantly different from zero, however, coefficients of *SCHIPOL* and *UTRECHT* are not significantly different from zero. Jointly, these five additional controls significantly add to the explanatory power of the regression ($F(5,1784)=42.42$). Given that their coefficients were insignificant, *SCHIPOL* and *UTRECHT* were dropped from subsequent analysis. Also *ROTTERDAM* was not included in the spatial analysis. Given that the city of Rotterdam is part of the study region, distance of every zip code to the harbour is accounted for in the weight matrices.

b. Spatial Lag Estimates

Estimation of spatial lag models requires a decision regarding the specification of the weight matrix, *W*. Many defendable alternatives can be developed (see Anselin 1988 for further discussion). For example, *W* could be specified as a contiguity matrix with elements of ones and zeros indicating whether the borders of zip code areas touch each other. Or, elements of *W* could be chosen to indicate whether any pair of zip code areas is located within some number of kilometres of each other. These alternatives, however, were not pursued in order to focus more directly on specifications of W that posit different mechanisms by which growth might be transmitted across space and allow that growth effects decay with distance. Estimates presented below are based on two weight matrices. W_1 is a relatively sparse matrix that allows for growth in a particular zip code-industry to be directly transmitted only to that same industry in other zip codes. Thus, this matrix has predominantly zero elements. Remaining elements in W_1 show the reciprocal of distance (in kilometres) between pairs of zip codes. W_2, on the other hand, is a relatively dense matrix that allows growth in a zip code-industry to be directly transmitted to all other industries, including those in the same zip code. Thus, elements of W_2 reflect the reciprocal of distance (again, measured in kilometres). Industries in the same zip code were assumed to be less than one kilometre apart. W_1 is similar to the GKSS formulation, except elements are distance weights, rather than by employment shares. W_2 was selected mainly to determine whether zip code-industry growth rates are determined by spatial lags that operate through all industries in all locations.

Spatial lag estimates using row-standardised wieght matrices W_1 and W_2 are presented in the first two columns of Table 2. Estimation is by maximum likelihood. The coefficient of determination, presented only to give a crude indication of goodness-of-fit, is somewhat higher in the column (2) regression as compared with the column (1) regression (0.161 vs. 0.123). However, spatial autocorrelation in the residuals is detected in the column (2) regression using the Lagrange Multiplier test (LM(SE)) described by Anselin (1988, pp.105-106). In consequence, this equation was re-estimated as the spatial error model shown in equation (3).

(3) $$y=X\beta+\varepsilon \quad \text{where} \quad \varepsilon=\lambda W\varepsilon+\mu$$

In equation (3), y, X, β, and ε are defined just as they were in equation (2), however, the error vector, ε, is the outcome of a spatial autoregressive process involving a weight matrix (W) and a spatial autoregressive coefficient (λ). Results of estimating the spatial error model by setting $W=W_1$ are presented in column (3) of Table 2. Coefficient estimates of the explanatory variables are quite similar to those obtained for the spatial lag model. The estimate of the spatial error parameter (λ) is 0.48, which is significantly different from zero at conventional levels. An LM test (LM(SL)), however, shows that a spatial lag is present in y. Together, the regressions in column (2) and column (3) suggest that a mixed spatial lag-spatial error model should be estimated. This complex specification is not pursued in light of the computational burdens involved and discussion below is directed mainly to the spatial lag estimates in columns (1) and (2). Also, in all three regressions reported in Table 2, the

Breusch-Pagan (1979) test rejects the null hypothesis of homoskedasticity at conventional significance levels. Thus, heteroskedasticity remains a problem for further investigation.

Both the column (1) and column (2) regressions suggest the existence of growth spillovers between zip code areas. Using the weight matrix W_2, the estimate of the spatial lag parameter (ρ) was 0.08 and significantly different from zero at the 1% level. The corresponding estimate of ρ using the matrix W_1 was 0.45 and also was significantly different from zero at 1%. Implications of these estimates for the extent of spatial knowledge spillovers are treated more fully momentarily. Also, coefficients of the three knowledge externality variables *CONCENTRATION, COMPETITION*, and *SHARE* are not greatly affected by the choice of the weight matrix. Moreover, these coefficient estimates are the same sign and order of magnitude as the ordinary least squares regression results reported in Table 1. Thus, support for Jacobs externalities carries over to the spatial lag estimates. Coefficient estimates for the remaining control variables are broadly similar between the column (1) and column (2) regressions in Table 2 as well as between these two regressions and the ones reported in Table 1.

The spatial lag results can be used to estimate the extent to which knowledge spills between zip code areas. As shown in Section 2 in connection with equation (2), this calculation involves summing the elements in each column of $A=(I-\rho W)^{-1}$. In the regression using the weight matrix W_2 reported in column (1), column sums of A ranged from a low of 1.34 to a high of 2.46 and averaged 1.94. In the regression using the weight matrix W_1 reported in column (2), column sums of A exhibited almost no variation (recall that W_1 was quite sparse; in fact only 7% of its cells are non-zero) and averaged 1.82. These results suggest that, on average, more than half of any change in *CONCENTRATION*, *COMPETITION*, or *SHARE* would be felt in the zip code area where it occurred. Thus, a key finding here is that knowledge externalities associated with industrial diversity and competition between firms are highly localised. This finding regarding localisation of externalities is similar to the results of Jaffe, Trajtenberg, and Henderson (1993) and Anselin, Varga, and Acs (1997), but is obtained using implications from an analysis of growth, rather than a study of patent citations or the geographic distribution of innovations. Also, this outcome is particularly significant in the present context because of the small geographic size of the zip code areas studied and suggests that competitive, highly diversified industrial areas with relatively high knowledge generating capacity tend to grow at faster rates than other areas.

c. Sources of Growth

In the previous two subsections, we analysed the change in (log) employment over the period 1988-1997 arising from all establishments. Additional insight can be gained by disentangling the employment change arising from existing firms (*i.e.*, firms that already existed in 1988) and the change arising from new firms. Over the period 1988-1997, total employment in the six largest sectors in all 416 zipcodes (with more than 50 employees) decreased from 722,631 to 693,688 full time jobs. Given the fact that employment in incumbent firms decreased to 579,230, new establishments accounted for an increase in employment of 114,458. Many factors affect the latter, such as, for example, managerial decisions about firm location. In this analysis of the sources of growth, we treat the increase in employment in new establishments as a residual. We have a run similar regressions as in section 4b, but as dependent variable we used the change in (log) employment over the 1988-1997 period in establishments that were founded before 1988 and that survived at least up until 1997. By comparing these regression results with the results in section 4b, we can infer the contribution of new firms to employment growth.

The regression results using employment growth in 'old' establishments are presented in Table 3, where the first column represents the spatial lag analysis assuming within *and* between industry spatial spillovers and where the second column represents the same analysis assuming that spillovers can only occur within the same industry. Comparing these spatial lag regression results for 'old' firms with those for all firms (see the first two columns in table 2), it is clear that results are very similar. Concerning the spillovers, old establishments grow fastest (i) the more competition they face (see *COMPETITION*), (ii) the less specialised the industry (see *CONCENTRATION*), and (iii) the more diverse zipcode (see *SHARE*). Additionally, the coefficients for the variables controlling for initial differences are very similar in terms of both the sign of the coefficients and the coefficient values. Also the spatial variables are of approximately the same magnitude, although the ρ-coefficient for intra-industry spatial spillovers seems somewhat higher for old establishments (0.54) than for all establishments (0.45). The main difference between the two tables is impact of the rate of change of the industry wage rate. Wage increases are more detrimental for growth in the case of all establishments than in the case of old firms. Finally, note that similar to the analysis of growth in all establishments, the Lagrange Multiplier test on spatial error dependence indicates that there is spatial correlation in the error terms in the case of intra-industry spillovers (W_1). The regression results for the spatial error model presented in the third column that all conclusions remain the same. The spatial coeffcient (λ) is somewhat higher for old firms than for all firms, while the negative impact of wage rate increases is much lower for the former than for the latter.

Therefore, it can be concluded that employment growth arising from new establishments is driven by essentially the same mechanisms as growth in incumbent establishments, although spatial correlation seems slightly more important while higher rates of wage increases discourages establishing new firms more than it retards expansion of existing firms.

5. Conclusions

This paper brings together two important strands of literature on the relationship between knowledge spillovers and employment growth. The first strand tests for evidence of endogenous growth linked to knowledge and knowledge spillovers between economic agents within cities and the second tests whether knowledge spills over between economic agents in different locations. The link between these two topics is made by extending the work of Glaeser, Kallal, Schienkman, and Schliefer (1992) to develop a spatial lag model that allows employment growth in one location to affect growth in other locations. The empirical work presented focuses on the province of South-Holland, the Netherlands. An advantage of is that it is small enough that location-specific factors, such as those related to cost, demand, and resource endowments, do not differ greatly between locations in the province. In consequence, the choice of an area to study removes an important objection to previous studies on the role of knowledge externalities in urban growth.

A key finding reported here is that local industrial diversity, rather than industrial concentration, and increased local competition tend to promote growth. This outcome is consistent with results of Glaeser, Kallal, Schienkman, and Schliefer (1992), but are not in agreement with results presented by Henderson, Kuncoro, and Turner (1995) who find that industrial concentration plays a more important role than industrial diversity. These results apply to existing firms that operated in the region throughout the sample period Additionally, results in this study suggest that knowledge spillovers between locations can occur indirectly through the growth process. In other words, increased spillovers can lead to increased growth in one location and increased growth in that location can lead to increased growth elsewhere. However, empirical estimates suggest that less than half of the growth induced by knowledge spillovers is transmitted to other locations in this way. This finding is particularly significant in light of the small size of the regions studied and suggests that areas with a relatively high knowledge generating capacity may be expected to grow faster than others.

REFERENCES

Anselin, Luc. 1988. *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, Dordrecht.

Anselin, L. (1995), *SpaceStat. A software program for the analysis of spatial data (version 1.80)*. Regional Research Institute, West Virginia University, Morgantown WV.

Anselin, Luc, A. Varga, and Zoltan Acs. 1997. Local geographic spillovers between university research and high technology innovations, *Journal of Urban Economics* 42: 422-48.

Arrow, Kenneth J. 1962. The economic implications of learning by doing. *Review of Economic Studies* 29: 155-173.

Audretsch, David B. and Maryanne P. Feldman. 1996. R&D spillovers and the geography of innovation and production, *American Economic Review* 86: 630-40.

Beardsell, Mark and J. Vernon Henderson. 1999. Spatial evolution and the computer industry in the USA, *European Economic Review*, 43: 431-56.

Black, Duncan and J. Vernon Henderson. 1999. Spatial evolution of population andindustry in the United States, *American Economic Review* 89: 321-27.

Blanchard, Olivier J. and Lawrence F. Katz. 1992. Regional evolutions, *Brookings Papers on Economic Activity* 1-76.

Bostic, Raphael, W. Joshua S. Gans, and Scott Stern. 1997. Urban productivity and factor growth in the late nineteenth century, *Journal of Urban Economics*, 41: 38-55.

Breusch, T. and A. Pagan. 1979. A simple test for heteroskedasticity and random coefficient variation, *Econometrica* 47: 1287-94.

Ellison, Glenn and Edward L. Glaeser. 1999. The geographic concentration of industry: Does natural advantage explain agglomeration, *American Economic Review* 89: 311-16.

Feldman, Maryanne P. and David B.Audretsch. 1999. Innovation in cities: Science-based diversity, specialization, and localized competition, *European Economic Review*, 43: 409-29.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman and Andrei Shleifer (1992), Growth in cities. *Journal of Political Economy* 100, pp. 1126-1152.

Glaeser, Edward L. 1998. Are cities dying? *Journal of Economic Perspectives*, 12: 139-60.

Henderson, J. Vernon, A. Kuncoro, and M. Turner. 1995. Industrial development in cities, *Journal of Political Economy* 103: 1067-85.

Henderson, J. Vernon. 1997. Externalities and industrial development, *Journal of Urban Economics* 42: 449-71.

Houston, David B. 1967. The shift-share analysis of regional growth: a critique, *Southern Economic Journal* 33, 577-81.

Jacobs, Jane. 1969. *The economy of cities*. New York: Vintage.

Jaffe, Adam B. 1989. Real effects of academic research, *American Economic Review*, 79: 957-70.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 63: 577-98.

Kim, Sukkoo. 1999. Regions, resources, and economic geography: Sources of U.S. regional comparative advantage, 1880-87, *Regional Science and Urban Economics* 29: 1-32.

Lucas, Robert E. 1988. On the mechanics of economic development. *Journal of Monetary Economics* 22: 3-42.

Lucas, Robert E. 1993. Making a miracle. *Econometrica* 61: 251-72.

Marshall, Alfred. 1890. *Principles of economics*. London: Macmillan.

Porter, Michael. 1990. *The competitive advantage of nations*. New York: Free Press.

Romer, Paul M. 1986. Increasing returns and long run growth, *Journal of Political Economy*, 94: 1002-1037.

RPD (1998), *Werklocaties 1998*. Rapportage bedrijventerreinen en voorradige kantoorlocaties (IBIS), Rijksplanologische Dienst, 's-Gravenhage.

Van Oort, F.G., G.A. Van der Knaap and W.F. Sleegers (2000), New firm formation, employment growth en the local environment: empirical observations in South-Holland. Chapter 8 in: J. Van Dijk and P. Pellenbarg (eds.), *Demography of firms and the space economy*. Netherlands Geographical Studies, Amsterdam/ Groningen.

WMD (1999), *Woonmilieudatabase 1998. Toelichting*. ABF-Onderzoek & RPD, Delft/Den Haag.

APPENDIX A

Data for the 1988-1997 regional analysis (South-Holland) of endogenous growth in location-industries

The dependent and most of the explanatory variables are derived from the longitudinal dataset of the Firm Register South-Holland (BZH). The variables EMPLOYMENT GROWTH, CONCENTRATION, COMPETITION, SHARE, and EMPLOYMENT88 are calculated from these data. The data concerning agriculture were derived from the Agricultural Statistics for Municpalities (Landbouwtelling) of the Central Bureau of Statistics (CBS), and localised to 4-digit ZIP-codes on the basis of the CBS Land Use Statistics (Bodemstatistiek, CBS). Various other sources have been consulted to construct the remaining variables. The wage data have been obtained from CBS Labour Statistics.

The index and dummy variable of 4-digit ZIP-codes characterised by more than average issue of new business premises is constructed making use of the Inventory of Industrial Sites (RPD 1998). The distinctions between rural and urban areas (URB) and the distinction between working areas and (predominantly) residential areas (WORKAREA) are obtained from the Living Environment Database (WMD 1999). This distinction urban-non urban is based on the density of addresses as registered by the Central Bureau of Statistics (CBS), using a potential function for its final estimation. The distinction in activity dominance per locality is based on the employment and residential functions, as well as the provision and accessibility of (daily) services (retail-shops, recreation, health-care, administrative functions and public transportation). All variables measuring physical distances (such as the distances between the zipcodes and the distance to Rotterdam, Utrecht and Schiphol) were constructed using Atlas*GIS, ArcInfo and ArcView geographical information systems, as well as internal calculations (of distance- and weight matrices) in the statistical package SpaceStat (Anselin 1995).

TABLE 1

DETERMINANTS OF ZIP CODE-INDUSTRY GROWTH: INITIAL SPECIFICATION

(t-values are presented in parenthesis)

EXPLANATORY VARIABLE	MEAN	(1)	(2)
CONSTANT	----	1.318 (3.63)	0.178 (0.35)
CONCENTRATION	4.823	-0.016 (-4.49)	-0.016 (-4.52)
COMPETITION	1.129	0.182 (6.39)	0.198 (6.53)
SHARE	0.590	-0.740 (-3.73)	-0.800 (-3.90)
GROWTH	0.082	0.770 (3.40)	0.925 (3.98)
WAGE88	46.256	-0.014 (-4.31)	-0.013 (-3.77)
ΔWAGE	0.293	-2.267 (-3.28)	-1.800 (-2.55)
EMPLOYMENT88	402.132	-6.484E-06 (-0.15)	-5.113E-05 (-1.07)
WORKAREA	0.262		0.217 (3.38)
SITES	0.151		0.213 (3.64)
ROTTERDAM	21465.48		1.581E-05 (3.03)
SCHIPOL	51393.01		7.741E-06 (1.64)
UTRECHT	52615.18		2.230E-06 (0.70)
SUMMARY STATISTICS			
N		1797	1797
R^2		0.126	0.143

TABLE 2
SPATIAL LAG AND SPATIAL ERROR ESTIMATES
(t-values are presented in parenthesis)

EXPLANATORY VARIABLE	SPATIAL LAG (W_2)	SPATIAL LAG (W_1)	SPATIAL ERROR (W_1)
CONSTANT	1.174 (5.94)	0.820 (2.85)	1.475 (2.95)
CONCENTRATION	-0.017 (-9.68)	-0.014 (-7.82)	-0.014 (-7.44)
COMPETITION	0.212 (7.57)	0.172 (6.25)	0.184 (6.33)
SHARE	-0.799 (-4.13)	-0.670 (-3.58)	-0.691 (-3.57)
WAGE88	-0.017 (-5.31)	-0.008 (-2.54)	-0.017 (-2.87)
ΔWAGE	-2.686 (-5.05)	-1.242 (-2.33)	-2.533 (-2.59)
EMPLOYMENT88	-1.142E-05 (-0.24)	-5.907E-05 (-1.28)	-4.510E-05 (-0.94)
SITES	0.179 (2.67)	0.211 (3.49)	0.227 (3.52)
WORKAREA	0.150 (2.54)	0.199 (3.49)	0.225 (3.85)
SPATIAL COEFFICIENT	0.080 (2.36)	0.451 (9.27)	0.476 (9.13)
SUMMARY STATISTICS			
N	1797	1797	1797
R^2	0.123	0.161	0.094
LM(SE)	0.73	76.38	
LM(SL)			88.65
LM(BP)	285.01	297.03	298.93

TABLE 3
SPATIAL LAG ESTIMATES FOR GROWTH IN INCUMBENT ESTABLISHMENTS
(t-values are presented in parenthesis)

EXPLANATORY VARIABLE	SPATIAL LAG (W_2)	SPATIAL LAG (W_1)	SPATIAL ERROR (W_1)
CONSTANT	1.412	0.499	0.892
	(4.31)	(1.61)	(1.34)
CONCENTRATION	-0.016	-0.013	-0.013
	(-8.14)	(-6.53)	(-6.18)
COMPETITION	0.199	0.159	0.181
	(6.30)	(5.24)	(5.563)
SHARE	-1.031	-0.869	-0.901
	(-4.72)	(-4.18)	(-4.18)
WAGE88	-0.015	-0.005	-0.014
	(-4.16)	(-1.50)	(-1.78)
ΔWAGE	-1.860	-0.531	-1.477
	(-3.10)	(-0.92)	(-1.13)
EMPLOYMENT88	-9.6911E-06	-6.463E-05	-4.0516E-05
	(-0.18)	(-1.26)	(-0.76)
SITES	0.193	0.244	0.267
	(2.55)	[i](3.39)	(3.74)
WORKAREA	0.137	0.203	0.232
	(2.06)	(3.21)	(3.59)
SPATIAL COEFFICIENT	0.086	0.539	0.568
	(3.42)	(12.54)	(12.77)
SUMMARY STATISTICS			
N	1797	1797	1797
R^2	0.094	0.150	0.068
LM(SE)	0.50	42.12	
LM(SL)			76.32
LM(BP)	321.89	305.18	303.00[9]

i

9