Paper to be presented at the

35th DRUID Celebration Conference 2013, Barcelona, Spain, June 17-19

How important are agglomeration effects for plant performance? Empirical evidence for Germany

Michaela Fuchs
Institute for Employment Research
Regional Research Network
michaela.fuchs@iab.de

Oliver Ludewig
Institute for Employment Research
Regional Research Network
oliver.ludewig@iab.de

Abstract

The question whether agglomeration effects are of importance for plant performance has a long tradition in regional science. This paper asks if agglomeration effects influence the performance of plants in Germany and, if so, in which direction. We aim at contributing to the still sparse empirical studies in this field of research by adding three aspects to the existing evidence. First, while earlier papers looked only at few sectors of the economy or only at manufacturing, we extend our analysis to both the manufacturing and the services sector. Second, we contribute to the still sparse literature that identifies agglomeration effects while controlling for the internal structure of the establishments. Third, we provide novel plant-level evidence on regional agglomeration effects for Germany To this we incorporate a rich set of plant characteristics that are likely to influence productivity. We estimate plant-level production functions augmented by regional characteristics and controlling in detail for plant-specific features. The analysis is conducted both within a static and a dynamic panel framework. We use the IAB Establishment Panel, a large-scale German establishment survey covering around 16,000 establishments each year. The results emphasize the strong impact of location on plant performance both in the manufacturing and in the services sector, supporting the idea that plant productivity might indeed be enhanced by agglomeration effects at force in dense areas. Furthermore, a learning effect in the sense that plants become more productive by being located in dense regions becomes visible.

Jelcodes:R30,L25

How important are agglomeration effects for plant performance? Empirical evidence for Germany

Abstract

The question whether agglomeration effects are of importance for plant performance has a long tradition in regional science. This paper asks if agglomeration effects influence the performance of plants in Germany and, if so, in which direction. We aim at contributing to the still sparse empirical studies in this field of research by adding three aspects to the existing evidence. First, while earlier papers looked only at few sectors of the economy or only at manufacturing, we extend our analysis to both the manufacturing and the services sector. Second, we contribute to the still sparse literature that identifies agglomeration effects while controlling for the internal structure of the establishments. Third, we provide novel plant-level evidence on regional agglomeration effects for Germany To this we incorporate a rich set of plant characteristics that are likely to influence productivity. We estimate plant-level production functions augmented by regional characteristics and controlling in detail for plant-specific features. The analysis is conducted both within a static and a dynamic panel framework. We use the IAB Establishment Panel, a large-scale German establishment survey covering around 16,000 establishments each year. The results emphasize the strong impact of location on plant performance both in the manufacturing and in the services sector, supporting the idea that plant productivity might indeed be enhanced by agglomeration effects at force in dense areas. Furthermore, a learning effect in the sense that plants become more productive by being located in dense regions becomes visible.

Keywords: firms, productivity, externalities, agglomerations, panel data

JEL classifications: R30, R12, L25

preliminary version – please do not quote

1. Introduction

Does it matter for plant performance if the plant is located in an agglomeration or in a rural area? According to theoretical considerations from regional science dating back to Alfred Marshall (1890), the answer is clearly yes. A major argument is that in large and dense regions, agglomeration effects come into force that are associated with region-specific external effects and increase the productivity of plants located close to each other. These agglomeration effects provide an explanation why economic activities are distributed unevenly in space and why the geographic concentration of firms and people remains persistent over time.

A large body of empirical literature has been devoted to quantify the relationship between agglomeration effects and productivity (for an overview, see Rosenthal/Strange 2004). However, most of the existing papers look at regional aggregates (e.g. Glaeser et al. 1992, Henderson et al. 1995), while the theoretical arguments rest on the establishment level. Only since Henderson's (2003) seminal paper there is an increasing effort to look into these relationships at the micro level that is at the level of firms, establishments or plants. Most of the literature, however, focuses on the manufacturing sector or even sub-sectors (e.g. Henderson 2003, Saito/Gopinath 2009, Anderson/Lööf 2009, Baldwin et al. 2010), the Anglo-Saxon countries (e.g. Henderson 2003, Maré/Graham 2010) and use administrative data (e.g. Maré/Graham 2010, Henderson 2003, Saito/Gopinath 2009, Baldwin et al. 2010, Cingano/Schivardi 2004), which allows little insight in the organization of the plant. Summarizing this literature there is strong focus on the manufacturing sector and the Anglo-Saxon countries. Evidence for the whole range of sectors is rare and other countries than the Anglo-Saxon ones are seldom investigated. Specifically to our knowledge there is no study analyzing agglomeration effects on the establishment level for Germany. The only recent paper looking at productivity in German regions uses regionally aggregated data (Eckey et al. 2010).

In addition to these findings, there is a huge body of literature showing the importance of the internal structure of an establishment, e.g. industrial relations, part-time work or the owner structure, for its productivity (e.g. Ichniowsky et al. 1997, Black/Lynch 2001, Ludewig/Sadowski 2009). However, these elements are largely missing in studies looking at agglomeration effects on plant level.

The present paper contributes to the question of agglomerations influence productivity by adding three aspects to the existing evidence. First, while earlier papers looked only at few sectors of the economy or only at manufacturing, we also include the services sector. Second, we contribute to the still sparse empirical literature that identifies agglomeration effects while controlling the internal structure of the establishment by using a rich set of plant characteristics that are likely to influence productivity. Additionally, we use for the production factor labor the number of employees in full-time equivalents. In taking explicitly into account the growing importance of part-time work, this gives a more precise account of the labor input. Third, we are among the first in providing plant-level evidence on regional agglomeration effects and productivity for Germany.

In order to quantify the impact of geography on plant performance we estimate plant-level production functions augmented by regional characteristics and controlling in detail for plant-specific features. The analysis is conducted both in a static and a dynamic panel framework. The static model answers the question if the output level of a plant depends on the regional economic structure, controlling for plant characteristics. The dynamic model accounts for endogeneity problems that can evolve in the static models and focuses on the temporal persistence of a plant's performance subject to the surrounding geographic features. A significant effect in the dynamic models can also point towards a learning effect, meaning that plants become more productive when located in large agglomerations.

Our analysis is based on the IAB Establishment Panel. It is an annual representative employer survey at individual establishments in Germany covering a broach range of establishment characteristics. We use the years from 2001 to 2006 with roughly 20,000 observations for almost 4,000 establishments. The regional variables are are added on the NUTS3-level.

The paper is structured as follows. Section 2 discusses the theoretical background and the econometric setup. In section 3, the data and the variables are described. Section 4 is dedicated to the results, and section 5 concludes.

2. Theoretical background and econometric setup

2.1 Agglomeration economies and productivity

The 'lumpiness of space' that discloses the large concentration of economic activities in just few regions can by now seen as a stylized fact (see, e.g., Krugman 1991). Explanations of why economic agents benefit from being close to each other reach back as far as to Alfred Marshall (1890). They are often called the three Marshallian forces (for an overview, see Rosenthal/Strange 2004 and Puga 2010). First, labor pooling allows better matches, reduces search costs, training costs and risks for both employers and employees. Second, knowledge spillovers induce product and process innovations and thus increase productivity. Spillovers especially of tacit knowledge are the more likely the more employees of the same industry interact with each other and switch employers (Jaffe et al. 1993). The third effect stems from forward and backward linkages. The common usage of an input or a supplier can for example free up economies of scale, of which all establishments in that specific region and industry benefit. All three forces can unfold themselves much better in larger markets with a large amount of economic agents than in rural areas.

Apart from these positive agglomeration effects there are also negative agglomeration effects working in the opposite direction. For example, in densely populated areas the overcrowding of places has unfavourable consequences: increasing land prices, traffic problems, environmental pollution, and an overstrained infrastructure etc. increase the cost of production and thus reduce productivity (Eckey et al. 2010).

In the course of time many empirical studies have been devoted to empirically confirm the role of agglomeration effects in general and specifically the three Marshallian forces. However, a huge part of them is based on regional aggregates (e.g. Glaeser et al. 1992, Henderson et al. 1995, Fuchs 2011). While such studies find evidence for positive agglomeration effects on aggregate measures like regional employment, they look at the wrong level of analysis, because agglomeration effects are expected to increase the productivity of plants located close to each other. (Rosenthal/Strange 2004, Cingano/Schivardi 2004). By definition, local external scale economies imply that plants are able to produce more output with the same inputs in larger, denser, urban environments (Puga 2010, 207). They shift the production function of establishments (Rosenthal/Strange 2004) and impact on the regional level only indirectly.

A still small but rising number of papers that try to identify agglomeration effects on the plant level has been published recently. Henderson (2003) finds in his paper for four high-tech and five machinery goods industries in the USA that there are substantial localization economies in the high-tech industries. Urbanization economies seem to dominate the machinery good industries. There is only weak evidence for dynamic effects. Moretti (2004) shows that productivity is higher in regions with higher shares of college graduates. Andersson/Löff (2009) identify substantial agglomeration effects for the Swedish manufacturing sector but they do not differentiate between localization and urbanization effects. Baldwin et al. (2010) find strong evidence for localization effects in the Canadian manufacturing sector. Cingano/Schivardi's (2004) results indicate for an Italian firm panel that there are substantial localization effects. Maré/Graham (2010) analyze agglomeration effects in New Zealand on the one-digit industry level using a huge firm level panel data set. Using employment density, they find substantial agglomeration elasticities, however they do not differentiate between localization and urbanization effects. In their study on Great Britain, Harris/Moffat (2012) control for both plant and firm effects in an attempt to provide more robust evidence on whether productivity is higher in cities.

For Germany there is to our knowledge only one study that analyzes the relationship between regional size and firm performance with firm-level data and conditioned on attributes of firms. Ehrl (2011) investigates the relative impact of alternative microeconomic agglomeration sources on firm's total factor productivity using German establishment and employment level data. However, he focuses on the various estimation techniques of total factor productivity from firm level production functions by using data from 2000 to 2005.

This paper attempts to provide a general connection between productivity and agglomeration economies. To this end we estimate in the following sections plant-level production functions including a general measure of the density of a region, while controlling for a rich set of plant characteristics

2.2 Econometric setup

There are several approaches to estimating productivity with micro-level panel data (see van Beveren 2012 for a survey). The starting point of our empirical model is a Cobb-Douglas establishment-level production function for plant *i* with the stock of capital and labor included as inputs (see also Cingano/Schivardi 2004). Henderson (2003) and Moretti (2004) report little differences in parameter estimation between Cobb-Douglas or more flexible forms of production functions. Our basic model is

$$Y_{it} = AK_{it}^{\alpha} L_{it}^{\beta}, \tag{1}$$

with the subscript i=1, 2,…, N referring to plant *i* and t=1, 2,…, T to year *t*. Y_{it} denotes value added, *A* the technology that is available for all plants, K_{it} is plant i's capital stock, and L_{it} expresses the amount of labor employed in plant i.

Other characteristics of the plant and the its location are assumed to be transmitted through the productivity term *A*. Hence, agglomeration phenomena 'shift' the production function of the establishments (see Rosenthal/Strange 2004). Similar modeling frameworks are applied by Henderson (2003), Moretti (2004), Andersson/Lööf (2011) and Baldwin et al (2010). In addition, assume that productivity depends on various plant, region and time components:

$$lnA = \gamma + \varphi lnX_{it} + \vartheta lnG_{rt} + \lambda_r + \eta_i + \delta_t + \varepsilon_{irt}, \quad (2)$$

where X_{it} is a vector of characteristics related to plant i, and G_{rt} is a vector of characteristics that are associated with region r. λ_r stands for fixed effects at the level of the region, and η_i represents the influence of plant-specific fixed effects. The inclusion is necessary because empirical evidence using plant- and firm-level panel data consistently shows that plants are heterogeneous and that their distribution is persistent over time (Harris/Moffat 2012, 772). General time-specific influences that are generated for example from business cycles are captured by δ_t., and ε_{irt} is an idiosyncratic error term. Substituting equation (2) into equation (1) in natural logarithms, we obtain our full model:

$$lnY_{it} = \gamma + \varphi lnX_{it} + \vartheta lnG_{rt} + \alpha lnK_{it} + \beta lnL_{it} + \lambda_r + \eta_i + \delta_t + \varepsilon_{irt} \ . \quad (3)$$

Equation (3) is our basic estimation equation that explains plant output by regional and plant characteristics in addition to the two factors of production. In containing variables for both the plant and the region, equation (3) combines information on two different levels of observation, with some of them not varying between plants or regions. This multilevel structure can result in inefficient estimates of the coefficients and in biased estimates of the standard errors especially of the variables for the higher level (Moulton 1990). In order to deal with this problem, in the static regressions clustering-robust linear regression techniques are used to estimate standard errors that recognize this clustering of the data. This method relaxes the independence assumption and requires only that the plant-level observations be independent across regions. By allowing any given amount of correlation within regions, clustering-robust techniques estimate appropriate standard errors when many observations share the same value on some but not all independent variables.

For estimating equation (3), we focus on methods for a short panel with data on many plants but few time periods. In order to make robustness checks for the basic relationship, we first run pooled OLS regressions under the assumption that the errors are independent over plants. We then move on to fixed effects (FE) estimation models. One of the main benefits of the FE estimator is its ability to control for time-invariant unobserved heterogeneity, which is important in our case. Hence, we impose time-independent effects for each plant that are possibly correlated with the regressor.

In the second step of our econometric analysis, we estimate equation (3) in a dynamic framework. Because of the large persistence in plant productivity (Blundell and Bond 2000), this possible source of endogenenity problems should be controlled for. Furthermore, the inclusion of a lagged dependent variable $Y_{i,t-m}$ dating back over an a priori unknown period of time (m=1,…,M) allows for the explicit modeling of adjustment processes and the influence of history on plant output. The dynamic model then becomes

$$lnY_{it} = \gamma + \varphi lnX_{it} + \sum_{m=1}^{M} \delta_m Y_{i,t-m} + \vartheta lnG_{rt} + \alpha lnK_{it} + \beta lnL_{it} + \lambda_{rt} + \eta_{it} + \ln \varepsilon_{it} \quad (4)$$

In dynamic panel models such as (4), OLS estimates are biased and inconsistent if the lagged dependent variable is correlated with the error term, as it is frequent in dynamic panels with a short time dimension (Nickell 1981). We therefore estimate equation (4) with a generalized methods of moments (GMM) estimator and use the two-step system GMM estimator proposed by Arellano/Bover (1995) and Blundell/Bond (1998) available in STATA 12. It allows for fixed effects and takes into account endogeneity of the right-hand side variables and selection bias by using their lagged values, both in differences and in levels, as instruments. In comparison to the difference GMM estimation technique of Arellano/Bond (1991) it allows the introduction of more instruments, which improves efficiency.[1] One further advantage that is important for our analysis is that, in contrast to the difference GMM estimator, it can include time-invariant regressors.

3. Data and variables

3.1 Data

For information on the individual plants we resort to the IAB Establishment Panel, an annual representative employer survey at individual establishments in Germany with at least one employee liable to social security (for details see Fischer et al. 2009). Approximately 16,000 establishments from all sectors of the economy and of all sizes are questioned on a large number of employment-related subjects, including employment development, business policy and development, innovations, wages and salaries, working times and general data on the establishment. The Establishment Panel was started in Western Germany in 1993 and in Eastern Germany in 1996. As a comprehensive longitudinal data set, it forms the basis for detailed research into the determinants of plant performance.[2] Details on the employees are obtained from the Linked Employer-Employee Data of the IAB, which matches the IAB Establishment Panel data with individual data from the IAB (for details seeAlda/Bender/Gartner 2005).

Data on the regional characteristics come from the IAB Establishment History Panel. It contains information about the branch of industry, the location and the employees of each establishment in Germany with at least one employee liable to social security or at least one marginal part-time employee

[1] Since the Arellano/Bond estimator is based only on an equation in differences, it is also called difference GMM estimator as opposed to the system GMM estimator (see, for example, Roodman 2009).

[2] English versions of the questionnaire can be downloaded under http://fdz.iab.de/en/FDZ_Establishment_Data/IAB_Establishment_Panel/IAB_Establishment_Panel_Working_Tools.aspx

on the reference date. The variable characterizing the plant's environment are calculated at the NUTS-3 level that comprises 439 *Kreise* and *kreisfreie Städte*.

We consider only those sectors that are subject to market-based forces, i.e. where our assumptions of a cost-minimizing firm are appropriate. Hence, plants that belong to the public sector are excluded. In addition, those plants are excluded that express their business volume by the budget volume (administration and property budget), i. e. publicly owned establishments that belong to market-oriented sectors. Sectors that are strongly dependent on geographical features (agriculture, fishing and mining) are not considered either. Finally, we exclude establishments with missing values for our relevant variables. The period of observation covers the years from 1996 to 2008. Our final panel data set comprises a total of 58,114 observations on 16,334 plants.

3.2 Variables

For the estimation of the basic production function (2), information on output, capital, and labour is required. We measure the dependent variable of plant output with its nominal value added in the respective year. Regarding the input of capital, the IAB Establishment Panel provides no direct information. As Müller (2008, 358) notes, this is a general problem of many establishment data sets. Measurement errors in capital stock, however, will lead to biased estimates and any inference based on such estimates could be misleading. In order to get a reliable measure of the capital stock, we calculate it according to a modified perpetual inventory approach specifically developed for this purpose by Müller (2008). The input of labor is denoted by the number of employees in an establishment, measured in full-time equivalents. This way, we take into account the growing importance of part-time work in Germany over the past years.

Our focus in this paper is not to single out through which channel agglomeration effects impact on plant productivity. We rather remain on a more general level of analysis and estimate the general impact of regional size on plant productivity. We hereby assume that the general features of large markets with a greater demand, where the various agglomeration effects come into force, make it possible for plants to adopt more efficient production processes. For measuring regional size we refer to a density measure and use the number of employees per km^2 (see also Puga 2010).[3] The motivation behind is that the effectiveness of agglomeration effects depends crucially on the number of people that not only live in a region, but that also participate actively in economic life.

In addition to the region-specific characteristics, we control for several plant-specific variables that are expected to influence plant performance. First of all, engaging in exports is an important opportunity for a plant to expand the market for its products. At the same time, it becomes exposed to negative

3 We have experimented with various measures of regional size, including only the number of employees, the number of plants, GDP and population per km^2. However, all measures are highly correlated with each other, with the number of employees per km^2 resulting to be the most precise measure.

demand shocks emanating from the export partners, so that a priori there is no clear-cut relationship. Exports also imply a higher degree of competition, because exporters are selling in additional markets with additional competitors. Competition puts pressure on the establishment to be more efficient and thus exporters should be more productive (see Wagner 2007). We use the share of sales achieved abroad on total sales in the last fiscal year as indicator of export orientation.

Plant productivity can also be indirectly influenced by the existence of a works council that has a strong impact on many decisions of the plant and the organisational structure (see, e.g., Blanchflower et al 1991). Empirical results stress a positive correlation between productivity and works councils (e.g. Hübler/Jirjahn 2003). The existence of a works council is captured by a dummy variable that takes the value of one if the establishment has a works or staff council or some other company-specific form of staff representation (staff spokesperson, round table conferences).

The influence of external enterprises can generate both positive and negative productivity effects (Bates 1995). On the one hand, the provision of technological or enterpreneurial know-how or networks with customers or suppliers can stimulate positive productivity effects. On the other hand, the influence becomes negative if an establishment is strongly influenced by strategies and decisions of the external partner. The IAB-Establishment Panel asks if the establishment surveyed is a) an independent company or an independent organization without other places of business, b) the head office of an enterprise or an organization with other places of business/offices/branches, or c) a place of business/office/branch of a larger enterprise or organization.[4] The answers are captured with the help of dummy variables. We also include a dummy variable if the enterprise is mainly or exclusively in foreign property. Because foreign-owned enterprises potentially have access to a broader range of experiences and technologies, they are expected to have a higher level of productivity (see also Henderson 2003 and Baldwin et al. 2008).

Plant performance also strongly depends on plant age. One can argue that, over time, plants age and slowly lose their ability to compete. On the other hand, plants can also improve with age due to learning effects (for an overview see Coad et al 2010). Hence, a priori the influence of age on plant performance is important, but if it is positive or negative needs to be answered empirically. We measure plant age with a dummy variable that takes on the value of 1 if the plant is young, i.e. if it was founded after 2001.

The innovation behaviour of a plant can also be interrelated with its productivity. The technical state of the plant's equipment is included as a rough proxy for innovations and market leadership. A plant that engages in innovative activities can be assumed to have a rather modern and up-to-date technical

[4] It is also asked if it is a middle-level authority of a multi-level company or a multi-level authority/organization. But since this includes basically the public sector, these enterprises are a priori excluded from the analysis.

state. The variable is measured with a scale ranging from 1 (very modern state) to 5 (outdated). Last, productivity is expected to vary significantly between sectors. We control for sector-specific effects by adding sectoral dummies at the 2-digit level of the WZ 2003 (corresponding to the NACE Rev.1 classification). Because of systematic differences between East and West Germany, we also include a dummy variable for East Germany. Table 1 gives an overview on all variables under consideration. Summary statistics are provided in table A2.

Table 1: Variable descriptions

Variable	Abbreviation	Description
		Dependent variable
Output	*y*	Value added (in Mill Euro)
		Basic production function variables
Labor	*labor*	Number of employees (in full-time equivalents)
Capital	*capital*	Capital stock (in Mill Euro)
		Region-specific variable
Regional density	*density*	Number of employees per km^2
		Plant-specific control variables
Export activities	*exports*	Share of sales abroad on total sales in percent
Works council	*counc*	Dummy: 1= works council
Dependency structure	*struct1*	Dummy: 1= independent
	struct2	Dummy: 1= branch
	struct3	Dummy: 1= head office
Property structure	*fprop*	dummy: 1= foreign property
Plant age	*age*	Number of years plant has been in operation based on year of entry
Technical state	*tech*	1 (state of the art) to 5 (obsolete)
Sector	*sector*	Dummies for NACE=15 to 74, respectively
East-West	*east*	Dummy: 1= Eastern Germany

4. Results

4.1 Static models

In this section, we concentrate on the results of the static specifications and estimate three models in total. In order to quantify the basic relationship between the two factors of production and plant output, we first estimate the influence of K_{it} and L_{it} only on Y_{it} (Baseline in table 1). Then, we add our measure of regional density (Baseline plus regional characteristics). Finally, in order to check for the consistency of the regional variable, we move to the full model that includes the plant variables (Baseline plus regional and plant characteristics):

$$lnY_{it} = \gamma + \varphi lnX_{it} + \vartheta lnG_{rt} + \alpha lnK_{it} + \beta lnL_{it} + \lambda_r + \eta_i + \delta_t + \varepsilon_{irt}\ . \quad (3)$$

Each model is estimated with OLS and with FE estimation techniques, resulting in six regressions. Table 2 presents the results of the static models.

Table 2: Production function estimates - static models

	Baseline		Baseline plus regional characteristics		Baseline plus regional and plant characteristics	
	OLS	FE	OLS	FE	OLS	FE
log labor	0.795 *** (114.34)	0.379 *** (14.22)	0.786 *** (67.10)	0.378 *** (14.24)	0.734 *** (57.83)	0.368 *** (12.02)
log capital	0.217 *** (40.27)	0.129 *** (5.54)	0.219 *** (23.18)	0.129 *** (5.54)	0.194 *** (21.47)	0.128 *** (4.88)
log density			0.232 *** (8.06)	-0.143 (-0.75)	0.211 *** (7.96)	-0.151 (-0.62)
exports					0.004 *** (7.87)	0.014 ** (2.04)
counc					0.237 *** (9.18)	0.042 (1.47)
struct2					0.252 *** (8.60)	-0.061 ** (-2.32)
struct3					0.174 *** (6.07)	-0.043 (-1.19)
fprop					0.064 * (1.82)	-0.015 (-0.32)
age					0.000 (0.16)	0.009 *** (2.71)
tech					-0.078 *** (-8.10)	-0.010 (-1.13)
east	-0.306 *** (-31.33)		-0.293 *** (-10.99)		-0.254 *** (-8.97)	
constant	8.451 *** (127.95)	11.432 *** (27.08)	8.394 *** (73.75)	11.473 *** (26.86)	8.904 *** (78.26)	11.734 *** (26.02)
N	29,389	22,908	29,389	22,908	25,817	19,827
R^2	0.86	0.84	0.86	0.83	0.86	0.80
R^2 within		0.07		0.07		0.07
R^2 between		0.85		0.83		0.81

Robust standard errors adjusted for clustering. Significance at the *p<0.10, **p<0.05, ***p<0.001 level. t-statistics in parentheses. Time and sector dummies included but not reported. The coefficients for log emp_dens are multiplied with 1,000.

The baseline results confirm the positive and highly significant influence of labor and capital on plant output. The OLS and FE coefficients mainly retain their size even after including the regional and plant variables, underlining the robustness of the underlying fundamental relationships. The number of employees per km^2 that acts as our indicator of regional density is positive and highly significant in the OLS regressions. This result supports our hypothesis that plant productivity is indeed a positive function of regional density and hence regional size. The FE regressions, however, show no statistical relationship between regional density and plant productivity. This might be due to the low time variance of the regional variable. Since the FE estimator uses only the within variance and forgoes the between variance (Baltagi 2001), this can lead to inefficient coefficient estimates. Similar results are also reported by Andersson and Lööf 2011. Since our preferred model is the System GMM case, we do not give too much emphasis on the FE results.

The OLS results from the full model clearly show that plants located in denser regions have a higher productivity, all else equal. The coefficient estimate of employment density is only slightly lower than in the model without plant characteristics. Most of the plant characteristics are positive and statistically highly significant in explaining plant productivity. Export activities and the level of output are strongly intertwined, as is the case for a works council. In comparison to independent plants, output is higher in branches or in head offices. Plant age does not seem to exert an influence, whereas output rises with the overall technical state of the plant. Finally, the location of a plant in the Eastern part of Germany significantly reduces productivity.

Results for the separate regressions on manufacturing and services are denoted in table A3 in the appendix. Since we want to focus on the impact of regional density under control of all variables, only estimates of the full models (Baseline plus regional and plant characteristics) are reported. Location clearly matters for both manufacturing and services plants, and the coefficient is of roughly the same size for both sectors. Again, the FE results are not significant for the density variable.

4.2 Dynamic model

This section presents results of the dynamic equation (4), which is our preferred model. Just as in the case of the static models, we first estimate the Baseline model and successively add the regional and the plant characteristics. The models are estimated with the system GMM estimator proposed by Arellano/Bover (1995) and Blundell/Bond (1998), and we report heteroskedasticity-robust standard errors calculated according to the mechanism by Windmeijer (2005). After trying out several specifications for the number of lags included, our optimal specification of equation (4) includes only one lag of the dependent variable:

$$lnY_{it} = \gamma + \varphi lnX_{it} + \delta lnY_{t-1} + \vartheta lnG_{rt} + \alpha lnK_{it} + \beta lnL_{it} + \lambda_{rt} + \eta_{it} + \ln \varepsilon_{it} \qquad (5)$$

Before turning to the results, the validity of the system GMM estimator should be checked. A comparison of the regression results for the lagged endogenous variable with the system GMM estimator on the one hand and the OLS and FE estimator on the other hand can serve as an indicator of its validity (Bond 2002, Roodman 2009). Bond (2002) notes that the OLS estimation results for equation (5) are biased upwards, whereas the FE estimation results are biased downwards. Accordingly, consistent GMM results should lie between those of the two former estimators.

Table A2 presents the results for the OLS, the FE, the difference GMM and the system GMM estimation methods for the Baseline model with the inclusion of the lagged dependent variable. It turns out that the system GMM estimator yields a value for the coefficient on y_{t-1} that lies between those of the OLS and the FE estimator. Hence, the OLS estimator gives an upper bound and the FE estimator a lower bound for the coefficient estimated with the system GMM technique. As a consequence, the system GMM estimator should generate consistent and unbiased results.

Table 3 summarizes the regression results of the dynamic models describing the dynamic relationship between productivity and location. In all three models, the estimates on lagged output highlight the persistency in plant performance. This implies that lagged output is a good predictor of contemporaneous output. Turning to our location indicator it becomes evident that, in contrast to the FE results, the coefficients are positive and highly significant. In quantitative terms they have about the same magnitude as in the OLS regressions. Since the dynamic framework controls for previous output levels, a causative interpretation of the results can be undertaken. According to our results, contemporaneous output is higher for plants located in regions with a higher density of employees. They thus suggest that density not only is positively related with plant output, but also propose that density contributes to higher output. This result can be interpreted in the sense that plant productivity might indeed be enhanced by agglomeration effects at force in dense areas.

Table 3: Production function estimates - dynamic models

	Baseline		Baseline plus regional characteristics		Baseline plus plant characteristics	
log output $_{t-1}$	0.317 (3.28)	***	0.308 (11.79)	***	0.316 (11.30)	***
log labor	0.410 (2.11)	**	0.424 (6.66)	***	0.477 (6.76)	***
log capital	0.168 (2.44)	**	0.168 (4.32)	***	0.151 (3.14)	***
log density			0.231 (5.25)	***	0.217 (3.83)	***
export					0.004 (1.79)	*
counc					0.032 (1.05)	
struct2					0.014 (0.42)	
struct3					0.038 (1.38)	
fprop					-0.022 (-0.44)	
age					0.003 (0.23)	
tech					-0.002 (-0.25)	
East	-0.327 (-1.34)		-0.216 (-3.37)	***	-0.200 (-1.53)	
constant	6.629 (6.48)	***	6.537 (10.93)	***	6.591 (8.05)	***
N	27,139		27,139		24,094	
# instruments	233		309		379	
Sargan (p-value)	0.004		0.007		0.012	
AR(1) (p-value)	0.000		0.000		0.000	
AR(2) (p-value)	0.159		0.250		0.119	

Windmeijer-corrected standard errors. Significance at the *p<0.10, **p<0.05, ***p<0.001 level. z-statistics in parentheses. Time and sector dummies included but not reported. The coefficients for log density are multiplied with 1,000. Labor, capital, and exports are treated as endogenous, density as predetermined.

Separate System GMM results for manufacturing and services are reported in table A3 in the appendix. The positive and significant coefficients on lagged output emphasize the persistency in plant performance also separately for the two sectors. Clearly, the density of the number of employees plays an important role for both manufacturing and services, albeit the estimate for the services sector is only significant at the 5% level.

Tables 3 and A3 include the Sargan test statistic and the Arellano-Bond test for autocorrelation in the first-differenced errors. The p-values of the Sargan test statistics in table 3 display rather low values, rejecting the null hypothesis that the instruments are valid. This is probably due to the large number of instruments, which can cause the Sargan test to overreject (see Hansen/Heaton/Yaron, 1996). Importantly, however, the conditions for the absence of second-order autocorrelation in the error terms are fulfilled. Hence, the main requirement for the moment conditions to hold is met, and the instruments used in the estimations can be regarded as valid.

5. Conclusions

In this paper we investigate the importance of agglomeration effects for the performance of plants while controlling for various plant-specific determinants. To this end we augment a Cobb-Douglas production function with labor and the capital stock as basic inputs by characteristics specific to the regional environment the plant is located in as well by plant-specific characteristics that are likely to influence plant productivity. Since we are interested in the general relationship between the performance of a plant and its location, we use employment density as a general indicator of regional size and potential of agglomeration effects. Both static and dynamic models are estimated, with the dynamic models explicitly considering endogeneity between agglomeration effects and plant productivity. The paper further adds to the still sparse literature on agglomeration and productivity at the micro level by investigating the manufacturing as well as the services sector while controlling for a rich set of plant-level characteristics.

The results emphasize the strong impact of location on plant performance both in the manufacturing and in the services sector, supporting the idea that plant productivity might indeed be enhanced by agglomeration effects at force in dense areas. Plants located in regions with a high density of employees are more productive than plants in less dense regions, even after controlling for export activities, works councils, the dependency and property structure, plant age, technical state, and time and sector trends. Furthermore, results from the System GMM estimations allow a causal interpretation. Accounting for previous performance levels, contemporaneous plant productivity results to be higher in denser regions. This hints towards a learning effect in the sense that plants become more productive by being located in dense regions.

In order to gain more and refined insights into the relation between plant productivity and location factors further research is clearly warranted. So far, we have used a rather rough indicator of the mechanisms that are supposed to work in agglomerations. While the results support a strong interrelation between the employment density of a region and plant productivity, we have no information on the exact channels through which the underlying agglomeration effects come into force. One promising way to tackle this question would be to develop refined measures suitable for quantifying effects that operate through the three different channels Marshall (1890) proposed.

References

Alda, H., S. Bender and H. Gartner (2005): The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB). Schmollers Jahrbuch 125, 327-336.

Andersson, M. and H. Lööf (2011): Agglomeration and productivity: evidence from firm-level data. Annals of Regional Science 46, 601-620.

Arellano, M. and S. Bond (1991): Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 227-297.

Arellano, M. and O. Bover (1995). Another Look at Instrumental Variable Estimation of Error Component Models. Journal of Econometrics 68, 29-51.

Baldwin, J., M. Brown and D. Rigby (2010): Agglomeration economies: microdata panel estimates from Canadian manufacturing. Journal of Regional Science 50, 915-934.

Baltagi, B. (2001): Econometric analysis of panel data. Chichester, UK, Wiley and Sons.

Bartelsman, E. and M. Doms (2000): Understanding productivity: lessons from longitudinal microdata. Journal of Economic Literature 38, 569-594.

Bates, T. (1995): A comparison of franchise and independent small business survival rates. Small Business Economics 7, 377-388.

Black, S. and L. Lynch (2001): How to compete: the impact of workplace practices and information on technology and productivity. Review of Economics and Statistics 83, 434-445.

Blanchflower, D., N. Milward and A. Oswald (1991): Unionism and employment behavior. The Economic Journal 101, 815-834.

Blundell, R., S. Bond (2000): GMM estimation with persistent panel data: an application to production functions. Econometric Reviews 19, 321-340.

Blundell, R. and S. Bond (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. Journal of Econometrics 87, 115-43.

Bond, S. (2002): Dynamic panel data models: A guide to micro data methods and practice. CWP 09/2002.

Cingano, F. and F. Schivardi (2004): Identifying the sources of local productivity growth. Journal of the European Economic Association 2, 720-742.

Coad, A., A. Segarra and M. Teruel (2010): Like milk or wine: does firm performance improve with age? Papers on Economics and Evolution 1006, MPI Jena.

Combes, P. (2000): Economic structure and local growth: France 1984-1993. Journal of Urban Economics 47, 329-255.

Combes, P.-P., T. Magnac and J.-M. Robin (2004): The dynamics of local employment in France. Journal of Urban Economics 56, 217-243.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga and S. Roux (2012): The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. IZA Discussion Paper No. 6502.

Eckey, H.-F., R. Kosfeld and N. Muraro (2009): Effects of population development on the economic level of the regions in Germany. Jahrbuch für Wirtschaftswissenschaften 60, 204-226.

Ehrl, P. (2011): Agglomeration economies with consistent productivity estimates. BGPE discussion paper 113, Erlangen.

Fischer, G., F. Janik, D. Müller and A. Schmucker (2009): The IAB Establishment Panel - things users should know. Schmollers Jahrbuch 129, 133-148.

Frenken, K., F. van Oort, and T. Verburg (2007): Related variety, unrelated variety and regional economic growth. Regional Studies 41, 685–697.

Fuchs, M. (2011): The determinants of local employment dynamics in Western Germany. Empirical Economics 40, 177-203.

Glaeser, E., H. Kallal, J. Scheinkman and A. Shleifer (1992): Growth in cities. Journal of Political Economy 100, 1126-1152.

Harris, R. and J. Moffat (2012): Is Productivity Higher in British Cities? Journal of Regional Science 52, 762-786.

Henderson, V. (2003): Marshall's scale economies. Journal of Urban Economics 53, 1-28.

Henderson, V., A. Kuncoro and M. Turner (1995): Industrial development in cities. Journal of Political Economy 103, 1067-1090.

Holmes, T. and J. Stevens (2004): Geographic concentration and establishment size: Analysis in an alternative economic geography model. Journal of Economic Geography 4, 227-250.

Hoogstra, G. and J. van Dijk (2004): Explaining firm employment growth: Does location matter? Small Business Economics 22, 179-192.

Hübler, O. and Jirjahn, U. (2003): Works councils and collective bargaining in Germany: the impact on productivity and wages. Scottish Journal of Political Economy 50, 471-491.

Ichniowski, C., Shaw, K. and Prennushi, G. (1997), The effects of human resource management practices on productivity: a study of steel finishing lines. American Economic Review 87, 291-313.

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

Jacobs, J. (1986): Cities and the Wealth of Nations, Harmondsworth.

Jacobs, J. (1969). The Economy of Cities. New York; Vintage.

Johansson, B. and J. Klaesson (2011): Agglomeration dynamics of business services. The Annals of Regional Science 47, 373-391.

Krugman, P (1991): Geography and trade. The MIT Press, Cambridge

Mameli, F., A. Faggian and P. McCann (2008): Employment growth in Italian local labour systems: Issues of model specification and sectoral aggregation. Spatial Economic Analysis 3, 343-360.

Marshall, A. (1890): Principles of Economics. London:Macmillan.

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

Müller, S. (2008): Capital stock approximation using firm level panel data. Jahrbücher für Nationalökonomie und Statistik 228, 357-371.

Nickell, S. (1981): Biases in dynamic models with fixed effects. Econometrica 49, 1417-1426.

Ottaviano, G. and D. Puga (1998): Agglomeration in the global economy: A survey of the 'New Economic Geography'. World Economy 21, 707-731.

Plümper, T. and V. Troeger (2007): Efficient estimation of time-invariant and rarely changing variables in finite sample panel analysis with unit fixed effects. Political Analysis 15, 124-139.

Roodman, D. (2009): How to do xtabond2: An introduction to difference and system GMM in Stata. The Stata Journal 9, 86-136.

Rosenthal, S. and W. Strange (2004): Evidence on the Nature and Sources of Agglomeration Economies. Handbook of Urban and Regional Economics 4, Amsterdam: Elsevier.

Saito, H. and M. Gopinath (2009): Plants' self-selection, agglomeration economies and regional productivity in Chile. Journal of Economic Geography 9, 539-558.

Slaughter, M. (2001): International trade and labor-demand elasticities. Journal of International Economics 54, 27-56.

Südekum J. (2006): Concentration and specialization trends in Germany since reunification. Regional Studies 40, 861-873.

van Beveren, I. (2012): Total Factor Productivity Estimation: A Practical Review. Journal of Economic Surveys 26, 98-128.

van Oort, F., M. Burger and J. Knoben (2012): Multilevel Approaches and the Firm-Agglomeration Ambiguity in Economic Growth Studies. Journal of Economic Surveys 26, 468-491.

Wagner, J. (2007): Exports and productivity: a survey of the evidence from firm-level data. The World Economy 30, 60-82.

Windmeijer, F. (2005): A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics 126, 25-51.

Appendix

Table A1: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
y	7.928	42.352	1,05	1.524.292
ln y	7,27	1,72	0,05	14,24
l				
ln l				
k	36.900.000	314.000.000	232,14	16.400.000.000
ln k	14,39	2,24	5,45	23,52
density	0,62	0,11	0,37	1,32
export	11,58	22,27	0	100
counc	0,35	0,48	0	1
struct2	0,15	0,36	0	1
struct3	0,11	0,31	0	1
fprop	0,08	0,26	0	1
age	0,05	0,22	0	1
tech	2,16	0,77	1	5

Table A2: Comparison of the results on y_{t-1}

	OLS	FE	Diff GMM	System GMM
log output$_{t-1}$	0.734*** (105.39)	0.161*** (10.20)	0.253*** (9.81)	0.317*** (3.28)
log labor	0.211*** (29.56)	0.324*** (12.70)	0.182** (2.38)	0.410** (2.11)
log capital	0.058*** (19.13)	0.092*** (3.87)	-0.054 (-0.54)	0.168** (2.44)
N	27,139	21,209	18,216	27,139

Robust standard errors in the OLS and FE regressions, in the GMM regressions Windmeijer-corrected standard errors. Significance at the *p<0.10, **p<0.05, ***p<0.001 level. t(z)-statistics in parentheses. Time and sector dummies included but not reported.

Table A3: Regression results for manufacturing and services

	Manufacturing			Services		
	OLS	FE	System GMM	OLS	FE	System GMM
log output $_{t-1}$			0.339 *** (9.04)			0.330 *** (7.84)
log labor	0.751 *** (36.44)	0.415 *** (6.76)	0.434 *** (4.76)	0.703 *** (39.43)	0.292 *** (7.40)	0.554 *** (6.21)
log capital	0.209 *** (14.65	0.147 *** (4.62)	0.117 ** (2.26)	0.189 *** (15.13)	0.054 (1.18)	0.058 (1.22)
log density	0.201 *** (5.61)	-0.000 (-0.39)	0.154 *** (3.11)	0.209 *** (6.93)	-0.265 (-0.73)	0.176 ** (2.31)
export	0.003 *** (5.20)	0.001 (1.61)	0.004 ** (2.34)	0.007 *** (5.65)	0.002 (1.07)	-0.003 (-0.65)
counc	0.215 *** (6.30)	0.055 * (1.67)	0.043 (1.02)	0.247 *** (5.47)	0.041 (0.76)	0.114 * (1.75)
struct2	0.194 *** (4.93)	0.034 (0.84)	-0.005 (-0.13)	0.317 *** (5.70)	-0.094 * (-1.77)	0.063 (1.10)
struct3	0.133 *** (3.31)	0.077 (1.60)	0.044 (1.37)	0.192 *** (4.16)	-0.049 (-0.87)	0.000 (0.00)
fprop	0.031 (0.81)	-0.011 (-0.19)	0.030 (0.62)	0.135 ** (1.98)	-0.014 (-0.15)	-0.029 (-0.23)
age	-0.002 (-0.94)	0.028 *** (5.26)	0.013 (1.57)	0.002 (1.04)	0.003 (0.57)	0.001 (0.04)
tech	-0.055 *** (-4.25)	0.005 (0.42)	0.003 (0.28)	-0.106 *** (-5.89)	-0.002 (-0.11)	0.008 (0.48)
east	-0.246 *** (-6.41)		-0.098 (-0.82)	-0.274 *** (-7.10)		-0.247 * (-1.67)
constant		10.577 *** (19.48)	6.225 *** (9.78)	9.190 *** (17.51)	13.209 *** (19.39)	6.920 *** (10.93)
N	12,019	12,019	11,409	9,932	7,370	6,750
R^2	0.89	0.84		0.80	0.55	
R^2 within		0.09			0.05	
R^2 between		0.86			0.55	
# instruments			374			373
Sargan (p-value)			0.216			0.645
AR(1) (p-value)			0.000			0.000
AR(2) (p-value)			0.100			0.138

Windmeijer-corrected standard errors. Significance at the *p<0.10, **p<0.05, ***p<0.001 level. t(z)-statistics in parentheses. Time and sector dummies included but not reported. The coefficients for log density are multiplied with 1,000. In the System GMM regressions, labor, capital, and exports are treated as endogenous, density as predetermined.