

**Uwe Blien, Kai Kirchhof, Oliver Ludewig:<sup>1</sup>**

(Institute for Employment Research, IAB)

## **Agglomeration effects on labour demand**

Draft paper prepared for presentation at the CAFÉ, Nuernberg 2006

17 March 2006

### **Abstract**

**Uwe Blien, Kai Kirchhof, Oliver Ludewig:**

(Institute for Employment Research, IAB)

## **Agglomeration effects on labour demand**

### Abstract

How do agglomeration effects influence the demand for labour? To answer this question, approaches on labour demand are linked with an analysis of the classic “urbanization effect”. We use models for static and for dynamic labour demand to find out, whether agglomerations develop faster or slower than other regions. Estimations of the static model show the influence of different degrees of regional concentration at the employment level. The model of dynamic labour demand is used to estimate the effect of different regional types on the growth rate of labour demand.

The empirical results (received with the linked employer-employee database of the IAB) on long-run or static labour demand indicate substantial agglomeration effects, since c. p. employment is higher in densely populated areas. In the dynamic model, however, labour demand in core cities grows slower than the average. This is not a contradiction. Labour demand is especially high in large cities, but the other areas are slowly reducing the gap.

---

<sup>1</sup> The authors thank L. Dirnfeldner, H. Sanner, J. Suedekum, and K. Wolf for very valuable advice and assistance. Any responsibility for the analysis and the presentation remains with the authors.

## 1. Introduction

Empirical and theoretical analyses on labour demand are often carried out without any specific reference to the regional dimension of the labour market. This dimension is, however, of considerable importance, as can be seen from a new debate about the effects of regional concentration on employment. The debate was started by seminal papers in the *Journal of Political Economy* by Glaeser et al. (1992) and Henderson et al. (1995). There is a new and expanding literature about different kinds of agglomeration (urbanization/ localization) effects on economic activity which derives novel results from ideas dating back even to Marshall. This literature includes contributions from the *New Economic Geography* (Krugman 1991) and from other theoretical and empirical work.

In this paper we intend a fusion of standard approaches on labour demand with the literature on agglomeration effects. This fusion has its advantages: In the literature on agglomeration effects it is normally not possible to control for the exact nature of the externality that gives rise to agglomeration effects. Here, a detailed analysis of labour demand could give new insights.

On the other hand a labour demand function might be not completely specified if the regional context of a firm is not included. For example, the effects of technological change might be completely different depending on whether the firm operates in a favourable environment or whether it is rather isolated. The diffusion of technological improvements and its effects on employment need to be studied with respect to the regional context.

Therefore, this paper uses an integrated approach: A labour demand function is estimated which is extended to take the regional context into account. The data requirements of this approach are rather vast, since data on three levels have to be put together: data on employees, on establishments and on regions. The models used have to take care of the multilevel problem which must be solved to understand the relation between individual organizations and their contexts. Since in this study workers are nested within establishments and establishments within regions, it is necessary to observe effects due to the clustering of observations and due to the interaction of levels.

For the analyses we use the linked employer-employee database of the IAB (called LIAB, see Alda, Bender, Gartner 2005). This includes the IAB Establishment Panel with currently about 16,000 establishments in each of the yearly waves. The IAB Establishment Panel is based on personal interviews with leading representatives of establishments in the years 1993–2005. The questionnaire was designed to make available a comprehensive set of information for analyses of the labour market. The sample is representative for Germany. The panel is linked with data of the employment statistics which includes information about all workers covered by social security. Information about regions is also included in the database. These variables indicate the degree of concentration of economic activity.

## 2. Background

Currently a debate is going on about the effects of different kinds of externalities on the regional development of productivity and employment. What economic structure supports employment growth at the local level? Glaeser et al. (1992) argue that a diversified economic structure is advantageous, whereas the study of Henderson et al. (1995) finds that own industry specialisation is the major engine of employment growth.

In this paper we are interested in answers to a related, but not identical, question. We intend to study the effects of the size of the respective agglomeration, i.e. we look at the classical “urbanization effect”. Due to the typology of Krugman (1991) this is the effect associated with the sheer size of the local agglomeration, without any regard to its specialisation or diversity. In the approaches of New Economic Geography the size of a local economy is associated with an externality, since the concentration of production generates a concentration of consumers and the latter is favourable for the concentration of production. Therefore, a cumulative causation process gives rise to a centre/periphery structure.

The assumptions of the New Economic Geography are restrictive. Many industries produce for the world market and the local agglomeration of consumers is not very important. Apart from this there are “deglomeration” – e.g. congestion – effects working in the opposite direction. In densely populated areas the overcrowding of places has unfavourable consequences. Increasing prices of housing, traffic problems, competition of firms for qualified labour etc. increase the cost of production. Therefore, it is an empirical question whether agglomerations develop faster or slower than the rural country. Empirical studies undertaken by Möller, Tassinopoulos (2000) and Suedekum, Blien, Ludsteck (2006) for Germany show that employment in city centres has smaller growth rates than in the rest of the country.

This research is relevant for an assessment of political measures. In recent years older concepts of “growth poles” have been revitalised under new headings. Common to all these concepts is the proposition that a successful development policy should be concentrated on the large cities. This is behind the new emphasis placed on “Metropolitan Regions” in European (and in German) development programmes. It is at least part of the “cluster” concept on regional growth, since one of the meanings given to the rather evasive cluster term is “pure agglomeration” (McCann 2005). There has been a change in the direction of regional assistance programmes, since these are now oriented towards the most likely growth engines of the country and not towards fair regional distribution of economic activities. The assumption is that there are secondary effects working in favour of the rural country. These include spillovers from the centres. The Metropolitan Regions are expected to pull the other parts of the country to higher levels of growth. But there is doubt about the effectiveness of all these programmes. How could an agglomeration produce spillovers effective for growth if its own growth rate is smaller than the one of the rest of the country?

In many empirical tests agglomeration effects are measured using a pure cross-section approach, as long-run employment growth rates are regressed on control variables that reflect

the regional industry composition in some base year.<sup>2</sup> It is thus assumed that a historical pattern from 10–30 years ago affects employment growth, but no real test is provided about the relevant time structure. To be able to do such test, one needs data of local industries for many consecutive years in order to make full use of the three dimensions of the panel (location, industry, time period). An additional advantage of panel techniques is the possibility to control for time-invariant fixed effects that cannot be easily disentangled from the impact of the local economic structure in a cross-section analysis. This literature normally uses aggregated data on individual workers. Many controlling variables measured at the level of establishments that are required to estimate a standard labour demand function are ignored.

We are interested in filling this gap. Our model of labour demand follows the classic work of Hamermesh (1986, 1993) and Nickell (1986). A production function with capital and labour as the two input factors and the common properties is assumed. A firm trying to minimize costs for a given output will set the optimal level of capital and labour so that the marginal productivity of each factor equals its price. Taking the ratio of these first order conditions one obtains that the marginal rate of technical substitution equals the factor-price ratio in the optimum. This result can now be used together with the output constraint to derive the demand functions for capital and labour.

A simple case for specifying a labour demand function for an empirical model is to use a linear homogeneous production function of the following kind:

$$Y = A[\alpha L^\rho + (1-\alpha)K^\rho]^{1/\rho}. \quad (1)$$

There Y is the output of a specific firm, L is labour and K is capital.  $1 > \alpha > 0$ ,  $1 \geq \rho \geq -\infty$  and A is a technology parameter. After taking the partial derivative with respect to labour and using the neoclassical assumption that the marginal product of labour is equal to wages this function yields the labour demand equation:

$$L = A^{1-\rho} \alpha^{1-\rho} w^{-1/\rho} Y. \quad (2)$$

Taking logarithms results in a first approach to the linear function of the empirical model:

$$\ln L = \sigma^2 \ln A + \alpha^2 - \sigma \ln w + \ln Y \quad \text{with: } \sigma = \frac{1}{1-\rho} \quad (3)$$

This is a very simple function, which could be easily estimated. A problem is that the assumptions about the production process might not be exactly met. For example, the production function might not exactly show constant returns to scale. Therefore, it is advisable to use an estimation strategy which is robust against violations of the basic assumptions. At any rate it is necessary to extend the estimated function with respect to regional characteristics and other controlling variables. Agglomeration effects could be thought to be working through the pa-

---

<sup>2</sup> Both Glaeser et al. (1992) and Henderson et al. (1995) are cross-sections, as well as the influential study on France by Combes (2000). Among this literature is also the paper by Blien and Suedekum (2005) on Germany (1993–2001).

parameter  $A$ . Depending on regional characteristics labour demand might be higher or lower than the average.

### 3. The Empirical Model

In our empirical work two different versions of the labour demand function are applied. One is the static version giving the demand in the long run. The other one is the dynamic function which includes lags of the endogenous variable. One basic difference between the two specifications is that within static models parameters are estimated that concern the change in labour demand due to the long-run effects of external changes, whereas the dynamic model shows the growth of labour demand. Appropriately adapted static models show agglomeration effects with respect to the level of labour demand, whereas from the dynamic model the response in terms of the growth rate can be obtained.

In many cases it is regarded as unavoidable to estimate dynamic models because normally there is inertia in the development of labour demand. Then, a correctly specified model would include the lagged endogenous variable. In this case the standard fixed effects estimator could not be used, because it gives biased and inconsistent results (Baltagi 2001). Instead a GMM-estimator has to be applied (Arellano, Bond 1991).

#### 3.1 Models for static labour demand

All these models have to be adapted for the question at hand. In the case of the static function the fixed-effects estimator is useful to control for unobserved heterogeneity. We apply a two-step procedure to identify the effects of regional agglomerations on the labour demand of establishments. In the first step we use the panel structure of the data to extract the establishment fixed effect from a usual static labour demand function. We do so using the common within estimator. This is the first step equation:

$$\ln L_{it} = \beta_0 + \beta_w \ln w_{it} + \beta_Y \ln Y_{it} + \beta_X \ln X_{it} + \gamma_t + \varepsilon_{it} + \nu_i \quad (4)$$

Here  $i$  is the index for the establishment and  $t$  the index of time.  $X$  is a vector of time-varying variables which are added to equation (3) as additional controls.  $\varepsilon_{i,t}$  is the usual error term.  $\gamma$  is a vector of time dummies for the influence of the business cycle and  $\nu_i$  is the establishment fixed effect which reflects all time-invariant effects specific to the establishments. This includes things like a favourable location, an especially talented owner and market position within the industry as well as the influence of the regional conditions as summarized in agglomeration or suburbanization effects. Therefore the effect of the variable  $A$  in equation (3) is included in the fixed effect  $\nu_i$ . Since most establishments do not change their respective region a second step is required to identify agglomeration effects. The fixed effects are regressed to type of regions, some spell indicators and other firm-specific and time-constant variables  $Z$ :

$$v_i = \beta_0 - \beta_r D_r + \beta_z \ln Z_i + \beta_s S_i + \eta_i \quad (5)$$

The  $D$ s are dummies which represent the type of the respective region. Formally, they partly replace the parameter  $A$  of the theoretical model, which could have positive or negative effects on employment. The  $D$ s should represent the information about the degree of agglomeration which is characteristic for the region.

Using unbalanced panel data we have to add a further set of special controls. Due to the unbalanced time structure the different  $v_i$  are determined on the basis of different observation spells. Some establishments are observed from 1995 to 2001, others from 1996 to 1999 and so on. Thus different conditions at certain points of time and different observations spells might influence the value of  $v_i$  for each firm. We control for this by defining a dummy variable for each spell length and an interacting term with the diverse wave dummies yielding 21 spell indicators ( $S$ ). These are added to the regression function of the second step.

Besides these spell dummies and our main explanatory variable, the regional type in which an establishment is located, we add a set of control variables  $Z$  which are fixed over time or quasi-fixed. Quasi-fixed variables are those which only change for very few establishments at a point of time or very seldom or by very small amounts, like the existence or not-existence of a works council, or the industry or fraction of certain employee groups. Whether a variable is quasi-fixed or free over time is a matter of degree.

One final remark on this procedure: In the first step the coefficient  $\beta_Y$  is expected to be close to one. This might be not the case if the variable  $Y$  does not vary much in time. In this case part of its effect is included in the fixed effect.

### 3.2 Models for dynamic labour demand

If there is considerable inertia in the adaptation process a dynamic model might be appropriate for labour demand. In this case the lagged endogenous variable is included:

$$\ln L_{it} = \beta_0 + \ln L_{i(t-1)} - \beta_w \ln w_{it} + \beta_Y \ln Y_{it} + \beta_X \ln X_{it} + \gamma_t + \varepsilon_{it} + v_i \quad (6)$$

In principle the same two-step procedure could be used as in the static model. But we change the procedure to obtain information not only about agglomeration effects on the level of labour demand but also on its growth. This could be done in the following way. With GMM the above equation is differenced to eliminate the fixed effects. In this case the equation is formulated in differences of logs, i.e. in approximations of growth rates. It would be informative to have the effect of agglomerations on the growth rate of labour demand. This could be done by including a specific trick introduced by Nickell et al. (1992). To avoid the elimination of the time-invariant variables, they included interactions of time-constant variables with a time index  $t$ . We do the same:

$$\ln L_{it} = \beta_0 + \ln L_{i(t-1)} - \beta_w \ln w_{it} + \beta_Y \ln Y_{it} + \beta_X \ln X_{it} + t\beta_r D_r + \gamma_t + \varepsilon_{it} + v_i \quad (7)$$

Now we gain the effect of a time-constant dummy variable representing the type of the respective region (in which the establishment  $i$  is located) on the growth rate of labour demand.

No second step is required. Since equation (7) is estimated by taking differences, the effect of a special degree of agglomeration on the growth rate of labour demand is estimated. This is more closely related to the current literature on agglomeration effects than the estimates obtained with the static model.

In a last remark we address the multilevel structure of the problem. Moulton (1990) is famous for showing that the inclusion of variables related to different levels of observation, here regions and establishments, could result in inefficient estimates of the coefficients and in biased estimates of the standard errors especially of the variables measured at the higher level. He recommends the inclusion of fixed effects for the higher-level units. This is redundant in our case since we include fixed effects for establishments. If there were no relocation of establishments, regional fixed effects would be perfectly multicollinear with establishment fixed effects. In our case with rare movement of establishments they are highly multicollinear.

#### **4. The Data**

We use the so called IAB Establishment Panel (IAB-Betriebspanel, see Bellmann 1997 and Kölling 2000) as one basic data source. It is extended to a employer-employee linked panel by linking it with the employment statistics of Germany. The IAB Establishment Panel is a general purpose survey based on a random sample giving longitudinal information in yearly waves for the time since 1993 in West Germany and since 1996 for East Germany. It contains a broad range of variables regarded as important in economic theory. It includes establishments of all sizes, and is not restricted to manufacturing. These basic structural elements correspond to some of the principles of an ideal set of establishment data suggested by Hamermesh (1993). An establishment as it is counted in the panel is the local plant of a firm. It might be identical with the entire firm or it might be a part of it.

Starting with 4,300 establishments, the sample size of the survey was extended in several steps. Currently, it covers about 16,000 establishments in its yearly waves. Most of the information is collected by trained interviewers. Only in some regions the sample size is extended by data collection through mailed questionnaires. The base population consists of all establishments with at least one employee covered to the compulsory social security system. Over 80% of the German establishments fulfil this condition. Since the survey is supported by the German employers' association and Federal Labour Agency (Bundesagentur für Arbeit), there is a rather high response rate of around 70% for initial contacts and about 80% for the annually repeated contacts. The establishment panel provides general information on the establishments, such as organizational practices, total sales, employment or the industrial relations within the establishment.

The second data set is the so called Employment Statistics (Beschäftigten-Leistungsempfänger-Datei). This is a database generated for administrative purposes and therefore especially reliable. Pensions are computed from the original data. All employees are included who are covered by the social security system. This database comprises information on gen-

der, wage, age, occupation and qualification of the employees. Thus a rich personalized database is generated.

The IAB Establishment Panel and the Employment Statistics are linked (forming the LIAB) by a unique establishment identification number. Thus it is possible to match the information of all employees covered by the social security system with the establishments of the IAB Establishment Panel. In doing so, we add the averages across an establishment of variables from the employment statistics as new variables to the establishment panel. Variables giving establishment characteristics, like total sales or existence of a works council, stem from the establishment data.

The establishment panel starts in 1993. We use data of the Employment Statistics Registry until 2002. Thus our time window is ranging from 1993 to 2002. However, some questions of the survey are backward looking, such as “What were your total sales last year?” Thus we have to transfer some of the information of  $t+1$  to  $t$ , generating missings for establishments not observed in  $t+1$ .

The panel is unbalanced due to panel mortality, missing values on some variables and new entrants to the panel. Therefore it is necessary to control the effects of different observation times and spell lengths. We do so by introducing time dummies in the first step analysis and the spell indicators described above in the second step analysis.

While this data set is rather large and representative for Germany it is not possible to use all observations. We exclude the agricultural and mining sector, non-profit organisations and state agencies as well as observations with missing values on variables used in the estimations. Establishments with only one or two observations are also excluded to get a broader base for the fixed effects estimator. This leaves us with 6,532 establishments observed over an average of 4.8 waves, giving a total of 31,509 observations. The minimum length of a spell is 3 years, the maximum length is 10 years. Table 1 gives the descriptive statistics of the variables used and indicates the source data set.

Let's take a closer look on the variables. As mentioned above, the wage variable is taken from the registry data and averaged across employees of each establishment. The qualification level of each employee is also provided by the registry data. The qualification levels are increasing from one (low skilled) to 6 (university degree). Employees without information about their qualification are put into the category 7. These are mostly unskilled persons. We calculated the shares of each qualification level for each establishment. The same procedure was conducted with the women's share and the share of part time employees.

We use also the industry classification of the registry data, since it is more detailed than the one of the establishment panel and since the IAB establishment panel is providing one set of industry classification until 1999 and another one from 2000 onwards. The industry classification is used to generate 77 dummies. The share of the service sector establishments, which is about 43% of all observations, is also calculated using the industry classification. The share of West German establishments (57% of all observations) is calculated on the basis of the employee data, which provides the regional location of the workplace. The industry structure



might be very important with respect to differing patters of product demand and technical progress which influence employment (cf. Blien, Sanner 2006).

**Table 1: Summary statistic of the data set**

	All establishments (6,532), 31,509 observations				Source
	Mean	Std. Dev.	Min	Max	
Employment	246	1,166	1	57,154	establishment data
Total sales	44,600,000	260,000,000	2,000	12,700,000,000	establishment data
Average wage	58.032	24.157	0	148	employee data
Women's share of employment	38.271	32.076	0	100	employee data
Share of part-time work	13.506	20.451	0	100	employee data
Qualification typ 1 (share)	7.469	15.844	0	100	employee data
Qualification typ 2 (share)	35.686	33.757	0	100	employee data
Qualification typ 3 (share)	35.912	37.438	0	100	employee data
Qualification typ 4 (share)	1.580	6.097	0	100	employee data
Qualification typ 5 (share)	2.649	7.576	0	100	employee data
Qualification typ 6 (share)	2.657	7.425	0	100	employee data
Works council	0.361	0.480	-	-	establishment data
Type 1 regions:	0.206	0.405	-	-	BBR
Type 2 regions:	0.112	0.315	-	-	BBR
Type 3 regions:	0.059	0.235	-	-	BBR
Type 4 regions:	0.052	0.222	-	-	BBR
Type 5 regions:	0.092	0.288	-	-	BBR
Type 6 regions:	0.180	0.384	-	-	BBR
Type 7 regions:	0.120	0.325	-	-	BBR
Type 8 regions:	0.093	0.291	-	-	BBR
Type 9 regions:	0.086	0.281	-	-	BBR
State of equipment / level 1 (share)	0.216	0.412	-	-	establishment data
State of equipment / level 2 (share)	0.489	0.500	-	-	establishment data
State of equipment / level 3 (share)	0.267	0.442	-	-	establishment data
State of equipment / level 4 (share)	0.025	0.157	-	-	establishment data
State of equipment / level 5 (share)	0.002	0.050	-	-	establishment data
Number of spells	5.598	2.129	3	10	
		Up to 9 year dummies			generated
		Up to 77 industry dummies			employee data

The establishment panel also provides very important variables. The employment of the establishment is one, also total sales. Another variable reflects an important feature of industrial relations in Germany. This is a dummy indicating the existence of a works council. It is coded 1 (a works council exists) and 0 (no works council). 36% of the observations have a works council. Since this variable is asked biannually, every second year is missing. We circumvent this problem by relying on the substantial inertia of such an institution and fill the missing values in t+1 with values for t. The state of equipment is a categorical variable which reflects the modernity of the real capital. It is ranging from one (state of the art) to five (out-dated).

We use one as reference category and insert dummies for the four remaining levels into (some of) our empirical specifications.

Spell length indicates the number of observations per establishment. The average based on all observations is 5.6. This is more than the average number of waves calculated above on basis of the number of establishments, because establishments with longer spells provide by definition more observations. Depending on the length of the spells and their starting point we define up to 21 identifiers of spells with different length and starting years. These spell identifiers enter as dummies into our estimation.

In addition to information about individual workers and establishments data on regions are used for the analysis. In fact, this is the most important information for the research question. To analyze the effects of economic concentration, appropriate regional units have to be defined first. If large or heterogeneous regions were used, the effects would be blurred. To avoid this problem we use districts (= “Landkreise und kreisfreie Staedte”, NUTS III regions), i.e. 439 small regions that are rather homogeneous. Districts are administrative units of the German government. Larger cities form their own districts. In rural areas districts combine small towns, villages and the area between them.

**Table 2: Characterization of regions**

Regional types	Description	Number of establishments
Type 1 regions:	Core cities in regions with major agglomerations	1337
Type 2 regions:	Very densely populated districts in regions with major agglomerations	698
Type 3 regions:	Densely populated districts in regions with major agglomerations	380
Type 4 regions:	Rurally structured districts in regions with major agglomerations	365
Type 5 regions:	Core cities in regions with conurbational features	593
Type 6 regions:	Densely populated districts in regions with conurbational features	1189
Type 7 regions:	Rurally structured districts in regions with conurbational features	778
Type 8 regions:	Densely populated districts in rurally structured regions	601
Type 9 regions:	Rurally structured districts in rurally structured regions	591

(Classification following Goermar and Irmen 1991)

To map agglomeration effects a widely used classification of German districts (Goermar and Irmen 1991) provided by the Federal Office for Building and Regional Planning (BBR) is adopted. As can be seen from Table 2 the classification is based on the density of the population and the centrality of the location. We define eight dummy variables indicating the types 2 to 9. Thus, we are using the core cities in regions with major agglomerations as the reference category. These are cities with at least 300,000 inhabitants.

The use of the typology in table 2 has advantages compared to the direct inclusion of single indicators like population density or population size. These variables often give an erroneous picture of the regional units. The definition of regions does not follow a stringent criterion, but historical idiosyncrasies and administrative purposes applied differently in different part of the country. Population density might vary very much for a core city, since in one case the surrounding country is included in others not.

## 5. Results

The model for static or long run labour demand gives a first impression of differences between the rural country and the agglomerations with respect to the level of employment. In order to identify these regional differences we apply a two-step procedure as described in section 3. In our first step we estimate a common fixed effects model (table 3a).

**Table 3a: Static labour demand, first step: fixed effects – all establishments**

Fixed effect regression of static labour demand		
	Number of observations	31,509
	Number of groups	6,532
	F(23,24954)	314
	Prob > F	0.000
	R-sq: within	0.224
Dependent Variable: Employment (logarithm)		
	Coef.	t-value
Total sales (logarithm)	0.320	73.820
Average wage (logarithm)	-0.033	-3.910
Women's share of employment (logarithm)	0.009	3.600
Share of part-time work (logarithm)	0.036	18.620
Qualification typ 1 (share / logarithm)	0.044	19.830
Qualification typ 2 (share / logarithm)	0.034	18.230
Qualification typ 3 (share / logarithm)	0.020	8.780
Qualification typ 4 (share / logarithm)	0.012	4.340
Qualification typ 5 (share / logarithm)	0.018	7.170
Qualification typ 6 (share / logarithm)	-0.005	-1.760
State of equipment / level 2	-0.003	-0.790
State of equipment / level 3	-0.027	-5.340
State of equipment / level 4	-0.033	-3.060
State of equipment / level 5	-0.054	-1.760
	9 Year dummies	
Constant	-1.268	-17.410
Source: own calculations, LIAB waves 1993-2002		

The coefficients of total sales and wages have the right sign; however, the coefficient of total sales is relatively small. As discussed above this might be due to the fact that the fixed effect is capturing part of this relationship. Estimating the same function without fixed effects yield

coefficients about 0.8 for total sales, thus, supporting our hypothesis. Since our focus is not the coefficients of the labour demand equation, we include fixed effects to control for unobserved heterogeneity.

In the second step the fixed effects estimated in the first step are regressed on the regional types described above and on some control variables (table 3b). To facilitate interpretation of the results we use a transformed version of the fixed effects. The first-step equation is in logs, therefore we use the exponentiated values of the fixed effects.<sup>3</sup> Additionally to our regional types we include the works council variable as well as 75 industry dummies and 21 spell identifier as time invariant control variables.

All coefficients of the regional dummies are negative and significant. The reference category is core cities in large agglomerations. Thus, *ceteris paribus* the employment level of establishments located there is on average higher than the level in other regions. This might concern employment in general. Another explanation would be that many firms localize their headquarters, central administrations, central development units in large cities, whereas plants with reduced functions are placed elsewhere. This might be due to the person-to-person contact that is required with units close to the external market. It is also necessary with development units which are appropriately placed in locations with other firms and universities.

**Table 3b: Static labour demand, second step: analysis fixed effects – all establishments**

OLS Regression of the fixed effects with heteroskedasticity robust standard errors		
	Number of obs =	6,532
	F( 85, 6132) =	8.67
	Prob > F	0.00
	R-squared	0.14
Dependent variable:		
Fixed effect	Coef.	t-value
Type 2 regions	-172.122	-2.790
Type 3 regions	-173.008	-2.620
Type 4 regions	-182.674	-4.440
Type 5 regions	-136.106	-2.550
Type 6 regions	-195.681	-4.180
Type 7 regions	-167.507	-3.580
Type 8 regions	-196.181	-3.630
Type 9 regions	-166.393	-3.910
Works council	320.320	15.640
	21 spell identifying dummies	
	7 dummies for spell length	
	7 dummies for spell starting point	
	75 industry dummies	
Constant	1,194	3.000
Source: own calculations, LIAB waves 1993-2002		

<sup>3</sup> The second step with fixed effects which are not transformed gives basically the same results.

Thus, the static analysis of labour demand gives agglomeration effects. Congestion effects seem to be smaller than the advantages of a large city, at least with respect to the employment criterion. On a first glance the agglomeration hypothesis is supported.

Now we look at the effects of agglomerations on employment growth in a dynamic model, applying the mentioned trick of Nickell et al. (1992). Using the dynamic approach has the additional advantage of taking care of possible inertia in labour demand of the individual firms. We estimate two versions of the dynamic panel model with the Arellano-Bond estimator. The first is a rather parsimonious model. We only include total sales, wages (both in logs), wave dummies and the regional types in addition to the lagged values of the dependent variable. Total sales and average wages are instrumented by lags of their own levels. Thus we are accounting for the predetermination of wages and sales.<sup>4</sup> This model specification is then applied to three different (sub-)samples: all establishments, only manufacturing and only services.

The results (table 4) for the whole sample and for services include coefficients for the region types which are positive indicating that average employment growth is greater for all establishments not located in core cities in regions with major agglomerations. However, for the whole sample only the coefficients on regional type 2 and 3 are significant. Thus establishments in areas in the vicinity of large agglomerations are growing especially fast (or are shrinking slower than average).

For the service sector almost all coefficients are significant. Employment in the service sector is developing better in all regional types than in the core cities. This effect is especially strong in densely populated districts in regions with conurbational features. These results show suburbanization processes.

The findings with respect to the manufacturing sector are inconclusive. The larger part of the coefficients is positive, but they are all insignificant.

However, these results might be affected by an omitted variable bias. Therefore we estimate a more comprehensive model. We include controls for the women's share, part time share, qualification levels, state of the equipment and industry.

---

<sup>4</sup> The predetermination assumption of these variables is supported by a substantial higher p-value of the Sargan test compared to a model with wages and sales as strictly exogenous variables.

**Table 4: A parsimonious dynamic model (one-step results)**

GMM estimates with heteroskedasticity robust standard errors						
	all establishments		manufacturing		services	
Number of obs =	10,771		5,919		4,779	
Number of groups =	3,897		2,171		1,716	
Wald test	chi2 (20) =	129.67	chi2 (20) =	295.85	chi2 (20) =	60.79
Dependent Variable: Employment (logarithm)	Coef.	z-value	Coef.	z-value	Coef.	z-value
Employment (log)						
LD	0.483	8.08	0.478	7.42	0.330	4.44
L2D	0.040	1.83	0.053	1.48	0.020	0.83
Total sales (logarithm)						
D1	0.147	2.59	0.212	2.90	0.086	1.34
LD	0.032	0.86	0.037	0.66	-0.005	-0.16
Average wage (logarithm)						
D1	-0.023	-0.33	-0.144	-1.34	-0.029	-0.40
LD	0.032	0.75	0.122	1.66	-0.011	-0.24
Type 2 regions	0.011	2.19	0.001	0.10	0.019	2.43
Type 3 regions	0.014	2.11	0.007	0.73	0.023	2.19
Type 4 regions	0.011	1.15	0.001	0.07	0.024	1.95
Type 5 regions	0.006	0.95	0.001	0.09	0.009	0.91
Type 6 regions	0.007	1.25	-0.007	-1.09	0.030	3.50
Type 7 regions	0.009	1.57	0.000	-0.06	0.023	2.28
Type 8 regions	0.006	0.91	-0.004	-0.57	0.018	1.93
Type 9 regions	0.007	1.03	-0.002	-0.21	0.019	1.75
	5 year dummies in each estimation					
Constant	-0.022	-2.78	-0.024	-3.45	-0.025	-3.03
Sargan test of over-identifying restrictions: (twostep)	chi2(99) = 107.81 Prob > chi2 = 0.2562		chi2(99) = 113.72 Prob > chi2 = 0.148		chi2(99) = 88.51 Prob > chi2 = 0.766	
Arellano-Bond test that average autocovariance in residuals of order 1 is 0:	H0: No autocorrelation: z = -6.16 Pr > z = 0.0000		H0: No autocorrelation: z = -4.81 Pr > z = 0.0000		H0: No autocorrelation: z = -4.68 Pr > z = 0.0000	
Arellano-Bond test that average autocovariance in residuals of order 2 is 0:	H0: No autocorrelation: z = 0.58 Pr > z = 0.564		H0: No autocorrelation: z = -0.59 Pr > z = 0.558		H0: No autocorrelation: z = -1.06 Pr > z = 0.287	
Source: own calculates, LIAB, waves 1993-2002						

Despite substantial changes in the model specification, the results are remarkably stable. For the total sample all coefficients of the regional types are positive and those for the type 2, 3 and 7 regions are significant. Thus, employment of establishments develops better outside the most populated areas and is strongest in the second and third most aggregated areas. For the

service sector this pattern is even more prominent. The coefficients of all regional types are positive and most are significant.

**Table 5: A comprehensive dynamic model (one-step results)**

GMM estimates with heteroskedasticity robust standard errors						
	all establishments		manufacturing		services	
Number of obs =	10,771		3,640		4,779	
Number of groups =	3,897		1,397		1,716	
Wald test =	chi2(100)=	1353	chi 2(75) =	5180	Chi2(64) =	204.78
Dependent Variable: Employment (log)	Coef.	z-value	Coef.	z-value	Coef.	z-value
Employment (log)						
LD	0.502	8.21	0.317	4.81	0.352	4.75
L2D	0.037	1.62	-0.029	-1.08	0.026	1.08
Total sales (log)						
D1	0.177	2.81	0.083	1.56	0.091	1.54
LD	0.049	1.22	0.031	0.48	-0.003	-0.10
Average wage (log)						
D1	0.017	0.20	-0.127	-0.97	0.002	0.02
LD	0.021	0.47	0.044	0.78	-0.037	-0.82
Women's share of employment (log)						
D1	0.011	1.29	-0.014	-0.85	0.010	0.88
Share of part-time work (log)						
D1	0.025	7.36	-0.025	-2.56	0.026	5.96
Qualification level 1 (share in logs)						
D1	0.015	2.28	0.010	0.99	0.007	1.01
Qualification level 2 (share in logs)						
D1	0.023	4.86	0.014	1.94	0.016	2.62
Qualification level 3 (share in logs)						
D1	0.015	2.38	0.032	2.65	0.003	0.53
Qualification level 4 (share in logs)						
D1	0.006	1.14	0.000	-0.04	0.010	1.55
Qualification level 5 (share in logs)						
D1	0.015	2.28	0.009	1.14	0.013	2.15
Qualification level 6 (share in logs)						
D1	-0.003	-0.45	0.008	0.88	-0.001	-0.08

**Table 5 continued**

Type 2 regions	0.012	2.35	0.000	0.05	0.020	2.60
Type 3 regions	0.016	2.43	0.008	0.85	0.023	2.35
Type 4 regions	0.016	1.75	0.032	2.46	0.022	1.79
Type 5 regions	0.005	0.73	0.003	0.31	0.010	0.98
Type 6 regions	0.009	1.68	0.008	1.01	0.028	3.28
Type 7 regions	0.012	2.06	0.008	0.92	0.018	1.94
Type 8 regions	0.009	1.47	0.009	0.86	0.021	2.30
Type 9 regions	0.008	1.14	0.014	1.26	0.016	1.52
state of equipment / level 2	-0.008	-1.65			-0.009	-1.34
state of equipment / level 3	-0.009	-1.36	-0.015	-2.58		
state of equipment / level 4			-0.022	-3.22	-0.021	-2.21
state of equipment / level 5	0.002	0.10	-0.035	-1.63	0.008	0.42
	-0.069	-0.89			-0.046	-0.74
			0.340	3.56		
			5 year dummies in each estimation			
Industry dummies	68		43		24	
constant	-0.022	-2.03	-0.025	-2.04	0.001	0.06
Sargan test of over- identifying restric- tions: (twostep)	chi2(99) = 109.00 Prob > chi2 = 0.231		chi2(99) = 113.24 Prob > chi2 = 0.155		chi2(99) = 90.53 Prob > chi2 = 0.716	
Arellano-Bond test that average autocor- variance in residuals of order 1 is 0:	H0: no autocorrelation: z = -6.46 Pr > z = 0.000		H0: no autocorrelation: z = -3.15 Pr > z = 0.002		H0: no autocorrelation z = -5.34 Pr > z = 0.000	
Arellano-Bond test that average autocor- variance in residuals of order 2 is 0:	H0: no autocorrelation z = 0.61 Pr > z = 0.544		H0: no autocorrelation: z = -0.44 Pr > z = 0.660		H0: no autocorrelation z = 1.26 Pr > z = 0.209	
Source: own calculates, LIAB, waves 1993-2002						

Evaluating the test statistics our specifications are mainly supported. The Sargan test of over identification (calculated by a two-step estimation) does not reject the assumption of the exogeneity of the instruments. The Arellano-Bond tests of autocorrelation indicate that in all cases there is as assumed autocorrelation of the first but not of the second order.

## 6. Conclusion

In this paper we do research on the open question about regional agglomeration effects on labour demand at the establishment level. For this purpose we use the LIAB, a German linked employer-employee database of the Institute for Employment Research (IAB). Applying two different empirical approaches we find that establishments within agglomerated regions c.p. have a higher employment level. Thus the Krugman hypothesis of agglomeration effects and local demand is confirmed to some extent.

This is underlined by the fact that the effect is primarily driven by services, which are related to local and regional needs – at least to some degree. The inconclusive evidence for the



manufacturing sector might be explained by the global nature of the demand for most of the goods.

However, these findings reflect the state of the observation period which is the result of past developments. To gain answers about current developments we estimate a dynamic model. In this context, employment growth rates are smallest for establishments within large agglomerations. Establishments in less populated areas grow faster (or shrink slower). Thus, in accordance with other studies about the German labour market, we observe deglomeration and suburbanization processes. This is driven by the service sector, which is plausible because service sector establishments are easier to relocate than manufacturing plants. Due to the general growth of services there are more opportunities to start new enterprises for which new locational decisions are required.

There is no conflict between the findings obtained with the static and with the dynamic model. Assuming that the level of employment reflects past decisions, the agglomeration effects of our first empirical approach are results of location decisions made a long time ago, when transportation and communication costs were much higher than today. But due to path dependence these former decisions still form the economic structure of today.

However, current developments are more strongly influenced by the current environment. Thus, due to low transportation and communication costs the congestion effects of agglomerations outweigh their advantages. Employment is primarily growing in establishments in less crowded areas. This implies that policy measures focusing on metropolitan areas might not follow the most promising approach.

## **References**

Alda, Holger; Bender, Stefan; Gartner, Hermann (2005): The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB). In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 125, H. 2. pp. 327-336.

Arellano, Manuel; Bond, Stephen (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, v. 58, iss. 2, pp. 277-297.

Baltagi, Badi H. (2001): *Econometric Analysis of Panel Data* (Second edition), West Sussex: John Wiley & Sons.

Bellmann, Lutz (1997): The IAB establishment panel with an exemplary analysis of employment expectations. *IAB Labour Market Research Topics*, No. 20, pp. 1-14.

Blien, Uwe; Sanner, Helge (2006): *Structural Change and Regional Unemployment*, IAB-Discussionpaper.

Blien, Uwe; Suedekum, Jens (2005): *Local economic structure and industry development in Germany, 1993-2001*. IAB-Discussionpaper.

Blien, Uwe; Suedekum, Jens; Wolf, Katja (2005): Local Employment Growth in West Germany: A Dynamic Approach. Paper presented at the EALE/ SOLE Conference 2005 in San Francisco, forthcoming in: Labour Economics.

Combes, Pierre-Philippe; Magnac, Thierry; Robin, Jean-Marc (2004): The Dynamics of Local Employment in France. *Journal of Urban Economics* 56, pp. 217-243. reprinted: Discussions Paper Series of Centre for Economic Policy Research No. 3912 (2004).

Glaeser, Edward L. (1992): Growth in Cities. *Journal of Political Economy*, v. 100, iss. 6, pp. 1126-1152.

Görmär, Wilfried; Irmen, Eleonore (1991): Nichtadministrative Gebietsgliederungen und -kategorien für die Regionalstatistik. Die siedlungsstrukturelle Gebietstypisierung der BfLR, In: *Raumforschung und Raumordnung*, 49/6, pp. 387-394.

Hamermesh, Daniel S. (1986): The Demand for Labor in the Long Run. In: O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, North-Holland Press, pp. 429-471.

Hamermesh, Daniel S. (1993): *Labor Demand*. Princeton and Chichester, U.K.: Princeton University Press.

Henderson, Vernon; Kuncoro, Ari; Turner, Matt (1995): Industrial Development in Cities. *Journal of Political Economy*, v. 103, iss. 5, pp. 1067-1090.

Kölling, Arnd (2000): The IAB-Establishment Panel. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*, Jg. 120, H. 2., pp. 291-300.

Krugman, Paul (1991): *Geography and Trade*, Cambridge (Mass.) etc.: MIT Press.

McCann (2005): *Industrial Clusters*. Paper presented at the Conference on Regional Development, University of Barcelona.

Möller, Joachim; Tassinopoulos, Alexandros (2000): Zunehmende Spezialisierung oder Strukturkonvergenz? Eine Analyse der sektoralen Beschäftigungsentwicklung auf regionaler Ebene. *Jahrbuch für Regionalwissenschaft*, Jg. 20, 1. pp. 1-38.

Moulton, Brent R. (1990): "An illustration of a pitfall in estimating the effects of aggregate variables on micro units", in: *The Review of Economics and Statistics*, 72: 334-338

Nickell, Stephen J. (1986): Dynamic Modells of Labour Demand. In: O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, North-Holland Press, pp. 473-522.

Nickell, Stephen; Wadhvani, Sushil; Wall, Martin (1992): Productivity Growth in U.K. Companies, 1975-1986. *European Economic Review*, v. 36, iss. 5, pp. 1055-1085.

Suedekum, Jens; Blien, Uwe; Ludsteck, Johannes (2006): What has caused regional employment growth differences in East Germany. *Jahrbuch für Regionalwissenschaft* 1.