

Gornig, Martin; Schiersch, Alexander

**Working Paper**

## Agglomeration economies and firm TFP: Different effects across industries

DIW Discussion Papers, No. 1788

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Gornig, Martin; Schiersch, Alexander (2019) : Agglomeration economies and firm TFP: Different effects across industries, DIW Discussion Papers, No. 1788, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/193165>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

1788

## Discussion Papers

Deutsches Institut für Wirtschaftsforschung

2019

# Agglomeration Economies and Firm TFP: Different Effects across Industries

Martin Gornig and Alexander Schiersch

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

#### IMPRESSUM

© DIW Berlin, 2019

DIW Berlin  
German Institute for Economic Research  
Mohrenstr. 58  
10117 Berlin

Tel. +49 (30) 897 89-0  
Fax +49 (30) 897 89-200  
<http://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:  
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:  
<http://ideas.repec.org/s/diw/diwwpp.html>  
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

# Agglomeration economies and firm TFP: different effects across industries

Martin Gornig<sup>a,b</sup>, Alexander Schiersch<sup>a</sup>

<sup>a</sup>*German Institute for Economic Research (DIW Berlin), Mohrenstr. 58, 10117 Berlin, Germany*

<sup>b</sup>*Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany*

---

## Abstract

This paper analyzes the effect of agglomeration economies on firms' total factor productivity. We propose the use of a control function approach to overcome the econometric issue inherent to the two-stage approach commonly used in the literature. Estimations are conducted separately for four industry groups, defined by technological intensity, to allow for non-uniform effects of agglomeration economies on firms given their technological level. In addition, R&D is included to account for the firms' own efforts to foster productivity through creating and absorbing knowledge. Finally, radii as well as administrative boundaries are used for defining regions. The results confirm differences in the strength and even in the direction of agglomeration economies: While urban economies have the largest effect on TFP for firms in high-tech industries, they have no effect on TFP in low-tech industries. For firms in the latter industries, however, the variety of the local economic structure has an impact, while this is irrelevant for the TFP of firms in high-tech industries. Only localization economies have a positive and significant effect on TFP throughout, but the effect increases with technological intensity of industries. Throughout, R&D is also found to have a positive effect that increases with technological intensity.

*Keywords:* total factor productivity, manufacturing firms, agglomeration economies, spatial concentration, structural estimation

*JEL classification:* R11, R12, R15, D24

---

## 1. Introduction

The increasing availability of firm-level data allows for testing the effects of agglomerations on firm performance, thereby enriching the academic literature on agglomeration economies dating back to [Marshall \(1920\)](#) and [Weber \(1929\)](#). Although the empirical evidence at the micro-level is not unambiguous, it generally confirms the presumptions: there is evidence of a positive effect on employment ([Henderson, 1986](#); [Glaeser et al., 1992](#); [Caragliu et al., 2016](#)) and wages ([Combes et al., 2008](#); [Matano and Naticchioni, 2012](#); [Faberman and Freedman, 2016](#)); firms also show higher labor productivity ([Melo et al., 2009, 2017](#); [Ahrend et al., 2017](#)) and more innovativeness ([Hervas-Oliver et al., 2018](#)). However, so far there are few studies at the firm-level analyzing the effect of agglomeration economies on the total factor productivity (TFP) of firms and the results of these studies are mixed.

Moreover, despite the many differences between the studies on firm TFP and agglomeration economies, the literature on this topic shares four common characteristics: Firstly, all studies apply a two-stage approach. The first stage envisages the estimate of TFP by means of a production function. With the exception of [Harris and Moffat \(2015\)](#), agglomeration economies are ignored at this stage. Thereafter, the TFP is regressed on the variables of interest in order to (partly) explain productivity or its development. This causes an issue: It is hard to justify that agglomeration externalities affect TFP (second stage) while simultaneously pretending that the same variables are irrelevant when estimating that very TFP (first stage). Secondly, industries are pooled in the second stage, which implies uniform effects of agglomeration economies on TFP across industries. A third common characteristic is, apart from [Cainelli and Ganau \(2018\)](#), the predominant use of administratively defined regions, such as NUTS-2 or local labor markets (LLMs), and the like. Finally, none of the studies account for the R&D activity at firm-level and its effect on TFP.

This paper contributes to the literature in several ways. Firstly, using ideas and methods from the productivity literature allows us to estimate the effect of various agglomeration indicators on TFP while estimating TFP and the production function itself. More specifically, a control function approach along the lines of [Akerberg et al. \(2015\)](#) is applied, with the agglomeration indicators directly incorporated into the law of motion that describes TFP. This overcomes the issue of the classic two-stage approaches. Secondly, we include the R&D activity of firms in the function explaining TFP in order to account for the firms' own efforts to create knowledge in order to foster their own productivity. Third, regions are once defined according to administrative

entities and once by distance. Comparing the findings help to assess whether the practice of using administratively defined regions is critical. Finally, we allow that agglomeration externalities can affect TFP differently in accordance with the technological intensity of industries.

35 The analysis is conducted using firm-level data from German manufacturing firms covering the 2003 to 2014 period. Our main results show that localization economies and urban economies positively affect TFP. Yet, the results also reveal that the effect is stronger for firms in high-tech industries than those in medium low-tech or low-tech industries. The effect of stronger competition is also not equal across firms. In low-tech and medium low-tech industries, stronger competition  
40 seems to unfold negative effects. In more technologically intensive industries, the effect of competition is either slightly positive or has no effect at all. Differences between low-tech and high-tech industries are also found for the variety within industries and across industries. While related variety impacts firm TFP in low-tech industries, it is essentially irrelevant for TFP in high-tech industries.

45 Although the analysis confirms the importance of agglomeration externalities, it also reveals a strong persistence of productivity. This persistence is most pronounced in low-tech industries and the least pronounced in high-tech industries. The analysis also confirms the positive effects of R&D, thus highlighting the importance of the firms' own efforts for increasing their TFP. Moreover, the effect of R&D increases with the technological intensity of industries. The use of different regional  
50 definitions does not lead to decisively different results, proving their robustness. Yet, regional proximity affects the strength of the effects of some agglomeration economies. Further robustness checks include estimations for SMEs and large enterprises only, again confirming the main findings.

The remainder of this paper is organized as follows. Section 2 presents the standard model applied to date and discusses its main econometric issues. The section also discuss related studies.  
55 Section 3 introduce the empirical approach used in the analysis. The data are described in section 4. The empirical results, robustness checks, and a discussion of the limitations are provided in section 5 while section 6 concludes.

## 2. State of the literature on agglomeration economies and firm TFP

### 2.1. Benefits of agglomeration economies

60 Whether spatial concentration of economic activities has advantages is an issue studied by regional economics since industrialization. The concentration of firms in dense areas is considered

beneficial for firms for a large variety of reasons and mechanisms (Duranton and Puga, 2004; Rosenthal and Strange, 2004). *Inter alia*, firms benefit from *sharing*. This refers to the fact that firms share access to externalities that exist in denser areas. For example, all companies benefit  
65 from a better infrastructure that is built and maintained when more firms are active in the same region. Moreover, each firm benefits from the access to local suppliers and service providers whose numbers and diversity increases as more firms are active. Both the sales and procurement markets are also bigger.

The advantages of a large labor market are known as *matching* and *labor pooling*. It refers to  
70 the fact that a large number of firms within the same industry are usually accompanied by a large market of labor qualified for the different jobs within this industry. This makes it easier for firms to find the necessary number of workers with the specific qualifications when needed. Moreover, the matching accuracy between the skills needed by companies and the skills of the job-seeking workforce is better the larger the local labor market.

Another positive externality of agglomeration areas is associated with *learning*. This refers to  
75 the creation, diffusion, and accumulation of new knowledge within and between industries. For all of these learning and innovation processes, agglomeration areas comprising many and diverse stakeholders offer benefits.

There are a variety of indicators to address these effects in the literature. Yet, it must be  
80 emphasized that no indicator perfectly identifies a single specific mechanisms. Following Marshall (1920), Arrow (1962), and Romer (1986), the positive externalities associated with the specialization of a region on a specific industry are known as *MAR externalities* or *localization economies*. Consequently, the localization economies variable measures the importance of a region for an industry within a country. However, those benefits that come from density and urbanization, which  
85 are shared by all firms regardless of their industry affiliation, are considered *urbanization economies* (Isard, 1956). Jacobs (1969) argues that technological spillovers have a particularly large effect if innovation and new knowledge are created in one industry, but then adapted and used within a different industry. This refers to the learning effect outlined above. Accordingly, the diversity of the sector structure in a region is beneficial. The literature developed this approach further by  
90 differentiating between *related* and *unrelated variety* (Frenken et al., 2007). Finally, Porter (1990) argues that innovation and knowledge transfer increases as a result of more intense competition in a region.

## 2.2. TFP, production functions and agglomeration economies

The starting point of previous analyses on the relationship between TFP and agglomeration  
95 effects is a production function such as:

$$Y = \Omega F(L, K; \beta) e^\epsilon, \quad (1)$$

where  $Y$  is the output,  $L$  denotes labor,  $K$  capital and  $\Omega$  the TFP. The vector of parameters is  $\beta$  while  $\epsilon$  is the error term. Implicitly or explicitly, it is furthermore assumed that TFP itself is described by a function such as:

$$\Omega = G(\mathbf{Z}; \gamma), \quad (2)$$

where  $\mathbf{Z}$  contains the variables for agglomeration economies and  $\gamma$  is the vector of parameters.

100 To date, a two-stage approach is applied to estimate the effect of agglomeration economies, that is to estimate  $\gamma$ . The first stage is estimating Eq. (1) in order to obtain the coefficients for the production inputs ( $\hat{\beta}$ ) and with these calculating TFP as  $\hat{\Omega} = Y - F(L, K; \hat{\beta})$ . This initial estimation is conducted by a variety of different econometric methods, which themselves can have two or more steps. The TFP, obtained from the estimation of Eq. (1), is used subsequently in the  
105 second stage of the process in order to estimate Eq. (2).

Such procedure causes two issues. The first is the well-known simultaneity issue first emphasized by [Marschak and Andrews Jr. \(1944\)](#). It is caused by the fact that firms know about their productivity, or at least have some idea about it, consequently choosing their inputs accordingly, while TFP is an unobserved variable for researchers. TFP is, therefore, part of the error term.  
110 Hence, the actual estimation equation in logs is  $y = f(l, k; \beta) + \varepsilon$  with  $\varepsilon = \{\omega, \epsilon\}$ , whereby  $\epsilon$  is the i.i.d. error that picks up true measurement error, while  $\omega$  captures the productivity of a firm.<sup>1</sup> Consequently, the inputs are correlated with the error term causing an endogeneity issue and, ultimately, leading to biased parameter estimates. It follows that all estimations ignoring this issue are at risk of biased coefficients for the agglomeration variables in the second stage because  
115 of incorrect input coefficients in the first stage that lead to incorrectly calculated  $\hat{\omega}$ . In other words, from  $\hat{\beta} \neq E(\beta)$  it follows that  $\hat{\omega} \neq E(\omega)$ , which causes biased estimates of  $\hat{\gamma}$ . Beginning with [Olley and Pakes \(1996\)](#), control function approaches have been developed to overcome the

---

<sup>1</sup>Note that the introduction of an intercept  $\beta_0$  is not a solution as it just captures the time-constant industry average. The time-varying and firm-specific parts of the TFP remain in the error term.



simultaneity issue.<sup>2</sup> They have become a powerful and popular technique for production function estimations (e.g. [Aw et al., 2011](#); [De Loecker and Warzynski, 2012](#); [De Loecker, 2013](#); [Doraszelski and Jaumandreu, 2013](#); [De Loecker et al., 2016](#)).

The second issue is the correct incorporation of the agglomeration variables in the estimation. [De Loecker \(2011\)](#) and [De Loecker and Warzynski \(2012\)](#) show that all relevant and observed variables driving TFP "are required to be included when considering [the] intermediate inputs [function] as proxy" ([De Loecker and Warzynski, 2012](#), p.2446). Otherwise, the control function is not properly approximating TFP and, as a result, part of the firm specific TFP remains in the error term, thus leading to biased input coefficients. This is in line with the reasoning in [Harris and Moffat \(2015\)](#). They emphasize that ignoring the agglomeration variables when estimating the production function in the first stage makes these variables "part of the random error term  $\hat{\epsilon}_{it}$ . Clearly, ... the coefficients on the factor inputs and thus  $\ln T\hat{F}P_{it}$  from such an approach [are] biased because of an omitted variable(s) problem" ([Harris and Moffat, 2015](#), p.1023). As a result, all second stage analyses have biased results.

We overcome the simultaneity problem by applying a control function along the lines of [Akerberg et al. \(2015\)](#). The second issue is addressed by incorporating agglomeration economies into the control function when estimating the input coefficients of the production function. In fact, as shown in the subsequent section, our approach sees the simultaneous estimation of both  $\hat{\beta}$  and  $\hat{\gamma}$ .

### 2.3. Related literature

There is a limited, but growing, micro-level empirical literature on the relationship between TFP and the different measures for agglomeration economies as outlined above.<sup>3</sup> The study of [Martin et al. \(2011\)](#), which serves as a template for a number of subsequent studies, uses data covering French manufacturing firms over the 1996 to 2004 period to analyze the effect of localization economies, industry diversity, intensity of competition, and urbanization economies on the TFP of firms. The first stage sees the estimation of a Cobb-Douglas production function by means of

---

<sup>2</sup>The literature has a long history of addressing the simultaneity issue. For a comprehensive presentation see [Akerberg et al. \(2007\)](#).

<sup>3</sup>We abstain from presenting studies using solely aggregated data, whether aggregated at the sector level, the regional level, or both, such as [Antonelli et al. \(2011\)](#) or [Beugelsdijk et al. \(2018\)](#). Studies including agglomeration variables as standard inputs into a production function ([Maré and Graham, 2013](#); [Melo et al., 2009](#)) are also not considered here. The reason is that regardless of whether the dependent variable is gross output, value added, or labor productivity, the underlying assumption in such models is that agglomeration economies directly affect the output of firms, just like labor or capital. By definition, this means that they do not affect TFP.

the method proposed by [Levinsohn and Petrin \(2003\)](#). This method is itself a two-step approach. Production function estimations are conducted at the two-digit industry level. After the coefficients are estimated and TFP is calculated, the agglomeration variables are regressed on the TFP in the second stage of the analysis. The agglomeration variables are calculated for 94 departments (NUTS 3 level) and 341 employment areas in continental France. Using OLS and the generalized method of moments (GMM), urbanization and localization are found to positively affect TFP in most regressions. In contrast, coefficients of the competition variable are ambiguous while industry diversity seem to have no effect whatsoever.

The approach has three main issues that also apply to the studies following [Martin et al. \(2011\)](#). First, the method of [Levinsohn and Petrin \(2003\)](#) does not solve the endogeneity issue of the production function estimation, as shown in the seminal and widely recognized study of [Akerberg et al. \(2006\)](#). Secondly, the assumption that agglomeration economics affects the TFP of firms is imposed, and this is then tested by estimating the relationship between the respective variables and TFP. Yet, simultaneously, the authors pretend that agglomeration economics are completely irrelevant when they estimate the production function and calculate the TFP. As discussed in the previous section, this, again, causes an endogeneity issue. Finally, the second stage of the analysis uses pooled data. This seems to be inconsistent. First the authors assume that industries differ so substantially in terms of the applied technology, the labor requirements, capital requirements, the output markets, and the procurement markets etc. that it is necessary to estimate separate production functions. But once it comes to the question of how the TFP of firms are affected by agglomeration variables, e.g. the concentration of an industry in a region, they assume that there are no differences whatsoever between firms from totally different industries.

The study of [Martin et al. \(2011\)](#) has a number of characteristics that are also common to other studies on this topic. Firstly, and as already emphasized, is the use of the method of [Levinsohn and Petrin \(2003\)](#) for estimating the production function and deriving TFP. This is also the case in [Combes et al. \(2012\)](#), [DiGiacinto et al. \(2014\)](#), [Cainelli et al. \(2015\)](#), and [Holl \(2016\)](#). These studies, as well as [Harris and Moffat \(2015\)](#) and [Cainelli and Ganau \(2018\)](#), also apply a Cobb-Douglas specification in the first stage. Moreover, the production function is generally estimated separately for each two-digit industry. However, the second stage analyses are not conducted separately, neither for each two-digit industry nor for industry groups; rather, the data are pooled for the entire manufacturing industry. This imposes the assumption that the effects of agglomeration

economies are uniform across all industries. All studies listed here, with the exception of [Cainelli and Ganau \(2018\)](#), use administratively defined regions, be it NUTS-2 regions, NUTS-3 regions, or aggregated NUTS-regions. Moreover, the company's own R&D activities are generally ignored, both when estimating the production function as well as in the second stage when agglomeration variables are used to explain the estimated first-stage TFP. Finally, the agglomeration variables are also ignored when estimating the production function, thus causing the aforementioned issue. An exception regarding the last two issues is [Harris and Moffat \(2015\)](#). They include agglomeration economies and R&D in the first stage production function estimation. However, they still ignore R&D in the function that explains the TFP of firms.

The overall picture emerging from existing studies is mixed, with evidence regarding the benefits of agglomeration economies going multiple ways. Using French data, [Combes et al. \(2012\)](#) find that firms in denser populated areas have higher TFP, which points to positive effects from urbanization economies, while selection due to stronger competition is irrelevant. This is partly in line with the findings of [Martin et al. \(2011\)](#) on urban economies and competition. Contrary results are found in [Cainelli et al. \(2015\)](#), which analyze Italian manufacturing firms and use the same variables definition for competition intensity, localization, urbanization, and diversity as [Martin et al. \(2011\)](#). They constantly find positive effects for competition while the coefficients for location economies are only positive and significant in a very few specifications. The coefficients for urban economies and diversity are usually not significantly different from zero. The results in [Harris and Moffat \(2015\)](#) suggest a positive relationship between the population density of regions in the UK, which serves as indicator for urban economies, and TFP growth in regions. [DiGiacinto et al. \(2014\)](#), using data from Italian manufacturing firms and defining three regions that capture urban centers, industry districts, and others, find that firms in the first two regions have significantly higher TFP. This is interpreted as support for the beneficial effects of urban economies and location economies. [Cainelli and Ganau \(2018\)](#), using distances to define regions, analyze the effect of intra-industry (localization economies) and inter-industry (Jacob) externalities on TFP. They find a positive coefficient for the intra-industry density measure on short and medium distances, but a negative one for the inter-industry density measure on the distances up to 15 km. This is interpreted as positive effects from localization economies and negative ones from variety. Finally, [Holl \(2016\)](#) uses data on Spanish manufacturing firms when analyzing the effect of employment density and population density at the municipal level as well as the distance to the next highway on the TFP

205 of firms. She finds that a nearby highway is beneficial as is higher density. The latter points to a positive effects of urbanization.

The studies that deviate the most in terms of their econometric approach are [Harris and Moffat \(2015\)](#) and [Cainelli and Ganau \(2018\)](#). [Harris and Moffat \(2015\)](#) acknowledges that ignoring the agglomeration variables when estimating the production function in the first stage of the process lead to biased input coefficients and, as a result, all second stage analyses would also have biased results. The authors overcome the issue by including the agglomeration variables together with further control variables when estimating the production function by means of GMM ([Blundell and Bond, 1998](#)).<sup>4</sup> The second part of the analysis consists of regressing the variables capturing urban economies and regional human capital together with a set of control variables on TFP growth by means of OLS. Note that TFP growth is no longer firm specific, but rather it is aggregated at regional level. We follow [Harris and Moffat \(2015\)](#) by explicitly controlling for agglomeration economies when estimating the production function. Our analysis deviate substantially by conducting the analysis separately for industry groups, by using different spatial definitions, by including R&D as driver of TFP, and by overcoming the classical two-stage approach as described in section 3.

220 [Cainelli and Ganau \(2018\)](#) is in line with most studies with respect to the first stage of the analysis, as they estimate a simple Cobb-Douglas production function per two-digit Italian manufacturing industry and obtain TFP from these estimates. While applying the method proposed by [Wooldridge \(2009\)](#) in order to overcome the simultaneity issue, they do not control for agglomeration economies or R&D effort; therefore the omitted variables problem, as outlined by [Harris and Moffat \(2015\)](#), remains an issue. However, while all studies listed in this section use administratively defined regions when constructing the agglomeration variables, [Cainelli and Ganau \(2018\)](#) deviate by defining three circles around the location of each firm. They use the radii  $0 \leq d_5 \leq 5$ ,  $5 < d_{15} \leq 15$  and  $15 < d_{30} \leq 30$  and calculate a density measure for the inter- and the intra-industry externalities for each circle. In the second stage, these six regional variables and additional controls are regressed on TFP growth, which captures the difference in TFP between 2009 and 2012, the start and the end of the observation period. Our analysis follows [Cainelli and Ganau \(2018\)](#), by

---

<sup>4</sup>*Inter alia*, they also include R&D when estimating the production function, but referring to a Griliches-type functional relationship between output and R&D ([Griliches, 1979](#)). This approach postulate the functional form  $Y = DC^\alpha L^\beta K^\gamma e^{\lambda t + u}$ , with  $K$  being the knowledge stock, which often is substituted with R&D. In such a specification,  $K$  directly affects the output but not the TFP. Consequently, firm-specific R&D is not included in the second part of their analysis.

applying radii in addition to administratively defined regions. It deviates by not using different circles simultaneously, by refrain from the classical two-stage approach as outlined in section 2.2, by overcoming the omitted variable bias when estimating the production function, and by including  
 235 the firms' own R&D effort together with all agglomeration variables in the analysis.

### 3. Model and estimation strategy

We follow the literature insofar as we assume that a production function  $F(\cdot)$  exists that transforms labor and capital inputs into outputs, as well as a function  $G(\cdot)$  that explains TFP and how it is affected by relevant variables. The function to be estimated, in logs, is shown in Eq. (3):

$$y_{it} = f(l_{it}, k_{it}, c_{it}; \beta) + \underbrace{\omega_{it}}_{g(Z; \gamma)} + \epsilon_{it}, \quad (3)$$

240 where  $i$  and  $t$  capture the firm and time dimension, respectively, to account for the panel structure of the data.

We deviate from the aforementioned studies by following the productivity literature for modelling and estimating the relationship between TFP and the covariates which capture the characteristics of agglomerations. As in [Aw et al. \(2011\)](#), [Doraszelski and Jaumandreu \(2013\)](#) or [De Loecker et al. \(2016\)](#), covariates that directly affect TFP are included in the law of motion explaining TFP.  
 245 As shown subsequently, the elasticities of functions  $f(\cdot)$  and  $g(\cdot)$  are estimated simultaneously using a control function approach along the lines of [Akerberg et al. \(2015\)](#).

Since the seminal paper of [Levinsohn and Petrin \(2003\)](#), control function approaches utilize the assumption of an intermediate input demand function ( $m_{it} = h_t(\dots, \omega_{it})$ ) with certain characteristics. These are, *inter alia*, that the intermediary input  $m_{it}$  is a fully flexible input, TFP ( $\omega_{it}$ )  
 250 is the only unobserved state variable in  $h_t(\cdot)$  and  $h_t(\cdot)$  is strictly monotonic in  $\omega_{it}$ . Because of these assumptions,  $h_t(\cdot)$  is invertible and allows to obtain a function that serves as proxy for the unobserved TFP in Eq. (1), i.e.  $\omega_{it} = h_t^{-1}(\dots, m_{it})$ .<sup>5</sup> Apart from the intermediate input, capital and labor are also included in the control function. The latter dates back to the seminal paper of  
 255 [Akerberg et al. \(2006, 2015\)](#) in which the authors prove that " $l_{it}$  is functionally dependent on  $k_{it}$ ,  $m_{it}$  and  $t$ " ([Akerberg et al., 2015](#), pp. 2423), which is why identification of the labor coefficient

<sup>5</sup>See [Levinsohn and Petrin \(2003\)](#) for the proof of invertability.

in the first step fails as it is done in [Levinsohn and Petrin \(2003\)](#). Moreover, [De Loecker and Warzynski \(2012\)](#) states that all variables observed and relevant to the TFP should be included in the proxy function. This leads to the proxy function  $\omega_{it} = h_t^{-1}(l_{it}, k_{it}, m_{it}, \mathbf{z}_{it})$ , where the vector  $\mathbf{z}_{it}$  captures additional variables. Furthermore, the "exact variables to be included in  $\mathbf{z}_{it}$  depend on the application but will definitely capture variables leading to differences in optimal input demand across firms" ([De Loecker and Warzynski, 2012](#), p.2446). If the general premise is correct that agglomeration economies affect companies and their productivity in many different ways, then the respective agglomeration variables belong into  $\mathbf{z}_{it}$ .

By substituting the unobserved TFP with  $h_t^{-1}$ , and after including the variables for agglomerations economies, the function to be estimated in the first-step of the control function framework is:

$$y_{it} = \varphi_t(l_{it}, k_{it}, m_{it}, r_{it}, \mathbf{z}_{it}) + \mathbf{X}\theta + \epsilon_{it}, \quad (4)$$

with  $\varphi_t(\cdot) \equiv f(l_{it}, k_{it}; \boldsymbol{\beta}) + h_t^{-1}(l_{it}, k_{it}, m_{it}, r_{it}, \mathbf{z}_{it})$ . Note that we also include R&D in function  $h_t^{-1}$ , because the firms' own R&D efforts positively affect its TFP, as shown by [Aw et al. \(2011\)](#), [Doraszelski and Jaumandreu \(2013\)](#), [Kancs and Siliverstovs \(2016\)](#), and others.<sup>6</sup> This also takes into account that "firms are neither equally equipped to receive knowledge nor homogeneously willing to serve as sources of spillover" ([Cainelli and Ganau, 2018](#), p.923). By including R&D, our analysis actually controls for this ability and willingness. This goes beyond previous studies on the relationship between agglomeration economies and the TFP of firms. At this stage, additional control variables ( $\mathbf{X}$ ) are included, such as legal form, year, and an east dummy.

Eq. 4 is estimated as first step of the control function approach using OLS. Because the functional form  $h_t^{-1}$  is unknown, we follow the literature and use a polynomial as a proxy. As pointed out before, the input coefficients obtained from this estimation are not identified, thus requiring a second step. Control function approaches impose the assumption that TFP in  $t$  is determined by past experiences summarized in the information set  $I_{it-1}$  and a shock to productivity ( $\omega_{it} = E[\omega_{it}|I_{it-1}] + \xi_{it}$ ). The shock  $\xi_{it}$  is assumed to be random, which is why it is uncorrelated with the information set ( $E[\xi_{it}|I_{it-1}] = 0$ ). Following the literature,  $\omega_{it}$  is governed by a first-order Markov process and also driven by other relevant state variables. Assuming a linear relationship

---

<sup>6</sup>There is a large literature on the relationship between R&D and the productivity of firms as well as the direct effect of R&D on firm performance that is not discussed here. [Ugur et al. \(2016\)](#) and [Møen and Thorsen \(2015\)](#) provide an overview and conduct meta analyses confirming the beneficial effect of R&D.

between R&D effort, agglomerations economies, and TFP, this yields:

$$\begin{aligned}\omega_{it} &= g(\omega_{it-1}, r_{it-1}, \mathbf{z}_{it}; \boldsymbol{\gamma}) + \xi_{it} \\ &= \gamma_0 + \gamma_1 \omega_{it-1} + \gamma_2 r_{it-1} + \gamma_3 z_{1,it} + \gamma_4 z_{2,it} + \dots + \xi_{it}.\end{aligned}\tag{5}$$

Estimating Eq. 5 is hampered by the fact that TFP is still unknown. Yet, from  $\varphi_t(\cdot) = f(l_{it}, k_{it}; \boldsymbol{\beta}) + h_t^{-1}(\cdot)$  and  $\omega_{it} = h_t^{-1}$  follows

$$\begin{aligned}\varphi_t &= f(l_{it}, k_{it}; \boldsymbol{\beta}) + \gamma_0 + \gamma_1 [\varphi_{t-1} - f(l_{it-1}, k_{it-1}; \boldsymbol{\beta})] + \gamma_2 r_{it-1} \\ &\quad + \gamma_3 z_{1,it} + \gamma_4 z_{2,it} + \dots + \xi_{it}.\end{aligned}\tag{6}$$

Eq. 6 is estimated by means of GMM in an iterative process that minimizes the error term using starting values for input coefficients and  $\hat{\varphi}_t$  calculated in the first step of the process. Note that by estimating Eq. 6, not only are the input coefficients ( $\hat{\boldsymbol{\beta}}$ ) obtained, but also, simultaneously, the coefficients for the agglomeration variables ( $\hat{\boldsymbol{\gamma}}$ ).

280 For the coefficients to be consistent, the variables need to be orthogonal to  $\xi_{it}$ , which is unobserved and, therefore, part of the error term. Identification is based on timing assumptions regarding the firms' decisions for the different inputs. [Akerberg et al. \(2015\)](#) show that contemporaneous labor is a rather flexible input and that this creates an issue: even if labor is considered less flexible than material<sup>7</sup> and even if firms decide upon its use after  $t - 1$  but before  $t$ , hence  
285 at  $t - b$  with  $0 < b < 1$ , it is still at least partly influenced by  $\xi_{it}$ . Therefore, the authors advise to use lagged labor as an instrument because the decision regarding its use is taken at  $t - b - 1$  and, thus,  $l_{it-1}$  is not correlated with  $\xi_{it}$ . The same reasoning holds for R&D (see [Aw et al., 2011](#); [Doraszelski and Jaumandreu, 2013](#); [Kancs and Siliverstovs, 2016](#)). Since [Olley and Pakes \(1996\)](#), it is assumed that a firm's decision on investing is taken in  $t - 1$  but fully implemented in  $t$ , which  
290 is why  $k_{it}$  is uncorrelated with  $\xi_{it}$ . Finally, we impose the assumption that location is exogenous to the firm, which is why all contemporaneous agglomeration variables are orthogonal to  $\xi_{it}$ . Hence, the following moment conditions apply:  $E[\xi_{it} | \mathbf{W}_t] = 0$  with  $\mathbf{W}_t = \{l_{it-1}, k_{it}, r_{it-1}, \mathbf{z}_{it}\}$ . Following the bulk of the literature,  $f(\cdot)$  is assumed to be a Cobb-Douglas function.

---

<sup>7</sup>This is a necessary assumption.

## 4. Data

### 295 4.1. Datasets

The analysis uses the AFiD-Panel manufacturing firms, which contains the *Cost Structure Survey for manufacturing firm (CS)*, the *Investment Census for manufacturing firms (IC)*, and the *Annual Report for manufacturing firms (AR)*. The data are collected and processed by the German *Federal Statistical Office* and the *Statistical Offices of the Länder*.<sup>8</sup> The three datasets  
300 are combined by the Statistical Office using unique firm IDs. [Konold \(2007\)](#), [Koch \(2007\)](#), and [Fritsch et al. \(2004\)](#) provide a detailed discussions of the datasets.<sup>9</sup> The combined dataset covers the 2003 to 2014 period. It is enriched with deflator series for material, value added, and capital as well as depreciation rates at the two-digit industry level. These data are provided by the Statistical Office. The georeference codes of municipalities, the assignment of communities to the individual  
305 administrative units, i.e. NUTS-2 or NUT3-regions, are provided by the *Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR)*.<sup>10</sup> In the following, we focus on data preparation and on those characteristics of the datasets that are relevant for the analysis.

The AR is a census of all manufacturing and quarrying firms with at least 20 employees. In some  
310 industries, the threshold is 10 employees because of the size structure of the respective industries.<sup>11</sup> It contains only a few items, among them, the number of employees in each firm. Due to its full coverage of the manufacturing industry, the employment numbers from the AR are used to calculate the different indicators for agglomeration economies as described in section 4.2.

The IC is also a census of all manufacturing and quarrying firms with a threshold of 20 or,  
315 in some industries, 10 employees. It covers the investment of firms, which are used to calculate the physical capital stock per firm. This is done with the perpetual inventory method (PIM) in the version used by the OECD. In this version, the initial capital stock is the average of two separately calculated initial capital stocks. The first utilizes the steady state assumption and, thus, is calculated as the ratio of the lowest investment observed over the depreciation rate. The second  
320 is the product between the observed labor and the capital-labor ratio at the two-digit industry level

---

<sup>8</sup>Hereafter, we use the term Statistical Office for simplicity.

<sup>9</sup>See also [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de) for details on the data and for data access.

<sup>10</sup>For more information, see [https://www.bbsr.bund.de/BBSR/EN/Home/homepage\\_node.html](https://www.bbsr.bund.de/BBSR/EN/Home/homepage_node.html).

<sup>11</sup>The share of employees working in firms with less than 10 total employees in the German manufacturing industry is 7 percent.



as provided by the Statistical Office.

The CS contains a full census of all firms with more than 500 employees and a representative sub-sample for firms with 20-499 employees. The latter is held constant for four years before a new sub-sample is drawn. The CS provides the data on the number of employees, value added,  
325 consumption of raw materials and supplies, R&D activities, as well as data on the legal form, the federal state, and the industry code.

For privacy reasons, firm-level data do not contain the individual address of firms. Yet, the data include the so called "*Amtlichen Gemeindeschlüssel*" (*AGS*). This is an 8-digit code that is unique for each municipality.<sup>12</sup> Thus, each firm can be assigned to a municipality and each municipality  
330 to a region.

#### 4.2. Indicators of agglomeration economies

Regions are defined in two ways. Firstly, we make use of administratively defined regions, namely the 96 "*Raumordnungsregionen*" (*ROR*).<sup>13</sup> These regions are made up of NUTS-3 regions and are constructed such that they capture an economic center and its surrounding. The affiliation  
335 of a municipality to an ROR is determined by commuter flows, political structures, central services, etc., and set by the BBSR. The BBSR provides the table containing the municipalities per ROR.<sup>14</sup> Each firm is assigned to one of the 96 ROR, based on the 8-digit AGS code of the municipality where the firm is located.

Such a definition comes at a cost though: the well-known modifiable areal unit problem (MAUP)  
340 ([Openshaw, 1984](#)). It also refers to the issue that all agglomeration indicators – when constructed for regions such as RORs (or NUTS-1 region, NUTS-2 regions, etc.) – ignore all information from neighboring regions and the firms therein. It means, for example, that the RORs that surround Berlin, which is the largest city in Germany and an own ROR,<sup>15</sup> will very likely have indicator values that are close to those in truly rural areas. However, Berlin is essential for the majority  
345 of firms located in these ROR, which encircle the city completely. It is not just the major sales market, it is also the place where a large number of their partners and competitors are located.

---

<sup>12</sup>Figure [A.1](#) in the appendix provides a visual overview of the municipality.

<sup>13</sup>Figure [A.2](#) in the appendix provides a visual overview of the ROR.

<sup>14</sup>The ROR are largely compatible with the German metropolitan areas as defined by the [OECD \(2012\)](#).

<sup>15</sup>While the concept of a center and its surroundings is generally applied for the definition of RORs, the BBSR ignores it with respect to Berlin, Bremen, and Hamburg. Each of these cities is an own ROR, but without the surrounding countryside. Political considerations related to the German Federalism underlie this issue. Nevertheless, the MAUP is a general issue that applies to all ROR, not just to Berlin.

The second approach for defining regions, therefore, adapts the idea of [Cainelli and Ganau \(2018\)](#) and make use of radii. The georeference codes of the municipalities are used instead of the georeference codes of the firms, because latter is not included in the dataset due to the privacy policy rules of the statistical office. The codes are provided by the BBSR and are part of the external information added to the data.

We apply the QGIS program and, essentially, draw a circle around each municipality using either a 10 km or a 25 km radius. All municipalities within the respective radius are part of the region that is defined by a municipality and the radius.<sup>16</sup> Besides addressing the MAUP issue, the use of a small region (10 km) also addresses the question if agglomeration externalities have different effects in smaller regions.

Based on [Martin et al. \(2011\)](#), we construct the following indicators for agglomeration economies for each region, regardless of the spatial definition applied:

*Localization economies.* The externalities that come with a high concentration of firms in the same industry in a region are understood as localization economies. The variable is defined as

$$LOC_{it}^{jk} = \frac{\sum_i EMP_{it}^{jk} - EMP_{it}^{jk}}{\sum_i \sum_k EMP_{it}^{jk}}, \quad (7)$$

where  $EMP_{it}^{jk}$  is the number of employees in firm  $i$  at time  $t$  in a two-digit industry  $j$  and region  $k$ . The denominator is the sum of all employees working in industry  $j$  at time  $t$  in the entire country. In other words, the variable measures the region's share of employment in an industry, corrected for the number of employees in the firm under consideration. The variable varies by firm, time, region and two-digit industry.

*Competition.* The variable  $POR_{it}^{jk}$  captures regional competition intensity among firms in an industry  $j$  at time  $t$  in region  $k$ . It is calculated by means of the inverse of the Herfindahl index of employment concentration:

$$POR_{it}^{jk} = \frac{1}{\sum_i \left( \frac{EMP_{it}^{jk}}{EMP_t^{jk}} \right)^2}, \quad (8)$$

---

<sup>16</sup>The georeference code refers to the center of a municipality. Hence, a municipality is only assigned to a region if its center is within the respective radius.

where  $EMP_t^{jk}$  is the sum of all employees working in industry  $j$  at time  $t$  in region  $k$ . The variable  
 370 varies across regions, industries and time, but is identical for all firms within a specific two-digit  
 industry in a region at time  $t$ .

*Urbanization economies.* Population density, measured as inhabitants per square kilometer, is reg-  
 ularly used as indicator for urbanization economies. For firm  $i$  at time  $t$  that is located in region  
 $k$  and active in the two-digit industry  $j$ , the urbanization economies variable is defined as:

$$URB_{it}^{jk} = \frac{Inhabitants_t^k}{(km^k)^2}. \quad (9)$$

375 Given this definition, all firms in a region, regardless of their industry affiliation, face the same  
 population density. Thus, the variable varies across time and regions.

*Related and unrelated variety.* In contrast to [Martin et al. \(2011\)](#), which only uses a single variable  
 to capture diversity, we follow [Basile et al. \(2017\)](#) and distinguish between related and unrelated  
 variety. The related variety is measured by means of the inverse of the Herfindahl index of employ-  
 380 ment using the shares of the four-digit industries  $o$  within a two-digit industry  $j$  in a region  $k$  at  
 time  $t$ . Accordingly, the variable is defined as:

$$VR_{it}^{jk} = \frac{1}{\sum_o \left(\frac{EMP_t^{ok}}{EMP_t^{jk}}\right)^2} \quad (10)$$

where  $EMP_t^{ok}$  is the sum of employment in each four-digit industry  $o$  in region  $k$  at time  $t$ ,  
 whereby only that four-digit industries  $o$  are considered that belong the two-digit industry  $j$  that  
 firm  $i$  belongs to. Simply put, it measures the diversity within a two-digit industry in a region.  
 385 The variable varies over time, region, and two-digit industry. The unrelated variety uses the shares  
 of two-digit industries  $j$  in region  $k$  at time  $t$ :

$$VUR_{it}^{jk} = \frac{1}{\sum_j \left(\frac{EMP_t^{jk}}{EMP_t^k}\right)^2}, \quad (11)$$

where  $EMP_t^k$  is the sum of all employees working in region  $k$  at time  $t$ , whereas  $EMP_t^{jk}$  is the sum  
 of all employees working in industry  $j$  at time  $t$  in region  $k$ . It measures the diversity of two-digit  
 industries within a region. The variable varies over time and regions.

The unprocessed dataset contains 467,397 observations for the 2003 to 2014 period. The observations from the Annual Report are used to construct the indicators for agglomeration economies. The production function estimations, however, are limited to those firms that are simultaneously part of the investment census and the cost structure survey, as these datasets contain the production function inputs and the value added. As shown in Table B.2 (column IS&CS), this limits the number of observations that can be used in the production function estimation to roughly 186,000. From this dataset, a few industries are dropped, such as mining, mainly because of too few observations.<sup>17</sup>

Table 1: Descriptive statistics

Variable	Description (unit)	P5	Mean	Median	P95	Std. Dev.	N
$Y_{it}$	Value added (1000 Euro)	836.11	22294.45	4546.03	67406.55	210720.74	174,863
$L_{it}$	Employees	26	290.13	89	859	2160.98	174,863
$K_{it}$	Capital (1000 Euro)	1544.6	46148.81	8810.37	134451.34	420673.86	174,863
$M_{it}$	Material (1000 Euro)	391.41	38993.36	5082.96	114146.8	484444.42	174,863
$R_{it}$	R&D	0	0.44	0	1	0.5	174,863
$URB_{it}^{jk}$	Urbanization	130.53	537.6	302.94	1519.2	596.82	174,863
$POR_{it}^{jk}$	Porter	2.22	16.79	10.62	50.04	18.73	174,863
$VUR_{it}^{jk}$	Unrelated variety	3.8	8.87	9.17	12.82	2.55	174,863
$VR_{it}^{jk}$	Related variety	1.24	4.02	3.47	8.4	2.24	174,863
$LOC_{it}^{jk}$	Localization	0.18	2.25	1.23	7.59	3.25	174,863

Source: Research Data Centres of the Federal Statistical Office and the Statistical Offices of the Länder, 2003-2014; own calculations.

Following Cainelli et al. (2015) those observations that are below the 0.5 percentile or above the 99.5 percentile in terms of the labor and capital productivity are dropped as outliers. This is done separately for each two-digit industry. Finally, all observation with missings in the relevant variables are also dropped. Material, value added, and capital (the respective investments) are deflated with respective series per two-digit industry. The final dataset used in the production function estimations is an unbalanced panel with 174,863 observations. Table 1 shows the main

<sup>17</sup>We follow Richter and Schiersch (2017) and drop observations from the mining industries, from manufacturing of tobacco products (C12), and from manufacturing of refined petroleum products (C19). The number of observations is insufficient in these industries, which leads to conflicts with the privacy policy rules applied in the Statistical Offices. We also drop manufacturing of other transport equipment (C30). The industry is quite heterogeneous as it includes shipbuilding, train manufacturing, aircraft construction, truck construction, as well as construction of rockets and satellites. This results in implausible estimates for labor and capital.

405 descriptive statistics.

The variables of interests are characterized by high variation. R&D activities are reported for just 44 percent of the firm/year observations. The descriptive statistics at the two-digit industry level reveal an even stronger heterogeneity, with a share of 10 percent in industry C10 (*Manufacturing of beverages*) and a share of 75 percent in industry C26 (*Manufacturing of computer electronic and optical products*). The population density ranges from less than 130 inhabitants per square kilometer to more than 1,500. The dataset contains municipalities with a value of 2.25 or less in the Porter variable, pointing to a high concentration within these municipalities. At the same time, the value of 50 signals a very low concentration as the respective value for the Herfindahl index is 0.02. Similar pictures of high heterogeneity can be found regarding the related and unrelated variety as well as the for the localization variable.

## 5. Results

### 5.1. Main Results

Following Eurostat, four groups of industries are created in accordance with the technological intensity: high-technology, medium high-technology, medium low-technology, and low-technology industries.<sup>18</sup> The estimations are conducted separately for each group as well as for the pooled data. In each case, industry dummies for two-digit industries are included as additional controls in  $\mathbf{X}$ .

Table 2 shows the estimation results when the indicators of agglomeration economies are constructed for RORs. The first two rows contain the coefficients for labor and capital. The remaining rows contain the coefficients for the variables in the law of motion ( $g(\cdot)$ ). The coefficients for labor and capital are significant and have a magnitude within reasonable ranges in all estimations.

Column (5) contains the results for estimations across all industries. All indicators for agglomeration economies as well as R&D and past TFP have significant coefficients. Importantly, the firms' own R&D activity prove to be an important driver of TFP, despite controlling for the knowledge effects that come through the agglomeration externalities. Hence, supporting firms in their R&D activities is a meaningful economic policy. Yet, even after controlling for the firms' own knowledge

---

<sup>18</sup>See Table B.1 for the list of industries and the assignment of industries to the different groups. We also refer to [http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:High-tech\\_classification\\_of\\_manufacturing\\_industries](http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:High-tech_classification_of_manufacturing_industries).

creation, agglomeration economies remain important too. Location and urbanization economies positively affect TFP. This also holds for related and unrelated variety, the latter, however, only at the 10% level. In contrast, the effect of additional competition is negative. Finally, the coefficient of lagged TFP reveals a high persistence of productivity.

Table 2: Elasticities of production inputs and variables in the law of motion; technology groups; regions: ROR

	(1)	(2)	(3)	(4)	(5)
	low-tech	medium low-tech	medium high-tech	high-tech	All
	production function $f(\cdot)$ – output elasticities ( $\hat{\beta}$ )				
labour	0.624*** (0.00691)	0.773*** (0.00524)	0.780*** (0.00556)	0.93*** (0.01349)	0.715*** (0.00353)
capital	0.410*** (0.00713)	0.282*** (0.00459)	0.286*** (0.00489)	0.136*** (0.01196)	0.332*** (0.00323)
	law of motion $g(\cdot)$ – elasticities regarding TFP ( $\hat{\gamma}$ )				
TFP <sub><i>t</i>-1</sub>	0.982*** (0.00108)	0.944*** (0.00165)	0.938*** (0.00175)	0.923*** (0.00403)	0.958*** (0.0008)
R&D	0.0043*** (0.00051)	0.0037*** (0.00047)	0.0064*** (0.00059)	0.0201*** (0.0017)	0.0067*** (0.00029)
URB	0.0000 (0.00029)	0.0026*** (0.00032)	0.0014*** (0.0004)	0.0053*** (0.00087)	0.0015*** (0.00019)
LOC	0.0006** (0.00024)	0.0009*** (0.00026)	0.001*** (0.00026)	0.0033*** (0.00064)	0.0005*** (0.00014)
POR	-0.0024*** (0.00033)	-0.0015*** (0.00032)	0.0029*** (0.00057)	0.0004 (0.0012)	-0.0019*** (0.00021)
VUR	0.0012* (0.00061)	-0.0022*** (0.00066)	0.002** (0.0008)	-0.0022 (0.00175)	0.0007* (0.00039)
VR	0.0013** (0.00056)	-0.0018*** (0.00063)	-0.0054*** (0.00082)	0.0014 (0.00191)	0.0012*** (0.00033)
Constant	0.111*** (0.00672)	0.42*** (-0.0122)	0.456*** (-0.0128)	0.666*** (-0.036)	0.281*** (0.00546)
N	43,911	42,172	41,286	8,025	135,394

*Notes: \* significance at  $p < .1$  level, \*\* significance at  $p < .05$  level, \*\*\*Significance at  $p < .01$  level. East/West, year, industry and legal form are controlled for in the first stage of ACF procedure. Hansen-Test results are not reported as the  $p$ -value is always close 1, as it needs to be the cases in GMM specifications with as many instruments as variables. The number of observations drops in comparison to Table 1 because two subsequent observations per firm are required due to the lag structure imposed.*

Beyond these average effects, the analysis, for the first time, allows for assessing the heterogeneity of agglomeration economies regarding their effect on TFP across different sectors within the manufacturing industry. To begin with, such heterogeneity is found with respect to the factors directly under control of the firms: the TFP in low-tech sectors is determined the most by

440 the previous period TFP, while TFP has a considerably lower persistence in high-tech sectors. A company's own research and development activities have, by far, the strongest effect on TFP in high-tech industries. In medium high-tech sectors, the impact of R&D is lower than in high-tech sectors, but still noticeably higher than in low-tech and medium low-tech industries.

Considerable differences are also found for urbanization. The level of urbanization has the 445 largest impact on TFP in high-tech industries, followed by medium low-tech and medium high-tech sectors. Among other potential reasons, this suggests that public research and education activities, which are concentrated in large population centers with high population densities, are likely to generate positive productivity effects. Another argument in favor of this interpretation is that urbanization benefits do not affect firm TFP in low-tech industries that rely little on such 450 environment. The MAR externalities have a similar sectoral pattern, with the strongest effect in high-tech industries and the lowest coefficient for low-tech sectors. The positive coefficients throughout support the presumption that clustering has positive externalities.

Large differences are found for the remaining indicators. While the regional intensity of competition has negative effects in low-tech and medium low-tech industries, there is no or even a 455 slightly positive effect in the remaining industry groups. It follows that regional competition is rather unimportant for most of the export-oriented and globally active industries. By contrast, the phenomena of ruinous competition seems to come to bear in rather domestically oriented, low-tech industries.

Large sectoral differences are also found for the diversity variables. A large intra-sectoral (VR) 460 and inter-sectoral (VUR) variety has a positive impact on firms in low-tech industries. As these are industries with rather simple technical and less knowledge-demanding requirements, they might actually gain from knowledge created by others and then adapted for use in a much more simplistic form given their less demanding operations. At the same time, the indicators for related and unrelated variety are not significantly different from zero in the high-tech industries. This might 465 be driven by the fact that these are usually firms with very specific technologies aiming at specific niche markets. For these firms, own knowledge creation seems to be more important.

Summing up, the results in Table 2 confirm that agglomeration externalities are relevant factors for the firms' TFP development. Yet, they also show that R&D remains an important driver for TFP. Moreover, the results reveal and highlight that agglomeration externalities have distinctly 470 different effects for companies in different industries and there is not just the one - positive or

negative - effect of a specific agglomeration externality.

Table 3: Elasticities of production inputs and variables in the law of motion; technology groups; regions: 10km radii

	(1)	(2)	(3)	(4)	(5)
	low-tech	medium low-tech	medium high-tech	high-tech	All
	production function $f(\cdot)$ – output elasticities ( $\hat{\beta}$ )				
labour	0.628*** (0.01327)	0.729*** (0.01091)	0.768*** (0.01019)	0.892*** (0.02437)	0.693*** (0.00588)
capital	0.427*** (0.0116)	0.303*** (0.00848)	0.291*** (0.00989)	0.154*** (0.02195)	0.361*** (0.00544)
	law of motion $g(\cdot)$ – elasticities regarding TFP ( $\hat{\gamma}$ )				
TFP <sub><i>t</i>-1</sub>	0.98*** (0.00179)	0.948*** (0.00273)	0.94*** (0.00302)	0.884*** (0.00859)	0.956*** (0.00137)
R&D	0.0017** (0.0008)	0.0033*** (0.00085)	0.0078*** (0.00105)	0.0303*** (0.00346)	0.006*** (0.00047)
URB	0.000 (0.0004)	0.0028*** (0.00048)	0.0013* (0.00065)	-0.0013 (0.00189)	0.0017*** (0.00028)
LOC	0.0003 (0.00031)	0.0011*** (0.0004)	0.0016*** (0.00048)	0.0073*** (0.00147)	-0.0001 (0.00021)
POR	-0.0007** (0.00029)	-0.0015*** (0.00032)	0.0006* (0.00035)	0.0007 (0.00109)	0.0000 (0.00018)
VUR	-0.0004 (0.00046)	-0.0008 (0.00056)	-0.0011 (0.00079)	-0.0002 (0.00219)	-0.0011*** (0.00033)
VR	0.0012*** (0.0003)	0.0003 (0.00034)	-0.0011*** (0.00037)	-0.0041*** (0.00114)	0.0003* (0.00019)
Constant	0.115*** (-0.0105)	0.374*** (-0.0198)	0.44*** (-0.0226)	1.118*** (-0.0814)	0.276*** (0.0092)
N	16,000	14,825	14,941	2,559	48,325

Notes: \* significance at  $p < .1$  level, \*\* significance at  $p < .05$  level, \*\*\*Significance at  $p < .01$  level. East/West, year, industry and legal form are controlled for in the first stage of ACF procedure. Hansen-Test results are not reported as the  $p$ -value is always close to 1, as it needs to be the cases in GMM specifications with as many instruments as variables.

As pointed out before, the results shown so far might be affected by the chosen regional definition. The analysis, therefore, not only distinguishes between sectors, but also makes use of different spatial delineations. Table 3 shows the estimation results when regions are demarcated by the distances, here by a radius of 10 kilometers. Overall, the results using radii do not differ much from those using a functional-administrative spatial delineation. Urbanization and localization advantages have, if statistically significant, the theoretically expected positive effect on TFP. In fact, the comparison with Table 2 reveals that the effects of localization is larger for high-tech and medium high-tech industries, suggesting that a higher concentration of similar firms in the immediate vicin-



480 ity is particularly important for the localization effects to come into effect. This supports the idea of setting up clusters, especially for high-tech industries. At the same time, unrelated variety is barely detectable in the different industry groups. Hence, a large variety of different industries in close proximity to a firm is not relevant for its TFP. The coefficients of the remaining variables are mostly in the same direction and range as in Table 2. Overall, the comparison between Table 3  
485 and Table 2 reveals that different regional definitions are not decisively changing the results. Yet, it also shows that regional proximity affects the strength of the effects.

### 5.2. Robustness

The use of different regional definitions serves as a first robustness check. The similar results confirm the general notion that agglomeration economies affect the TFP of firms, but they also  
490 highlight that these effects differ between industries. Additional robustness checks include the separate estimations for SMEs and large firms as well as the use of a different radius.

Figure 1 compares the coefficients of each variable in function  $g(\cdot)$  for estimations conducted with the full sample (All), a sub-sample for firms with less than 100 employees (SME) and a sub-sample for firms with more than 100 employees (LE).<sup>19</sup> In these three estimations, the regions are  
495 defined as ROR. In addition, the figure contains the coefficients of estimations in which the regions are defined according to radii (10 km and 25 km). For reasons of space, Figure 1 only contains the coefficients for estimations that include all industries. The estimation results for each industry group are shown in Figure A.3 to Figure A.6. The dots in the figures represent the point estimates, while the 1.65 times the standard deviations defines the whiskers.<sup>20</sup> These whiskers show whether  
500 two coefficients are different from each other or from zero.

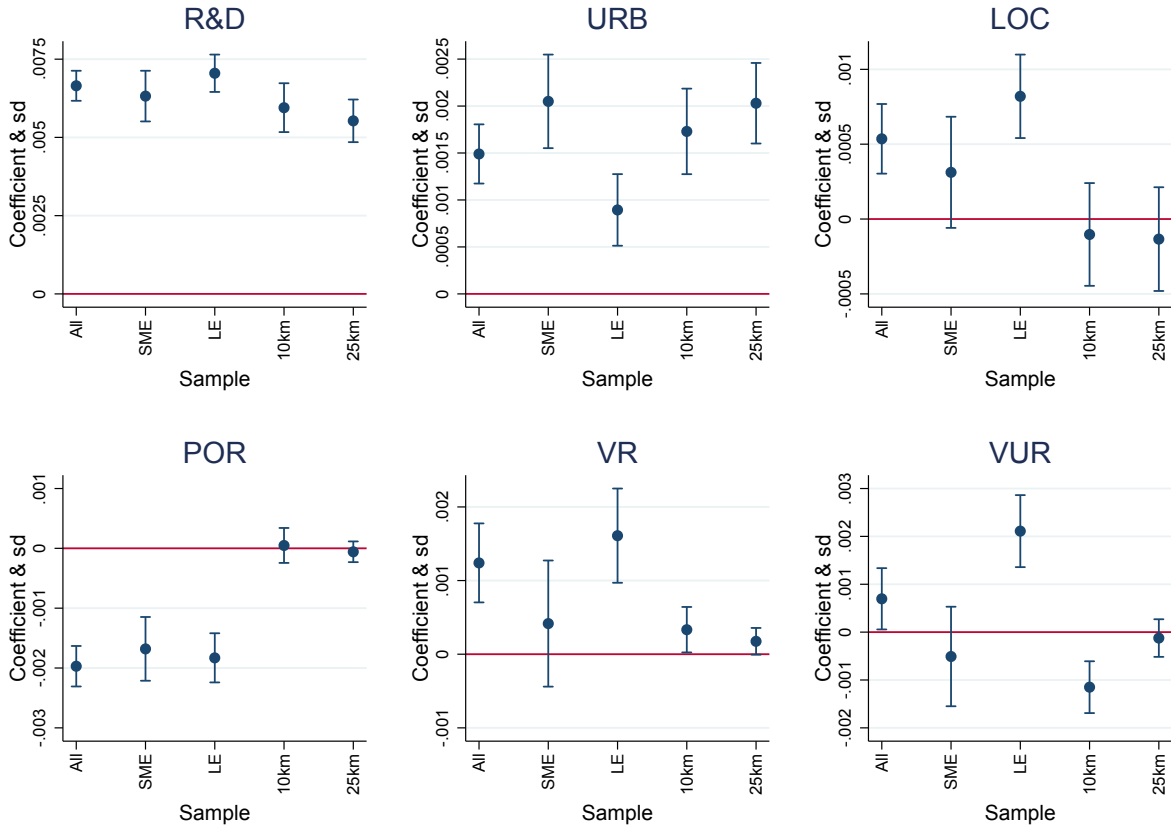
As the first panel in Figure 1 shows, the coefficients of R&D are always significant and positive with similar magnitudes in the different estimations. The coefficients for urban economies are also always positive and significant. Yet, the effect is stronger for SMEs compared to large firms. The coefficient of localization economies is insignificant in the subsample of SME and when radii are  
505 used to define the regions. However, as already discussed when comparing the results in Table 3 and Table 2, this is only the case in estimations for the entire manufacturing industry (column 5).

---

<sup>19</sup>The threshold is set such that the two subs-samples are more or less of equal size. Applying the Eurostat thresholds for SMEs and large firms was not meaningful. The numbers for large companies was too low in many industries, resulting in instable estimation, non-convergence of GMM and similar issues.

<sup>20</sup>The multiplier is chosen such that 90% of all observations lie within the range of  $\bar{x} \pm 1.65 \cdot \sigma_x$ . Accordingly, values outside this range are not drawn from the distribution defined by  $\bar{x}$  and  $\sigma_x$  with a probability of 10%.

Figure 1: Point estimates and 1.65 of sd; manufacturing industry; regions: ROR & radii



The coefficient for localization is positive and mostly significant in estimations for industry groups. This holds for different size definitions as well as spatial definitions as revealed in Figure A.3 to Figure A.6. The same reasoning applies to the coefficients of the Porter variable. In Figure 1, the Porter variable is negative and significant in all estimations based on ROR, but insignificant if radii are used. At the industry group level, this difference between the spatial definitions is no longer as pronounced as in Figure 1. Apart from the medium-high-tech industries, the coefficients of the Porter variable are either all negative and significant or all insignificant (see Figure A.3 to Figure A.6).

The coefficient of related variety in Table 2 is significant only because of its effect on large firms. As Figure 1 reveals, the coefficient is no longer significantly different from zero for SMEs. In contrast, for different regional definitions, the coefficient remains to be positive and significant,

which supports the results in Table 2. The coefficient for unrelated variety, on the other hand, is less stable. This is also the case when comparing the coefficients in estimations for separate industries (see Figure A.3 to Figure A.6).

Summing up, by and large the robustness checks confirm the main results. In addition, they highlight two points: First, if there are differences, then they are most pronounced between estimations for different spatial definitions. This indicates that agglomeration economies are not strictly the same given different regional proximity. This supports the findings of Cainelli and Ganau (2018). Secondly, while agglomeration economies have slightly different effects for SMEs, these differences are often not very pronounced. SMEs, thus, benefit in the same way as large firms from agglomeration economies.

### 5.3. Limitations

Studies on agglomeration economies face several econometric issues and limitations, some of them directly related to the use of regionally defined variables, while other are more general. These are, mainly: prices biases, selection of surviving firms, reverse causality, and the chosen regional unit. These are briefly discuss in what follows.

The omitted price bias is an issue common to all production function estimations that make use of deflated monetary values instead of physical quantities. If input or output prices differ between firms, but industry price series are applied for deflating monetary variables, as it is commonly done, the estimations are exposed to the risk of the *input price bias* or the *output price bias*. See Klette and Griliches (1996), Foster et al. (2008), Collard-Wexler and Loecker (2016), Grieco et al. (2016), and De Loecker et al. (2016) on this issues. Our analysis is only partially confronted with the problem that comes from different input prices because the input variables used in  $f(\cdot)$  or  $g(\cdot)$  are, apart from capital, non-monetary.<sup>21</sup> The output price bias stems from unobserved differences in output prices between firms. This affects the error term and can lead to biased coefficient if inputs are correlated with these price differences. Given that input and output price information at the firm-level is missing in our data, we cannot rule out that our estimations are affected by either or both of these issues. Following De Loecker and Goldberg (2014), if capital is indeed subject to the input price bias, the respective coefficient would be most likely underestimated and, consequently,

---

<sup>21</sup>Material is less of an issue because its coefficient is not estimated due to the fact that the analysis relies on a value-added production function. Material is only used in the proxy function, which is not designed to provide consistent coefficients and is not used for calculating the TFP or for interpretation.

TFP would be overestimated. In this case, the coefficients of the variables in function  $g(\cdot)$  might be upward biased. Yet, the direction of the output price bias is unclear and can well offset the input price bias (De Loecker, 2011; Foster et al., 2008). Moreover, the capital and labor coefficients show reasonable magnitudes, with the capital coefficients rather at the upper end of comparable estimations in the literature. Hence, while we cannot explicitly rule out that our analyses are affected by price biases, there is little indication that this issue is so massive that our main results are no longer valid. Another general problem is the selection bias, which results from the fact that only surviving companies are observed (Olley and Pakes, 1996). As shown by Levinsohn and Petrin (2003), however, the selection bias is a relevant issue in balanced panel but not in unbalanced panels. Because making use of the latter, we do not consider selection bias to be a relevant issue in our analysis.

Reverse causality addresses the question of whether higher TFP can positively affect agglomeration characteristics. It is, in other words, a question of whether the relationship between productivity and agglomeration economies is strictly unidirectional. While Graham et al. (2010) find support for a bidirectional relationship, Cainelli et al. (2015) find no empirical evidence that TFP affects localization economies etc. We argue that TFP could only affect regions in the longer run, which is why our analysis should be less affected by this problem.

Finally, the chosen geographical unit is another well-known issue of studies on agglomeration economies. Our analysis makes use of different regional definitions to overcome it. As the previous section proves, our main results are not affected by it. However, the analysis face the limitation that the data only contain the coordinates of the municipalities. Municipalities which partly belong within a radius are ignored as long as the coordinate, which usually refers to the center of a municipality, is not within the radius. Thus, our radii based spatial definition is imperfect.

## 6. Conclusion

Using data from German firms in the manufacturing sector, this study analyzes the effect of agglomeration externalities on firm-level total factor productivity (TFP). The analysis is conducted separately for industry groups that are defined by their technological intensity in order to allow for non-uniform effects of agglomeration economies. Furthermore, R&D activities and past TFP are allowed to affect current TFP. The econometric approach overcomes the hitherto widely used procedure of separating the estimation of the production function, in order to obtain TFP, from

the estimation of the effect of agglomeration externalities on TFP. We discuss the econometric issue resulting from such an approach and propose the use of a control function approach that simultaneously estimates output elasticities as well as the effects of several agglomeration externalities on TFP.

580 The results reveal that TFP, as the bundle of all past experiences and firm characteristics, only changes slowly. This is reflected in the strong persistence of productivity. However, the more dynamic and demanding the technological environment firms operate in, the more dynamics one sees in TFP. Consequently, persistence is most pronounced in low-tech manufacturing industries and the least pronounced in the high-tech manufacturing industries. Furthermore, the results 585 confirm the positive effect of R&D, thus confirming the importance of the firms' own efforts for increasing their TFP. In line with the findings for persistence, the effect of R&D increases with the technological intensity of industries.

Despite the importance of firm inherent aspects, the analysis confirms that agglomeration economies remain important for TFP. The analysis also reveals, though, that the effect differs 590 strongly between firms in industries with different technological intensities. Urban economies are found to have the largest effect on TFP for firms in the high-tech industries. TFP of firms in low-tech industries, on the other hand, are not affected at all by urban economies. For firms in the latter industries, however, the variety of the local economic structure has an impact, while this is irrelevant for the TFP of firms in high-tech industries. The effect of stronger competition is also 595 not equal for all firms. In low-tech and medium low-tech industries, stronger competition seems to unfold negative effects. Only localization economies have positive and significant effect on TFP throughout, but the effect increases with the technological intensity of industries.

These findings should be seen in the light of the potentials and limitations of regional economic policies. They show that the level of TFP in firms is affected not just by R&D, which is supported 600 through various subsidy programs at national level, but also the regional environment. Thus, regional economic policies can provide additional support and create a favorable environment. *Inter alia*, cluster policies are an instrument increasingly used for this purpose in regional economic policies in the US and in Europe (Delgado et al., 2010; Falck et al., 2010; Sternberg et al., 2010; Duranton, 2011). Our industry-specific results suggest that supporting the development of spatially 605 concentrated high-tech clusters is especially promising. Firstly, firms in these industries benefit the most from the external effects of agglomeration of firms of the same industry in a region

(localization benefits). Secondly, additional positive effects can be exploited if the high-tech clusters are located in densely populated regions, since this type of industry also benefits strongly from general urbanization advantages such as education and research infrastructure.

610 Yet, manufacturing encompasses more than just high-tech manufacturing. In Germany, just over 10% of value added in manufacturing is generated by high-tech industries and just under 40% by medium-high-tech industries (Gornig and Schiersch, 2016). Consequently, almost half of the value added is generated in low-tech and medium-low-tech manufacturing industries. Moreover, countries such as Germany might be able to exploit the re-industrialization potential of low-tech  
615 industries as a result of digitization (Koren, 2010). Our analysis shows that spatial effects stemming from related variety positively affect the TFP of firms in low-tech industries. Consequently, regional development strategies aiming at these industries should try to bring a corresponding broadness into such clusters.

### Acknowledgments

620 We gratefully acknowledge the funding received by the Hans-Böckler-Stiftung as part of the research project "Industry in the City". We would also like to thank the Advisory Board of the project for the valuable suggestions and hints. Furthermore, we gratefully acknowledge the extensive support of the team at the Research Data Centre of the Federal Statistical Office Berlin-Brandenburg.

## References

- 625 Akerberg, D., Benkard, C., Berry, S., Pakes, A., 2007. Econometric tools for analyzing market outcomes. In: Heckman, J. J., Leamer, E. E. (Eds.), *Handbook of Econometrics*. Vol. 6A. Amsterdam: North-Holland, Ch. 63, pp. 4171–4276.
- Akerberg, D., Caves, K., Frazer, G., 2006. Structural identification of production functions, MPRA Paper No. 38349.
- 630 Akerberg, D. A., Caves, K., Frazer, G., 2015. Identification properties of recent production function estimators. *Econometrica* 83 (6), 2411–2451.
- Ahrend, R., Farchy, E., Kaplanis, I., Lembcke, A. C., 2017. What makes cities more productive? Agglomeration economies and the role of urban governance: Evidence from 5 OECD countries. *OECD Productivity Working Papers*, No. 6.
- 635 Antonelli, C., Patrucco, P. P., Quatraro, F., 2011. Productivity growth and pecuniary knowledge externalities: An empirical analysis of agglomeration economies in European regions. *Economic Geography* 87 (1), 23–50.
- Arrow, K. J., 1962. *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton University Press, Ch. Economic welfare and the allocation of resources for invention, 640 p. 609–626.
- Aw, B. Y., Roberts, M. J., Xu, D. Y., 2011. R&D investment, exporting, and productivity dynamics. *American Economic Review* 101 (4), 1312–44.
- Basile, R., Pittiglio, R., Reganati, F., 2017. Do agglomeration externalities affect firm survival? *Regional Studies* 51 (4), 548–562.
- 645 Beugelsdijk, S., Klasing, M. J., Milionis, P., 2018. Regional economic development in Europe: The role of total factor productivity. *Regional Studies* 52 (4), 461–476.
- Blundell, R., Bond, S., 1998. Initial conditions and moments restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.
- Cainelli, G., Fracasso, A., Marzetti, G. V., 2015. Spatial agglomeration and productivity in Italy: 650 A panel smooth transition regression approach. *Papers in Regional Science* 94 (Supplement 1), S40–S67.
- Cainelli, G., Ganau, R., 2018. Distance-based agglomeration externalities and neighbouring firms' characteristics. *Regional Studies* 52 (7), 922–933.
- Caragliu, A., de Dominicis, L., de Groot, H. L., 2016. Both Marshall and Jacobs were Right! 655 *Economic Geography* 92 (1), 87–111.
- Collard-Wexler, A., Loecker, J. D., 2016. Production function estimation with measurement error in inputs. NBER Working Paper No. 22437.

- Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63 (2), 723–742.
- 660 Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2012. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80 (6), 2543–2594.
- De Loecker, J., 2011. Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica* 79 (5), 1407–1451.
- De Loecker, J., 2013. Detecting learning by exporting. *American Economic Journal: Microeconomics* 5 (3), 1–21.
- 665 De Loecker, J., Goldberg, P. K., 2014. Firm performance in a global market. *Annual Review of Economics* 6 (1), 201–227.
- De Loecker, J., Goldberg, P. K., Khandelwal, A. K., Pavcnik, N., 2016. Prices, markups, and trade reform. *Econometrica* 84 (2), 445–510.
- 670 De Loecker, J., Warzynski, F., 2012. Markups and firm-level export status. *American Economic Review* 102 (6), 2437–71.
- Delgado, M., Porter, M. E., Stern, S., 2010. Clusters and entrepreneurship. *Journal of Economic Geography* 10 (4), 495–518.
- DiGiacinto, V., Gomellini, M., Micucci, G., Pagnini, M., 2014. Mapping local productivity advantages in Italy: Industrial districts, cities or both? *Journal of Economic Geography* 14, 365–394.
- 675 Doraszelski, U., Jaumandreu, J., 2013. R&D and productivity: Estimating endogenous productivity. *The Review of Economic Studies* 80, 1338–1383.
- Duranton, G., 2011. California dreamin’: The feeble case for cluster policies. *Review of Economic Analysis* 3 (1), 3–45.
- 680 Duranton, G., Puga, D., 2004. Microfoundations of urban agglomeration economies. In: Henderson, J. V., Thiss, J.-F. (Eds.), *Handbook of Regional and Urban Economics*. Vol. 4. North-Holland, Amsterdam.
- Faberman, J. R., Freedman, M., 2016. The urban density premium accross establishments. *Journal of Urban Economics* 93, 71–84.
- 685 Falck, O., Heblich, S., Kipa, S., 2010. Industrial innovation: Direct evidence from a cluster-oriented policy. *Regional Science and Urban Economics* 40 (6), 574–582.
- Foster, L., Haltiwanger, J., Syverson, C., 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review* 98 (1), 394–425.
- Frenken, K., Oort, F. V., Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. *Regional Studies* 41, 685–97.
- 690



- Fritsch, M., Görizig, B., Hennchen, O., Stephan, A., 2004. Cost structure surveys for Germany. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften* 124 (4), 557–566.
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., Shleifer, A., 1992. Growth in cities. *Journal of Political Economy* 100 (6), 1126–1152.
- Gornig, M., Schiersch, A., 2016. Weak investment poses a threat to industry in Europe. *Intereconomics* 51 (5), 272–277.
- Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P., Noland, R. B., 2010. Testing for causality between productivity and agglomeration economies. *Journal of Regional Science* 50 (5), 935–951.
- Grieco, P. L., Li, S., Zhang, H., 2016. Production function estimation with unobserved input price dispersion. *International Economic Review* 57 (2), 665–690.
- Griliches, Z., 1979. Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics* 10, 92–116.
- Harris, R., Moffat, J., 2015. Total factor productivity growth in local enterprise partnership regions in Britain, 1997–2008. *Regional Studies* 96 (6), 1019–1041.
- Henderson, J. V., 1986. Efficiency of resource usage and city size. *Journal of Urban Economics* 19 (1), 47–70.
- Hervas-Oliver, J.-L., Sempere-Ripoll, F., Alvarado, R. R., Estelles-Miguel, S., 2018. Agglomerations and firm performance: Who benefits and how much? *Regional Studies* 521 (3), 338–349.
- Holl, A., 2016. Highways and productivity in manufacturing firms. *Journal of Urban Economics* 93, 131–151.
- Isard, W., 1956. *Location and Space-Economy; a General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure*. Technology Press of Massachusetts Institute of Technology, Wiley.
- Jacobs, J., 1969. *The Economy of Cities*. New York: Vintage.
- Kancs, A., Siliverstovs, B., 2016. R&D and non-linear productivity growth. *Research Policy* 45, 634–646.
- Klette, T. J., Griliches, Z., 1996. The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics* 11 (4), 343–361.
- Koch, A., 2007. Neue Datenquelle „Unternehmensregister“: Mehr Informationen über den Mittelstand ohne neue Bürokratie. Institut für Mittelstandsforschung. Bonn.
- Konold, M., 2007. New possibilities for economic research through integration of establishment-level panel data of German official statistics. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften* 127 (2), 321–334.

- 725 Koren, Y., 2010. *The Global Manufacturing Revolution: Product-Process-Business Integration and Reconfigurable Systems*. Wiley.
- Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70 (2), 317–341.
- Maré, D. C., Graham, D. J., 2013. Agglomeration elasticities and firm heterogeneity. *Journal of Urban Economics* 75, 44–56.
- 730 Marschak, J., Andrews Jr., W. H., 1944. Random simultaneous equations and the theory of production. *Econometrica* 12 (3/4), 143–205.
- Marshall, A., 1920. *Principles of Economics*. Macmillan and Co., Ltd.
- Martin, P., Mayer, T., Mayneris, F., 2011. Spatial concentration and plant-level productivity in France. *Journal of Urban Economics* 69, 182–195.
- 735 Matano, A., Naticchioni, P., 2012. Wage distribution and the spatial sorting of workers. *Journal of Economic Geography* 12 (2), 379–408.
- Melo, P. C., graham, D. J., Levinson, D., Aarabi, S., 2017. Agglomeration, accessibility and productivity: Evidence for large metropolitan areas in the US. *Urban Studies* 54 (1), 179–195.
- 740 Melo, P. C., Graham, D. J., Noland, R. B., 2009. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics* 39, 332–342.
- Møen, J., Thorsen, H. S., 2015. Publication bias in the returns to R&D literature. *Journal of the Knowledge Economy* 8 (3), 987–1013.
- OECD, 2012. *Redefining “Urban”: A New Way to Measure Metropolitan Areas*. OECD Publishing.
- 745 Olley, G. S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64 (6), 1263–1297.
- Openshaw, S., 1984. The modifiable areal unit problem. *Concepts and techniques in modern geography* 38.
- Porter, M. E., 1990. *The Competitive Advantage of Nations*. Free Press.
- 750 Richter, P. M., Schiersch, A., 2017. CO<sub>2</sub> emission intensity and exporting: Evidence from firm-level data. *European Economic Review* 98, 373–391.
- Romer, P. M., 1986. Increasing returns and long-run growth. *The Journal of Political Economy* 94 (5), 1002–1037.
- Rosenthal, S. S., Strange, W. C., 2004. Evidence on the Nature and Sources of Agglomeration Economies. Vol. 4. Elsevier, Ch. 49, pp. 2243–2291.
- 755

Sternberg, R., Kiese, M., Stockinger, D., 2010. Cluster policies in the US and Germany: Varieties of capitalism perspective on two high-tech states. *Environment and Planning C: Government and Policy* 28 (6), 1063–1082.

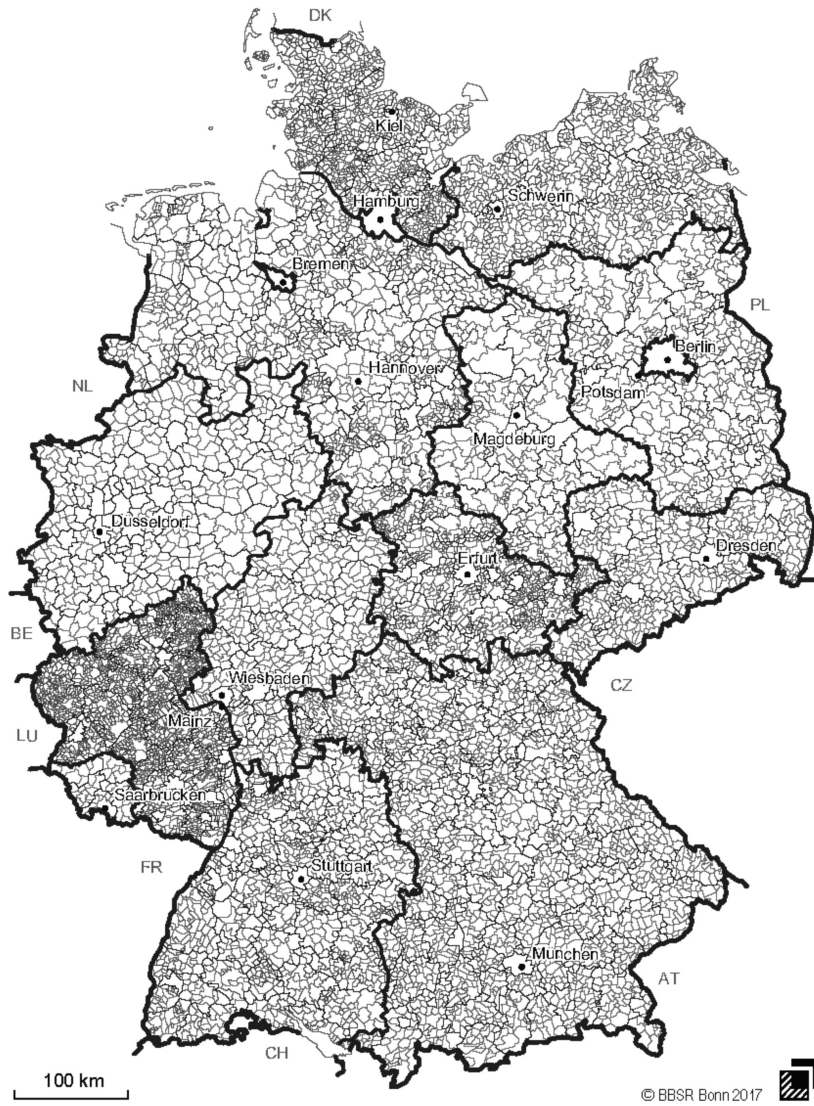
760 Ugur, M., Trushin, E., Solomon, E., Guidi, F., 2016. R&D and productivity in OECD firms and industries: A hierarchical meta-regression analysis. *Research Policy* 45.

UN, 2008. International standard industrial classification of all economic activities - revision 4.

Weber, A., 1929. *Theory of the Location of Industries*. Chicago University Press.

Wooldridge, J. M., 2009. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters* 104 (3), 112–114.

Figure A.1: Municipalities in Germany, 2015



Source: BBSR

Figure A.2: ROR in Germany, 2015



Source: BBSR

Figure A.3: Low-tech industries; point estimates and 1.65 of sd; regions: ROR & radii

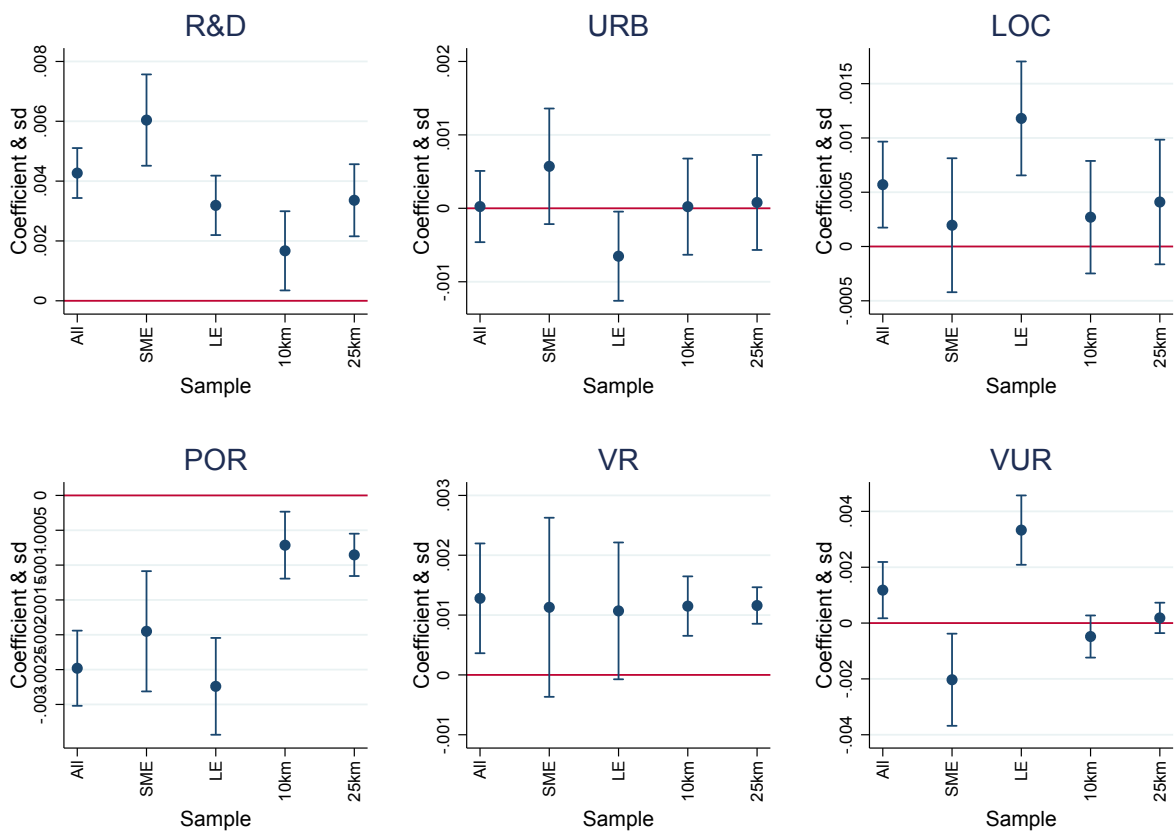


Figure A.4: Medium low-tech industries; point estimates and 1.65 of sd; regions: ROR & radii

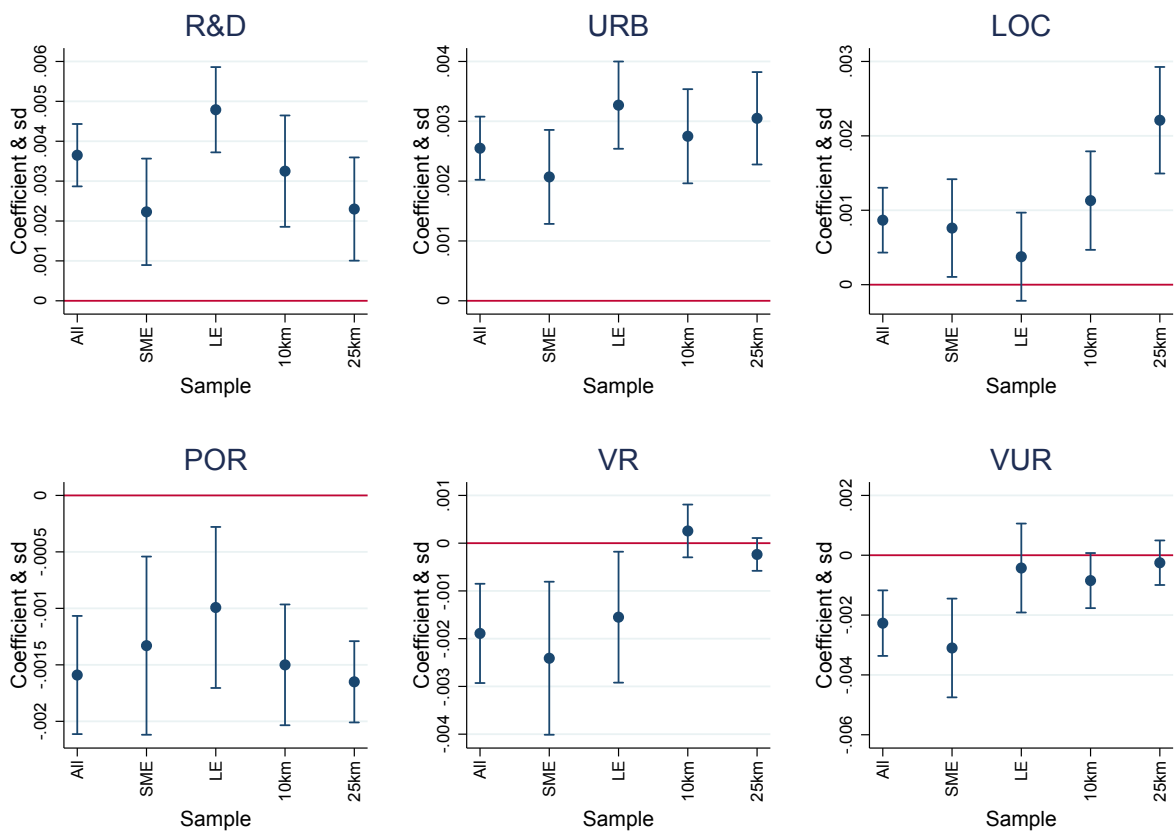


Figure A.5: Medium high-tech industries; point estimates and 1.65 of sd; regions: ROR & radii

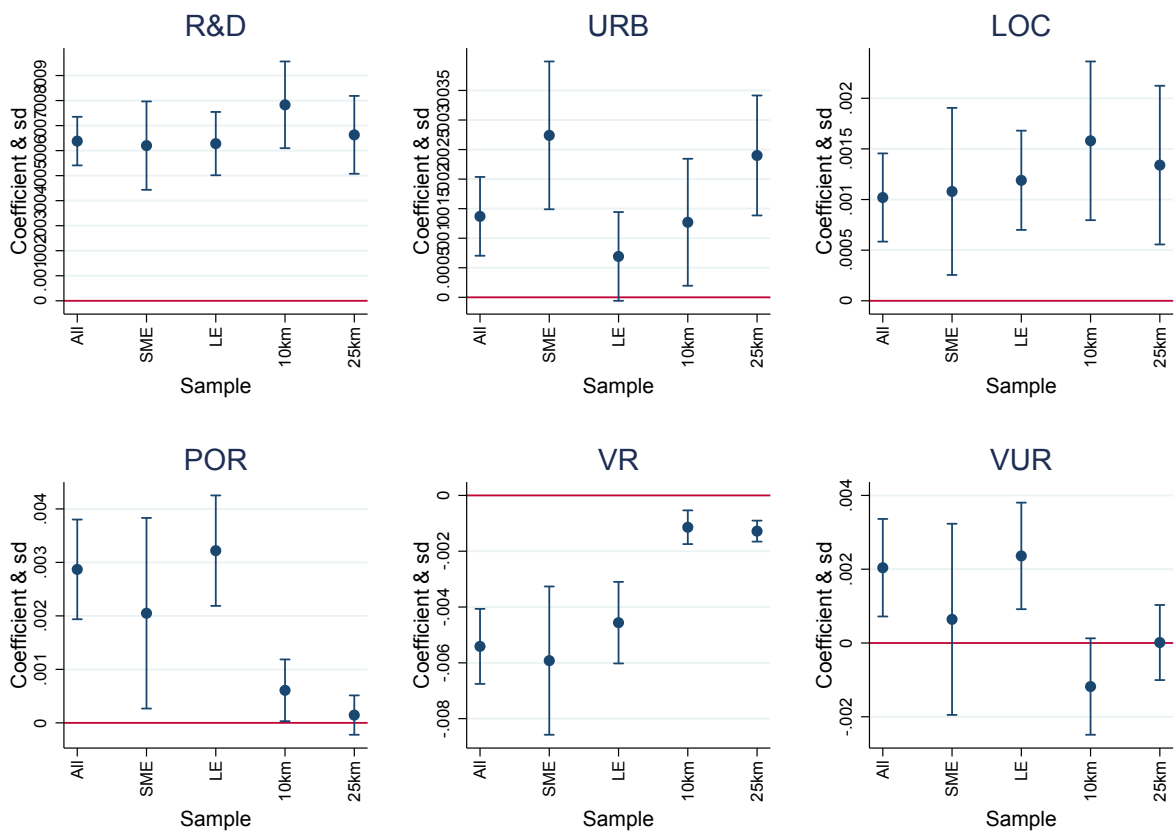
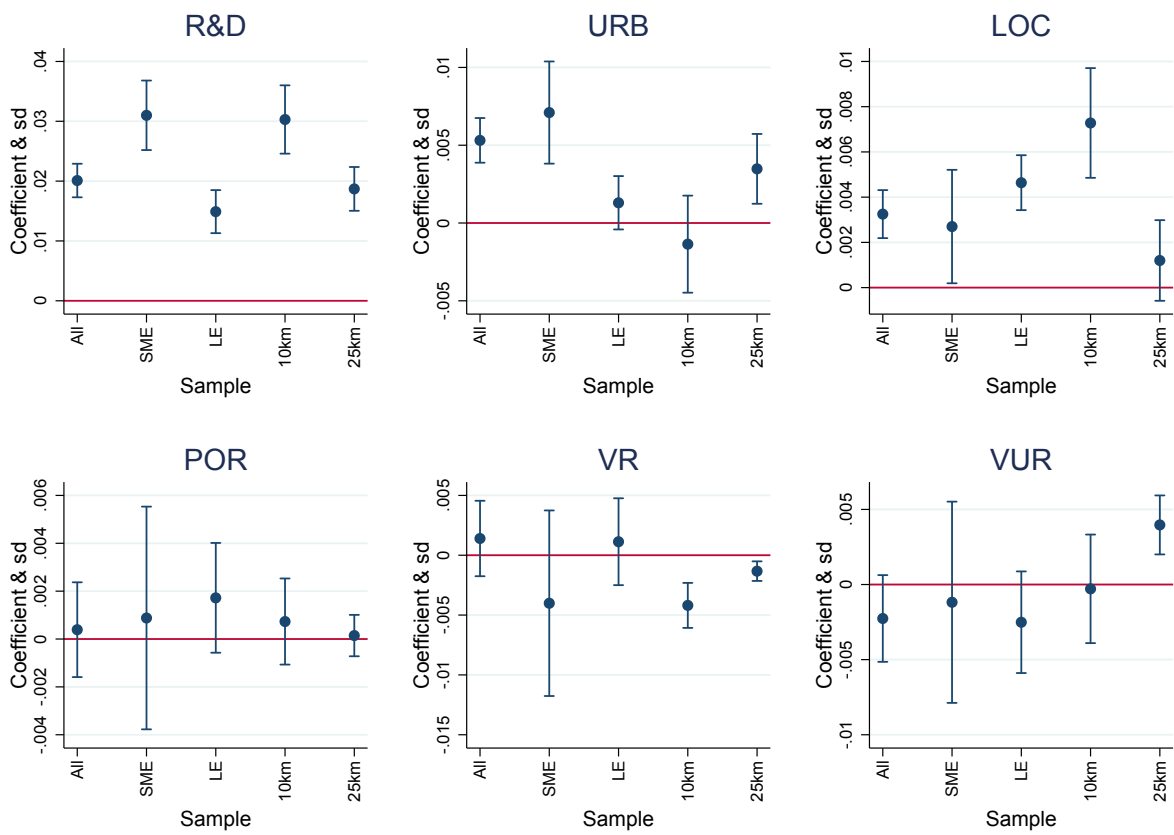




Figure A.6: High-tech industries; point estimates and 1.65 of sd; regions: ROR & radii



## Appendix B Tables

Table B.1: Sector codes

Industry	Industry Code ISIC Rev. 4	Technology group
Manufacturing	10t33	-
Manufacturing of food products	10	low-tech
Manufacturing of beverages	11	low-tech
Manufacturing of textiles	13	low-tech
Manufacturing of wearing apparel	14	low-tech
Manufacturing of leather & related prod.	15	low-tech
Manufacturing of wood & prod. of wood etc.	16	low-tech
Manufacturing of paper & paper prod.	17	low-tech
Printing & reprod. of recorded media	18	low-tech
Manufacturing of chemicals & chemical prod.	20	medium high-tech
Manufacturing of basic pharm. prod. & pharm. prep.	21	high -tech
Manufacturing of rubber & plastics prod.	22	medium low-tech
Manufacturing of other non-metallic mineral prod.	23	medium low-tech
Manufacturing of basic metals	24	medium low-tech
Manufacturing of fabricated metal prod., exc. mach. & equip.	25	medium low-tech
Manufacturing of computer, electronic & optical prod.	26	high -tech
Manufacturing of electrical equipment	27	medium high-tech
Manufacturing of machinery & equipment	28	medium high-tech
Manufacturing of motor vehicles, trailers & semi-trailers	29	medium high-tech
Manufacturing of furniture	31	low-tech
Other Manufacturing	32	low-tech
Repair & installation of machinery & equipment	33	medium low-tech

Source: [UN \(2008\)](#).

Table B.2: Number of observation per dataset

year	CS	IC	AR	CS&IS	Total
2003	16,728	37,714	40,036	15,813	41,110
2004	16,099	37,680	39,694	15,522	40,266
2005	15,420	37,015	38,941	14,920	39,500
2006	14,884	36,985	38,191	14,515	39,007
2007	14,532	37,255	38,346	14,265	38,658
2008	17,304	37,815	38,822	16,818	39,358
2009	16,735	36,338	37,539	16,341	38,006
2010	16,226	35,979	37,044	15,903	37,432
2011	15,766	36,277	37,163	15,523	37,485
2012	17,148	36,771	37,638	15,696	39,172
2013	16,639	37,001	37,784	15,685	38,851
2014	16,187	37,054	37,688	15,452	38,552
, Total	193,668	443,884	458,886	186,453	467,397

Source: *AFiD Panel Manufacturing Firms*.