WIDER Working Paper 2019/93

Agglomeration and productivity in South Africa

Evidence from firm-level data

Hammed Amusa,[1] Njeri Wabiri,[2] and David Fadiran[3]

November 2019

Abstract: Using comprehensive, anonymized tax administrative data for the 2008–14 period, we examine firm-level productivity in South Africa. Measures of firm-level productivity are included in a spatial autoregressive model that assesses spillovers from total factor productivity originating from agglomeration economies and the spatial diffusion of productivity shocks. We find that across South Africa's firms, intermediate inputs have the highest impact on firm productivity. The results from the spatial analysis indicate that for a firm in a particular region, its clustering with other firms, having increased market power, and an extended length of stay in a particular region have a greater impact on productivity than do market conditions and firm-specific characteristics associated with firms located in neighbouring regions or municipalities.

Keywords: agglomeration, productivity, South Africa

JEL classification: C23, D24, R11, R12

---

[1] African Development Bank, Abidjan, Côte d'Ivoire; [2] Human Sciences Research Council, Pretoria, South Africa; [3] School of Economics, University of Cape Town, South Africa. Corresponding author: h.amusa@afdb.org

This study has been prepared within the UNU-WIDER project Southern Africa—Towards Inclusive Economic Development (SA-TIED)

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-729-3

https://doi.org/10.35188/UNU-WIDER/2019/729-3

Typescript prepared by Gary Smith.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Beginning in the 1980s, socio-economic and political imperatives in South Africa led to a gradual breakdown of the restrictive anti-urban regime. Following South Africa's democratic transition in 1994, the country slowly recovered from the tough controls on urbanization and resumed its urban transition. Freedom of movement contributed to rapid urbanization that was largely driven by rural–urban migration that caused the population of black South Africans in urban areas to increase from 10.3 million in 1980 to over 20 million by 2000 (Louw 2003). As with other parts of the world, the main driver of rural—urban migration in post-1994 South Africa has been differences in economic opportunity. South Africa's cities are areas of concentration of economic activities and have consistently outperformed the rest of the country in terms of economic growth. This performance can be attributed to 'agglomeration economies' that allow for the exploitation of economic assets—including sharing of infrastructure, services, and information, well-connected infrastructure and a relatively better educated workforce (Turok 2012).

In the mid-2000s, the increasingly popular notion of placing South Africa as a developmental state led to the initiation of a number of national projects focused on socio-economic need and targeted at specific geographical economies within South Africa. Within these initiatives, government aimed to facilitate significant investments in urban economies and secondary cities, with the objective of addressing economic growth and unemployment. The launch of the National Development Plan (NDP) in 2012 gave additional impetus to the government's plans to overcome constraints to maximizing the potential of urban economies. In 2012, the National Treasury initiated the City Support Programme (CSP) to address spatial inequalities and development challenges at the city level. In collaboration with key government departments, the project aims to harness human, financial, and institutional capacity to improve planning, implementation, and innovation in order to ensure that cities are more inclusive and liveable, productive, and environmentally, financially, and economically efficient, while also being resilient to shocks. Similarly, with the seventh iteration of the Industrial Policy Action Plan (IPAP; 2015/16–17/18), South Africa plans to maximize the opportunities provided by the linkages and multipliers that exist between its key labour-absorbing economic sectors in order to (1) extract full value from South Africa's enormous resource endowment, and (2) ramp up export competitiveness. Included in this objective is a need to promote increased diversity of South Africa's industrial sector and for such diversity to enhance employment opportunities in an environment of rapid urbanization (Amusa and Ngozo 2017).

The strategies described above highlight the growing recognition by policymakers that accelerated urbanization and the structure of economic activity within South Africa's urban centres could help drive rapid economic growth and reduce the burdens of poverty and inequality. However, for this to become reality, it is imperative that implementing authorities have a full understanding of the nature of dynamic interactions between economic activities across major urban centres and outlying areas, and the extent to which the diversity or concentration of industrial activity enhances regional economic performance. Initiatives such as the CSP and IPAP assume that the economic development of urban economies will occur via state intervention aimed at addressing the apartheid spatial economic structures. However, events such as the global recession of 2008 and subsequent loss of markets and employers in the vulnerable sectors of mining and manufacturing are a reminder of the need for proposed interventions to be sensitive to how geographical dynamics of production and location of firms, as well as the structure of economic activities within urban areas, influence regional economic growth and development, and also to have a full understanding of the nature of dynamic interactions between economic activities across major urban centres and outlying areas, and the extent to which the diversity or concentration of industrial activity enhances regional productivity.

Empirical studies have identified agglomeration economies—the externality arising from spatial concentration of economic activities—as a key determinant of regional productivity gaps (e.g. Combes et al.

2012). Despite the urbanization trend in South Africa and the potential for agglomeration economies, research examining the relationship between agglomeration and productivity across South Africa's regions remains thin. Most empirical studies that have examined the industry location effects have largely focused on the determinants of output growth at the level of industries and regions (e.g. Fedderke and Wollnik 2007; Naudé and Krugell 2006b), or assessed whether location can be used to explain the differences in firm efficiency (e.g. Krugell and Rankin 2012). In this regard, the main contribution of this paper is to fill a gap in the literature by examining the relationship between agglomeration and productivity using firm-level data sourced from administrative tax statistics compiled by the National Treasury and the South African Revenue Service (SARS). The value-added of the work carried out in this paper lies in the explicit modelling of productivity effects of firm concentration through the use of econometric techniques based on the endogenous nature of agglomeration and locational productivity, as well as spatial spillovers across area-specific borders.

Using the comprehensive firm-level dataset over the period 2008–14, this paper follows an empirical approach similar to Ding et al. (2016) and Sangalli and Lamieri (2015), and examines the agglomeration–productivity nexus in two steps. First, estimates of total factor productivity (TFP) are derived from a standard Cobb–Douglas production function that, besides factor inputs (of labour and capital), is extended to include variables capturing a set of characteristics that allow for TFP comparisons across firms. Measures of agglomeration are included to control for the impact of location (region and/or city) effects on firm productivity.

Firm output and productivity are often location-specific, where factors such as the availability of physical infrastructure, the need to locate close to sources of raw materials, and decisions on adopting research-based technology from similar firms play important roles. Such location-specific factors make inter-linkages between firms inevitable and may cause productivity of a particular firm to be dependent on the productivity of firms located within the same geographical region. Noting the possibility of such spillover or neighbourhood effects, the second step in the analysis employs a geospatial econometric analysis to test the existence of spatial patterns in the agglomeration–productivity relationship. The spatial econometric approach is extended to include firm age as well as measures of industry concentration. Firm age is introduced to analyse whether, relative to older firms, access to better technology enhances the productivity of younger firms (vintage capital effects), or whether positive productivity gains are experienced through leaning-by-doing as a firm ages and becomes more experienced with operating and managing the production process (Harris and Moffat 2015). Both firm age and industry concentration are incorporated in the production density model in order to examine the effects of industry agglomeration (and congestion) on productivity. Additionally, with the geo-referencing of the firm-level administrative dataset, we are able to complement longitudinal data estimation techniques with a geospatial econometric approach, a framework that allows for the mapping and exploration of spatial links between regional size and firms' productivity.

To the best of our knowledge, there is no parallel work in South Africa and indeed in the rest of sub-Saharan Africa that has sought to utilize a firm-level approach in the investigation of the relationship between agglomeration and productivity across firms or regional economic activities. The focus of the present work thus represents one of the first attempts to provide information useful in answering two important research questions not currently considered in the South African literature. (1) Given South Africa's urbanization trend, are firms located in larger regions/cities more productive when controlling for attributes of individual firms? And (2) do firms benefit from spillovers generated by agglomeration of economic activities? In turn, an understanding of these questions should contribute to long–term policy decisions around (1) possible options that could involve trade-offs between the pace of short-run growth and the creation of systems that are better able to withstand economic shocks; and (2) design of appropriate spatial concentration/diversification initiatives to foster geography-specific external scale economies that enhance productivity and allow higher levels of regional development.

The rest of the paper is organized as follows: in Section 2 we provide a literature review that briefly describes the approach of extant studies to the analysis of the links between productivity and agglomeration. Studies applying spatial methodology in their empirical analysis are also highlighted. Section 3 discusses the SARS–National Treasury data and relevant variables employed in this study. Section 4 outlines the estimation procedure used to calculate baseline productivity estimates at the firm level, discusses the results, and compares productivity across firm groups and South Africa's municipalities. Lastly, Section 5 provides a way forward on the spatial econometric approach that will represent a key aspect of this study's future research.

2 Literature review

The analyses of the benefits from agglomeration economies are strongly influenced by the early works of von Thünen (1826) and Marshall (1890). Taking detailed observations on patterns of agricultural activities around cities in pre-industrial Germany, von Thünen (1826) developed the earliest location model. Employing general assumptions of perfect competition in a single market surrounded by farmland, the location model advanced the theory that agricultural production is a function of distance from the market. Based on the model, farmers surrounding the market can be expected to produce crops that have the highest market value (highest rent) in order to generate the maximum net profit (the location, or land, rent). Producing in specific areas where the locational rent is the highest would result in a system of concentric rings, with every ring specializing in different agricultural activities based on transportation costs, weight, and perishability.

More insight into the concept of geographical economies was provided by Marshall (1890), who sought to explain the concentration of cutlery and hosier industries in parts of England, France, and Germany. Marshall noted that *industrial districts*—the concentration of specialized industries in specific regions, arose from three key factors: (1) easy access to infrastructure (water) and raw materials (iron and coal deposits); (2) an available market mainly from patronage of royal courts; and (3) concerted efforts by authorities/rulers aimed at encouraging skilled labour (artisans) to settle in specific areas (Naudé 2003). The combination of these factors created pronounced geographical clusters in which the industrial structure was characterized by a skilled pool of labour (due to skills transfer from migrant/settler labour to locals), a diversity of supporting auxiliary industries (e.g. financial, technical, and transportation), and firm specialization in the various stages and processes of production. Marshall (1890) attributed the productive efficiency present in a geographic cluster to the fact that unlike single firms operating outside a particular cluster, spatial concentration allowed industries to benefit from linkages that promoted knowledge spillovers via information sharing between and among firms about innovations, new technologies, and markets (Cody 2006).

Marshall's explanation of industrial clustering and agglomeration economies was premised on observations made within a single industry cluster. The rapid industrialization of the economies of Western Europe and the United States in the first half of the twentieth century prompted efforts to explain agglomeration economies in a multi-industry and multi-location setting. In this regard, notable extensions to Marshall's notion of agglomeration economies were carried out by Ohlin (1933) and Hoover (1937, 1948), who classified agglomeration economies into two subcategories—localization economies and urbanization economies. The former describes specific economies that relate to firms engaged in similar or interlinked activities, resulting in the spatial agglomeration of related firms. In such cases of localized industry clusters, individual firms are affected by co-location with industries within the sector to which that specific firm belongs. On the other hand, urbanization economies describe the spatial concentration of economic activities involving unrelated firms and where industry agglomeration arises from either the size of a regional economy or the extent of urban concentration. Such regional economies permit diverse firms to share a range of public utilities/infrastructure, as well as specialized business and transport services, all of which are typically provided by the state or market (Malmberg et al. 2000; Parr 2000).

Following this classification, modern extensions to the theory of agglomeration economies sought to understand whether diversity or specialization of economic activities better promotes productive efficiency. Building on Marshall's (1890) idea of industry agglomeration, Arrow (1962) and Romer (1986) put forward a concept, later formalized by the seminal work of Glaeser et al. (1992) as the *Marshall-–Arrow-–Romer* (MAR) model. This model argues that proximity matters and that gains from knowledge spillovers may only be realized among same-industry concentrated firms. A key additional prediction of the MAR theory is that local monopoly is better than competition as it limits the transmission of information, thus protecting ideas and allowing for rents from innovation to be internalized by the innovator(Beaudry and Schiffauerova 2009). In contrast to the MAR model, Jacobs (1969) argues that diversity of industries within a geographic region may act as a viable mechanism for expanding the scope for innovation and economic growth. The diversity of geographically proximate industries fosters opportunities for firms to share and learn ideas and best/efficient productive practices across industries. Using the theory of comparative advantage, (Porter 1990, 1998) examines firm productivity and innovation capability with a framework in which location is the main driver of competitive advantage within the context of a globalizing economy. In this framework, the cluster is defined as a group of interconnected industries in a particular field, linked by specialized buyer–supplier relationships or connected by similar and complementary technologies or skills. Linkages within geographic clusters promote localized increasing-returns effects that translate into competitive advantages via reductions in transport costs and increases in the flow of information. In turn, these competitive advantages allow constituent firms within a cluster easier access to specialized inputs (supplies and labour), facilitating higher levels of productivity and firm capacity for innovation as well as stimulating new business formation in a manner that supports innovation and expands the cluster (Gjeslvik 2014). Based on Marshall's theorization of external economies, Krugman (1991) combines a monopolistic competition model and pecuniary externalities associated with forward–backward linkages to show that large-scale agglomeration economies are the result of an interaction between economies of scale that create increasing returns, transportation (or geographic transaction) costs, and regional market potential (Simonis 2002). In Krugman's theoretical framework, spatial agglomeration can emerge in the absence of inter-industry linkages as the co-location of firms is a feature of an endogenous process arising from technological lock-in, historical accident, or coordinated policy or political decisions. If such decisions cause a particular geographical area to become specialized in a specific type of economic activity, then potential access to greater regional markets and increasing returns to scale may stimulate more firms to locate to the area (Karlsson et al. 2005).

Empirical tests examining the influence of the different types of agglomeration on firm productivity have reached various conclusions. Braunerhjelm and Borgman (2004) analyse the relationship between concentration (in manufacturing and service sectors) and growth across Swedish regions for the period 1975–99. Their analysis indicates a positive and significant relationship between geographic concentration and labour productivity growth. They also find that greater levels of industry concentration and regional entrepreneurship stimulate productivity growth than skill levels and scale economies. Boschma and Iammarino (2009) examined the effect of agglomeration economies as well as the breadth and nature of international linkages on regional economic growth. Finding strong evidence that economic activity within related industries and regions contributes to regional economic growth, they conclude that regional growth is not influenced by being well-connected to the outside world or having a high variety of knowledge flowing into a particular region. Making a distinction between the size of the industrial sector and the density of the urban economy, Ke (2010) studied the effects of industry agglomeration and congestion on urban productivity across 617 Chinese cities in 2005. The results show that industrial production is an important cause of higher productivity in large industrial cities and cities in neighbouring regions, with the higher productivity resulting from concentrated industrial production allowing for further agglomeration. Results derived from a spatial econometric model show that intercity spillovers are a key contributor to increased agglomeration and productivity in neighbouring cities. Using survey data, Schmit and Hall (2013) analyses the importance of various firm, market, and agglomeration factors affecting New York food processors. Their results indicate that firm growth is related to important inter-industry market conditions. In addition, the clustering of similar manufacturers has important effects on firm revenue growth, with the benefits of firm clustering increasing significantly with the level of local urbanization. Hafner (2013) evaluates the impact of agglomeration economies on the clustering of European and German firms. Overall, empirical evidence provides support for the existence of inter-industry economies for all types of European firms and specifically for German knowledge-intensive firms.

A number of previous studies have analysed the agglomeration phenomena in South Africa. Most of these studies (e.g. Brand 2017; Lewis and Bloch 1998; Mahofa 2017; Naudé and Krugell 2006a; Turok and Borel-Saladin 2013) have either focused on spatial distribution of economic activity or examined the economic effects of urbanization and the role of agglomeration in fostering growth of economic zones and city-region economies. The notable exception to this trend are the studies by Fedderke and Wollnik (2007) and Krugell and Rankin (2012). Employing industry data at the provincial level, Fedderke and Wollnik (2007) consider a panel of regional distribution of firms within South Africa's manufacturing sector to evaluate the spatial structure of production. Their results indicate that high plant-internal scale economies, intensity in the use of human capital, and high industry-specific productivity gradients between regions are linked with greater geographical concentration of an industry. They also find lower levels of concentration in industries with strong inter-firm linkages and report that scale economies are the most important pro-concentration force. Krugell and Rankin (2012) employ data from the 2003 and 2007 Investment Climate Assessment World Bank Enterprise Surveys, and examine whether location can explain differences in the efficiency of South African manufacturers. The estimates from a firm-level production function incorporating location-specific explanatory variables show that place-specific size of the home market, level of education, and share of exports are found to be positively related to firm-level productivity. They also find that there is a negative relationship between manufacturing specialization and efficiency, and an indication that firms gain more benefits from diversity and urbanization economies than from specialization and localization economies. To the best of our knowledge, however, no previous study using South African administrative data has analysed the relationship between regional size and productivity with firm-level data and conditioned on attributes of individual firms.

3 Empirical framework

3.1 Productivity estimation

In establishing the link between agglomeration and firm productivity, agglomeration economies are typically treated as a kind of technology component that serves to shift the firm's production or cost functions. Following Krugell and Rankin (2012) and Rankin (2001), we note that at the firm level a typical specification of the production function would be

$$Y = g(U)f(X) \tag{1}$$

where Y is the the output level of the firm, X is a vector of factor inputs (such as land, labour, capital, and materials), and $g(U)$ is a vector of influences on production that arise from agglomeration economies, which captures the firm's environment and allows for the influence of agglomeration. Accordingly, we use firm-level data to provide an empirical representation of the type of production function outlined above, allowing for estimation of the effect of agglomeration on firm productivity. This is achieved by specifying a variant of the translog production function which includes a primary production function along with a set of inverse input demand equations. Using this particular approach, we can sketch out a reasonably complete specification of the production technology of firms which allows us to analyse the agglomeration–productivity nexus. From the literature review, industry agglomeration is generally assumed to improve firm productivity through localization or urbanization economies. Martin et al. (2011) suggest that when firm-level data is available, a natural empirical approach to follow is the estimation of a standard Cobb–Douglas production function, of the type:

$$Y_{it} = A_{it} K_{it}^{\beta_k} L_{it}^{\beta_l} M_{it}^{\beta_m} \tag{2}$$

where Y_{it} is the value–added of firm i at time t, A_{it} is the Hicks neutral efficiency level or TFP, K_{it} denotes capital stock, L_{it} is the total labour force of firm i at time t, and M_{it} is the cost of intermediate inputs. The parameters β_k, β_l, and β_m are the output elasticities of capital, labour, and intermediate goods, respectively. Log-linearizing the production function, one obtains:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \varepsilon_{it} \tag{3}$$

where $\ln A_{it}$ is defined by the term α_{it}, and where lower case denotes the log of variables in Equation 2 and $\ln(A_{it}) = \beta_0 + \varepsilon_{it}$. The constant term β_0 is a measure of the average efficiency level across firms and over time, with ε_{it} a composite error term that embeds a productivity shock that is only observable to a firm affected by its occurrence. Decomposing ε_{it} into its observable and unobservable components yields the following equation:

$$y_{it} = \beta_o + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \omega_{it} + \upsilon_{it} \tag{4}$$

where $\ln(A_{it}) \equiv \beta_0 + \omega_{it}$ represents firm-level productivity. ω_{it} is the component of the composite error term that is observed by the firm when making decisions on optimal factor input choices—that is the transmitted productivity component, and is thus correlated with capital, labour, and intermediate inputs. υ_{it} is an idiosyncratic error term that represents unexpected deviations from the mean due to measurement error, unexpected delays or other external circumstances (Beveren 2012). The main difference between ε_{it} and υ_{it} is that the former is a state variable that impacts the firm's liquidation and input decision rules, while the latter has no impact on the firm's decision problem and results from either a measurement error or a non-predictable productivity shock (Olley and Pikes 1996; Petrin et al. 2004). From Equation 4, TFP can be estimated as the level of output not attributable to factor inputs and reflects productivity associated with efficiency gains and technical progress. In this case, TFP is calculated as a residual term that is expressed as:

$$\ln(\hat{A_{it}}) \equiv \hat{\beta}_0 + \hat{\omega}_{it} = y_{it} - \hat{\beta}_k k_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_m m_{it} \tag{5}$$

Alternatively, when υ_{it} are assumed to be unobserved factors contributing to firm efficiency, then $\ln(A_{it}) \equiv \beta_0 + \omega_{it} + \upsilon_{it}$.

Estimation of the model specified in Equation 4 using a traditional econometric approach such as panel pooled ordinary least squares (OLS) relies on the assumption that inputs in the production function are independent of either productivity shocks or a firm's efficiency level. According to Marshack and Andrews (1944), this assumption is not always valid as profit-maximizing firms are able to observe productivity early enough to influence their input levels. Hence, firms will mobilize additional inputs (reduce the use of inputs) in response to positive (negative) productivity shocks. Sangalli and Lamieri

(2015) note the role of prior beliefs in input choices, with labour allowed to vary according to productivity shocks and firm decisions on capital inputs predetermined on the basis of investment decisions undertaken in the previous period. The possible endogeneity of inputs creates a simultaneity problem in the estimation of Equation 4 as the correlation between regressors and error term will yield biased and inconsistent estimates of the coefficients for the labour, capital, and intermediate inputs.

One approach to addressing potential biased and inconsistent OLS estimates is to estimate Equation 4 using a fixed-effects regression technique. One drawback of fixed-effects estimation is the assumption that the relationship between productivity and firm behaviour is time-invariant. Since this assumption may not necessarily hold, a class of semi-parametric models have been developed to address the econometric challenge posed by unobserved heterogeneity and simultaneity using micro-level panel data.

The two most popular of these models include the approaches of Olley and Pikes (1996) and Levinsohn and Petrin (2003) (hereafter OP and LP, respectively), which use a two-stage procedure where unobserved TFP is proxied by state variables such as investment or intermediate inputs, to account for the simultaneity problem. In this study, we adopt the LP method to estimate TFP. As He et al. (2017) note, estimating Equation 4 using the the LP method yields unbiased estimates of the production function, even when variable inputs are endogenous to local conditions or other time-varying unobservables that influence productivity.[1]

Where intermediate inputs are used to proxy for unobserved productivity level, then intermediate inputs can be expressed as a function of capital and productivity: $m_{it} = m_{it}(k_{it}, \omega_{it})$. Demand for m_{it} is assumed to be monotonically increasing in ω_{it}. This allows for the inversion of the intermediate demand function such that the unobservable productivity term is expressed solely as a function of two observed inputs:

$$\omega_{it} = \omega_{it}(k_{it}, m_{it}) \tag{6}$$

Using Equation 6, the production function in Equation 4 can be rewritten as

$$\begin{aligned} y_{it} &= \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \omega_{it} + \upsilon_{it} \\ &= \beta_l l_{it} + \varphi_t(k_t, m_t) + \upsilon_{it} \end{aligned} \tag{7}$$

where now $\varphi_t(k_t, m_t) = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \omega_{it}(k_{it}, m_{it})$

Estimation of Equation 7 proceeds in two steps. The first involves first-stage least squares regression:

$$y_{it} = a_2 l_{it} + \varphi(m_{it}, k_{it}) + \mu_{it} \tag{8}$$

[1] The choice of the LP method over that of OP is dictated by the data. The OP estimator employs a firm's investment decision and state of exit as proxies to overcome simultaneity. This implies that application of the OP method is not feasible when the data contains a large number of zero-investment observations or there is difficulty recovering reliable data on investments for small and medium-sized firms. The OP estimation also requires investment to be a strictly monotonic function of unobserved productivity. A large number of zero observations for the investment variable in South Africa's administrative tax data will require the unrealistic assumption that all firms with zero investment have the same level of unobserved productivity. In addition, information on the firm's exit does not truly reflect a change in operational state, and thus negates the validity of firm exit as a proxy for unobserved productivity. These problems are circumvented with the LP method.

where the consistent estimation of parameters of Equation 8 using the OLS method is made feasible by substituting a third-order polynomial approximation in k_{it} and mit in place of $\omega_{it}(k_{it})$. The estimate of $\varphi(k_{it}, m_{it})$ is then specified as

$$\hat{\varphi}_{it} = y_{it} - \hat{\beta}_l l_{it} \tag{9}$$

Equation 9 provides estimates of β_l and ψ_t. The second stage of the LP procedure yields estimates of β_k via a nonlinear regression of the form:

$$y_{it} = \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it} = \psi(\hat{\varphi}_{it-1} - \hat{\beta}_k k_{it-1}) + \upsilon_{it} \tag{10}$$

where $\psi(\cdot)$ is another polynomial of order three. He et al. (2017) note that Equation 10 is estimated by minimizing the sum of squared errors and once $\hat{a}_1$ and$\hat{a}_2$ are recovered, then the predicted value of TFP can be obtained as

$$\hat{\omega}_{it} = y_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_k k_{it} \tag{11}$$

3.2 The spatial model

Studies on firm productivity are limited in scope when they ignore possible externalities or spillover effects coming from other geographic partitions (particularly the nearest ones) constituting the economic system of the region (Davi and Barbaccia 2015). Given the spatial dimensions of South Africa's economic development, the empirical analysis of the relationship between urban agglomeration and firm productivity will need to account for the presence of spatial spillover effects. In this context, the analytical framework for the spatial analysis of productivity and agglomeration is described in Figure 1.

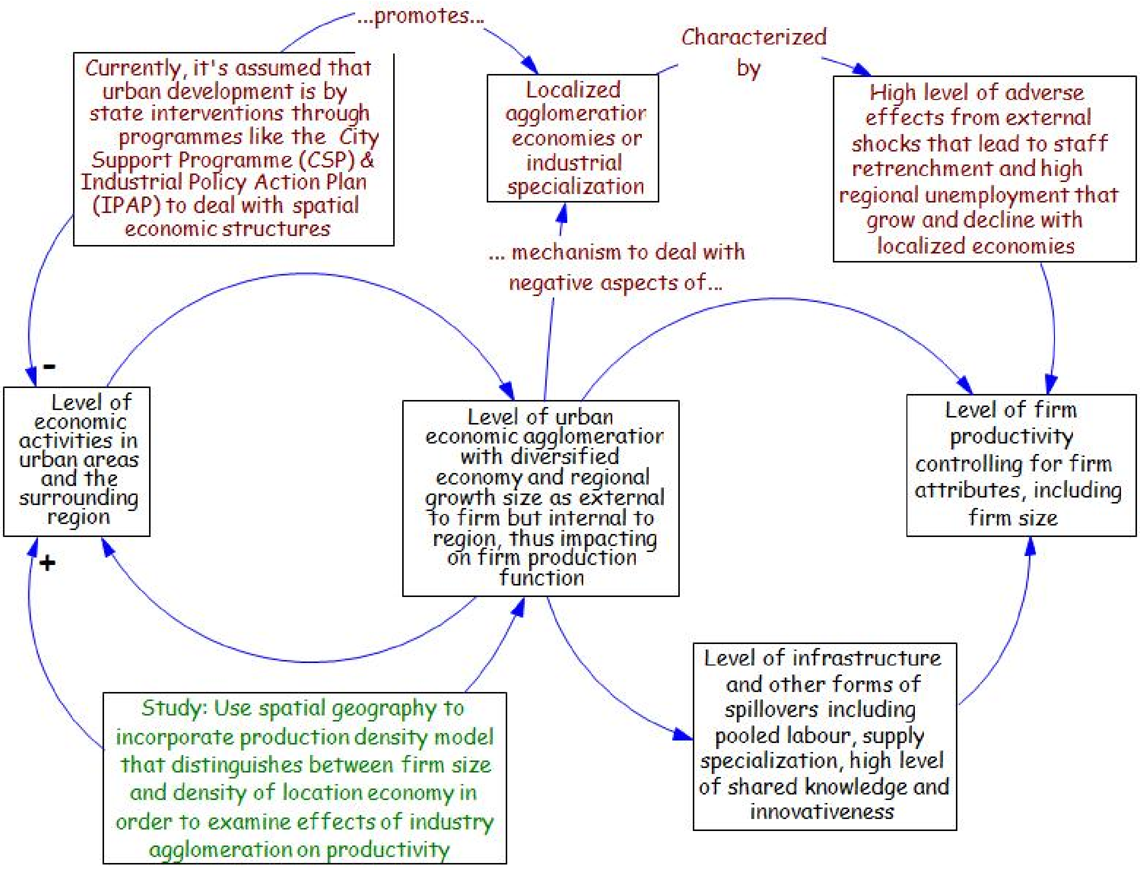
For the next step in the empirical analysis, we formalize the analytical framework into a complete spatial autoregressive model (SAR) in order to estimate the possible presence of TFP spillovers originating from a firm's geographical or sectoral neighbours. For a cross-section of n spatial units, the interactions between such spatial units can be expressed by a SAR model of the form:

$$\begin{aligned} \mathbf{y}_n &= \mathbf{X}_n \beta_n + \lambda_n \mathbf{W}_n \mathbf{y}_n + \mathbf{u}_n \\ \mathbf{u}_n &= \rho_n \mathbf{M}_n \mathbf{u}_n + \varepsilon_n \end{aligned} \tag{12}$$

where $\mathbf{y}_n$ denotes $n \times 1$ vector of observations (i.e. firm-level TFP estimates) of the dependent variable, $\mathbf{X}_n$ denotes the $n \times k$ matrix of exogenous regressors, $\mathbf{W}_n$ and $\mathbf{M}_n$ are $n \times n$ row-normalized spatial weight matrices with zeros in the main diagonal. n represents the number of firms in the sample, and the elements of the matrix $\mathbf{W}$ refers to 'spatial weight', measuring the strength of the relationship between neighbouring firms. $\mathbf{u}_n$ denotes the $n \times 1$ vector of regression disturbances while ε_n is an $n \times 1$ vector of innovations. λ_n and ρ_n are spatial autoregressive parameters, and β_n is a $k \times 1$ vector of unknown parameters (Kelejian and Prucha 2010). By stacking the equations over time periods, Equation (12) can be transformed into its panel representation, defined as:

$$\begin{aligned} \mathbf{y} &= \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \beta + \mathbf{u} \\ \mathbf{u} &= \rho \mathbf{M} \mathbf{u} + \varepsilon \end{aligned} \tag{13}$$

Figure 1: Hypothesized channels for exploring industry agglomeration and firm productivity in South Africa



Source: authors' creation.

Despite not having actual firm locations (x, y coordinates), with municipality as the spatial unit of analysis, we were able to aggregate the firms' TFP at the multiplicity level and to conduct a rigorous analysis of the inter-regional spillovers of local firm-level productivity by estimating a formal versions of Equation 12 expressed (in stacked form) as:

$$\ln(tfp)_t = \lambda \mathbf{W} \ln(tfp)_{it} + \mathbf{X}_{it} + (\beta \mathbf{W}(\mathbf{X}_{it})) + \varepsilon_{it} \tag{14}$$

where the $\ln(tfp)_{it}$ object contains (logged) levels of aggregate TFP for firms in location i at time t, where firms' TFPs were derived from estimation of Equation 5, λ is the SAR parameter associated with the spatial lag of TFP $\mathbf{W}\ln(tfp)_{it}$, with $\mathbf{W}$ accounting for the weighted contributions of the productivity levels pertaining to neighbouring firms j. $\mathbf{X}_{it}$ is a vector of exogenous covariates, $(\beta \mathbf{W}(\mathbf{X}_{it}))$ is included for the spatial distribution model (SDM) estimation, and represents the weighted lags of the covariates. ε_t is a pure idiosyncratic error term (He et al. 2017).

4 Data description and variable measurements

4.1 Data

The data used is from a recently compiled administrative dataset developed as a joint SARS–National Treasury–UNU-WIDER initiative. Known as the *SARS–NT Panel*, the compilation of the dataset on an annual basis from 2008 through 2014 makes use of four different sources: (1) company income tax (CIT); (2) employee income tax certificates (*IPR5* and *IT3a* submitted by employers); (3) value-added tax (VAT); and (4) customs data. For the purpose of this paper, only the first three sources are necessary. The data includes identity variables for firms (tax reference number: *Tax_RefNo*) and individuals (payee reference number: *PAYE_RefNo*). Matthee et al. (2018) notes that tax return forms completed by larger firms differed from those completed by others from 2013 onwards. However, the appended data addresses this issue, and the variables required for TFP calculations were adjusted accordingly.

For this study, there are two aspects to the data compilation and analysis. The first aspect is concerned with the calculation of TFP, which links directly with the CIT_IRP5 data. The data links each inputted individual's IRP5 data with a corresponding firm tax return, either through the company-completed CIT or the employer-completed IT3a. This allows for the calculation of TFP at the firm level. To calculate TFP, there are five main variables necessary: income/revenue or value-added, intermediate inputs, capital, labour, and costs. From the Levinsohn and Petrin (2003) approach to TFP calculation, the dependent variable can either be the firm revenue or the value-added reported by the firm. Both variables are available in the SARS–NT panel. However, the literature has shown that revenue is the better measure of firm output, and value-added should only be used as an alternative in the absence of firm revenue. Furthermore, the data has more instances of missing value-added observations relative to firm revenue. Thus, in this study, firm revenue is chosen as the dependent variable. In the case that any one of these variables is missing, the TFP for such a firm is not obtainable. Hence, firms for which one of the core TFP calculating variables is missing are dropped from the analysis.[2]

In a similar manner, the data has several different measures for labour, with some weighted and others unweighted. Six employment measures are available in the data: employee numbers based on the number of forms submitted (*IRP5_forms*), employee numbers based on selected codes but excluding pensioners (*IRP5_empl*), employee numbers inclusive of pensioners (*IRP5_empl3601*) and their respective weighted versions (*IRP5_forms_daysweight*, *IRP5_empl_daysweight*, and *IRP5_empl3601_daysweight*). The weighted measures are closer estimates to actual labour numbers, as there is the possibility of an individual being counted as a full-time, full-year employee with several firms in the same year, whereas this employee would actually have split their time between all those firms. With the weighted measure of employment, such multiple counting is accounted for, and weights assigned to adjust for the hours dedicated to a linked firm. For this study, the weighted version of *IRP5_empl* (*IRP5_empl_daysweight*), is included as the employment variable in estimating firm-level TFP.

The second way in which the spatial agglomeration analysis is carried out is to compile the data for location and agglomeration dynamics. This requires combining location data with the CIT and IRP5 data. The data with firm location contains information on each firm's PAYE reference number, as well as four locational information variables: province, district municipality, local municipality, and main place. Among the four available location variables, *main place* is the narrowest geographical level available. However, this variable has the largest number of missing observations of the four. Local municipality was thus selected as the indicative spatial unit. In order to merge the location information with the administrative tax data in the CIT_IRP5 dataset, the PAYE reference numbers were matched with their employees using the linked corresponding tax reference numbers. It is only after this matching has been carried out for each year that the location data can be matched to the CIT_IRP5 data. In total, the matched firms with location data resulted in about 500,000 firms. However, there were only about 10,000 firms from 2008, and thus the problem of missing observations persisted.

[2] The year 2008 is especially problematic as it has many firms with at least one required variable missing. Out of 500,000 firms with comprehensive geo-referenced data and for which TFP was calculated over the entire sample period, about 10,000 firms had complete observations in 2008. These firms account for a low proportion (about 2 per cent) of the entire dataset and were subsequently excluded from the spatial analysis.

The empirical analysis of spatial and agglomeration effects on firm productivity, the focus of this study, requires that a firm's municipal information is available consistently across the period 2008–14. However, given the low firm representation in 2008, which had only 82 municipalities represented out of the possible 213, there were unbalanced panel issues. Due to the low number of cases and spatial units represented in 2008, we did not include the 2008 data is the spatial panel model.

The vector of exogenous covariates, $\mathbf{X}_t$, in Equation 14, considers two types of independent variables: firm characteristics that have no impact on other firms, and market conditions that affect not only firms in a specific area, but also firms in neighbouring geographical regions. The literature (e.g. Feng and Wang 2019; Sheng and Song 2013) highlights the important role of firm-level characteristics in determining productivity. For the choice of firm idiosyncrasies, we use firm age (*age*) to capture the potential productivity gains associated with the operational context in which firms are located. More specifically, firm age is included to evaluate whether younger firms are more able to adapt new technology and produce more efficiently than older plants, or whether as firms age they improve their productivity over time by learning from their experience of operating within an industry (or market). Noting the information available in the tax administration database, we follow the suggestions of Harris and Moffat (2015) and consider two aspects of the local market environment: an index of industry (or market) concentration across firms, and a measure of industry agglomeration. The concentration index serves as a measure of market power, and is included to take into account competition effects. To estimate market concentration, we use the Herfindahl—Hirschman index (HHI) of gross sales in a particular industry relative to the value of output produced nationwide. The HHI is defined as:

$$HHI = \sum_{n=1}^{n} s_i^2 \tag{15}$$

where s_i is the share of gross sales by industry (or economic sector) i in overall gross sales by all industries (or economic sectors). The HHI takes a value between zero and unity, with smaller values indicating stronger competition. The HHI is thus able to capture the effect on concentration not only of the dimensional variability of firms, but also of their number.

Urban economic theory posits that firms gain productive advantages from locating in close proximity and the existence of such advantages can explain the formation and growth of cities and compact industrial areas (Graham 2007). The traditional approach to measuring agglomeration utilizes total population or total employment to provide an empirical measure of city size. However, this approach is criticized for its simplicity and its reliance on administrative areas that do not readily correspond to cities or urban densities/agglomerations.

We note that difficulties arise when one tries to set boundaries to define distinct metropolitan areas. For instance, while Johannesburg and Pretoria are nominally two separate cities, there is interaction between the two over relatively small distances that arguably prevents them from being truly distinct. Likewise, it is conceivable that a firm located outside the Johannesburg metropolis can still enjoy agglomeration benefits through proximity that arise from the scale of Johannesburg and its industries. The goal is to avoid placing a-priori restrictions on the geographic scope of agglomeration that correspond to some predefined boundaries. In light of these considerations, this study follows Graham et al. (2010) and obtains an agglomeration measure based on the concept of effective density. This concept combines the scale/size of the city or agglomeration with a relative proximity to activities, where applicable densities can be calculated for very small areas of the country. Specifically, we use municipal-level employment to construct a measure of agglomeration experienced by each firm. The total effective density of employment that is accessible to any firm located in municipality i is given by

$$U_i = \frac{E_i}{r_i} + \sum_{j}^{\iota \neq j} \frac{E_j}{d_{ij}} \tag{16}$$

where E_i is total official employment in municipality i, r_i is an estimate of the radius of municipality i, E_j is total employment in municipality j and d_{ij} is the distance between i and j. The effective density of the employment measure of agglomeration (U_i) is designed to contain an implicit transport dimension: proximity. The measure of proximity used is based on the straight-line distance and is calculated using Pythagoras' theorem and the x, y coordinates of the municipal centroid.[3]

Finally, spatial weights are an essential element in the generation of spatial autocorrelation statistics. To capture the existence of a neighbour relationship, we construct a spatial weight matrix $\mathbf{W}$ of reciprocal influences based on adjacent municipalities. $\mathbf{W}$ is thus derived using the contiguity method whereby weights indicate whether spatial units share an adjacent boundary or not.[4] For this purpose, we follow the example of He et al. (2017), with firms geo-referenced according to latitude and longitude coordinates. The coordinates are derived from the location of the spatial unit (in this case a municipality, m) pertaining to the main operational region of each firm in the SARS–NT panel. The generated weights are thus a function of shared municipal borders between firms in municipality i and a generic neighbour firm in municipality j.

Table 1 shows the variables used in the estimation of our production function as well as in the SAR model. Similar to Matthee et al. (2018) and Pieterse et al. (2018), from the CIT_IRP5 data, we obtain variable values as follows: firm output values are derived from the income statement's sales (turnover) variable; capital is computed from the balance sheet's non-current assets (i.e. fixed property, plant, and equipment plus other fixed assets where available); labour costs include (where available) wages, as well as contributions to insurance, the unemployment insurance fund and the skills development levy, pension, provident fund, medical scheme, professional body membership, and training. Finally, intermediate inputs are calculated from the income statement using the cost of sales variable where available and purchases otherwise. From the summary statistics in Table 1, two features stand out. First, on average, the number of employees hired by firms is 37, indicating a substantial proportion of small and medium-sized firms engaged in economic activities in South Africa. Second, over the sample period, the average firm age was just over 15 months, indicating extensive firm-level churn across South Africa's productive sectors.

[3] By following this approach, we generate *effective* density measures of agglomeration that have a number of desirable properties, including (1) allowing for a highly flexible spatial framework that is not constrained by predefined spatial units such as administrative areas or metropolitan definitions; (2) densities that incorporate an implicit transport dimension as they reflect the importance of the scale and proximity (accessibility) of economic activity to each firm; (3) applicability to estimations for very small areas of the country (i.e. the 500 municipalities) allowing for a high degree of spatial detail in analysis; and (iv) ability to serve as a useful proxy to represent the time and cost, as well as physical, dimensions of density.

[4] Prior to normalization, the spatial contiguity matrix expresses the existence of a neighbour relation as a binary relationship, with weights 1 and 0, with 1 indicating that municipal areas are adjacent.

Table 1: Variables, definitions and summary statistics

Variables	Mean	Std dev.	Min.	Max.
Gross sales/revenues (**Y**)	4.47E+07	1.14E+07	1	5.80E+10
Capital (**K**)	1.06E+07	1.02E+09	1	5.80E+11
Labour (**L**)	37	385	0	48307
Intermediate input (**M**)	3.24E+07	7.11E+08	1	2.32E+11
Labour cost (**W**)	4,337,706	8.01E+07	1	2.24E+10
Firm age (**Age**)	15.71	11.62	1	$72^{\dagger}$
Agglomeration (**Agglo**)	48,838,553	1,956,692	1,338,961	1.51E+07
Herfindahl–Hirschman index (**HHI**)	0.021	0.05	0.002	0.794

Notes: † derived by multiplying the number of months in a calendar year with the duration of a firm's existence as recorded in the CIT information used in compiling the tax administration database. With the exception of the measure of agglomeration, there were 491,309 observations for each variables. Observations for *Agglo* totalled 469,222.

Source: authors' calculations using data from SARS–NT panel dataset.

5 Empirical results

5.1 Estimates of the production function

Table 2 summarizes the results of parameter estimates of the baseline production function estimated using the OLS, fixed effects, and LP, as well as Wooldridge's (2009) two–step procedure. As earlier stated, the OLS approach does not account for the simultaneity in place between productivity shocks and the choice of inputs in the production function (i.e. labour input), while the fixed effects assumes a time-invariant relationship between productivity and firm behaviour. However, both represent useful upward benchmarks for the LP estimates.[5]

Table 2: Estimates of production function using OLS, FE, LP, and Wooldridge approaches

Coefficient (variable)	(1) OLS	(2) FE	(3) LP	(4) WLDRG
β_o	2.831 (0.01)***	6.101 (0.000)***		
β_k ($\boldsymbol{k}$)	0.03 (0.000)***	0.01 (0.000)***	0.017 (0.000)***	0.02 (0.000)***
β_l ($\boldsymbol{l}$)	0.07 (0.000) ***	0.09 (0.000)***	0.05 (0.001)***	0.18 (0.001)***
β_m ($\boldsymbol{m}$)	0.514 (0.000)***	0.450 (0.000)***	0.457 (0.04)***	0.194 (0.004)***
β_{wages} (labour costs)	0.343 (0.000)***	0.182 (0.000)***	0.241 (0.03)***	0.666 (0.005)***
Observations	488,968	488,968	488,968	318,080
R^2	0.905	0.582		
Tax reference numbers		149,849		
Groups			149,849	149,849

Notes: standard errors in parentheses. All variables in the estimations in columns 1–4 are in logarithms. The LP approach (column 3) reports results with the proxy variable (i.e. *intermediate inputs*) and the cost of labour (i.e. wages) as freely variable inputs. ***, ** and * indicate significance at the 1, 5, and 10 per cent levels, respectively. With the Wooldridge method based on a generalized method of moments (GMM) setup, observation are treated as lagged as instruments, hence the drop in sample size in column (4)

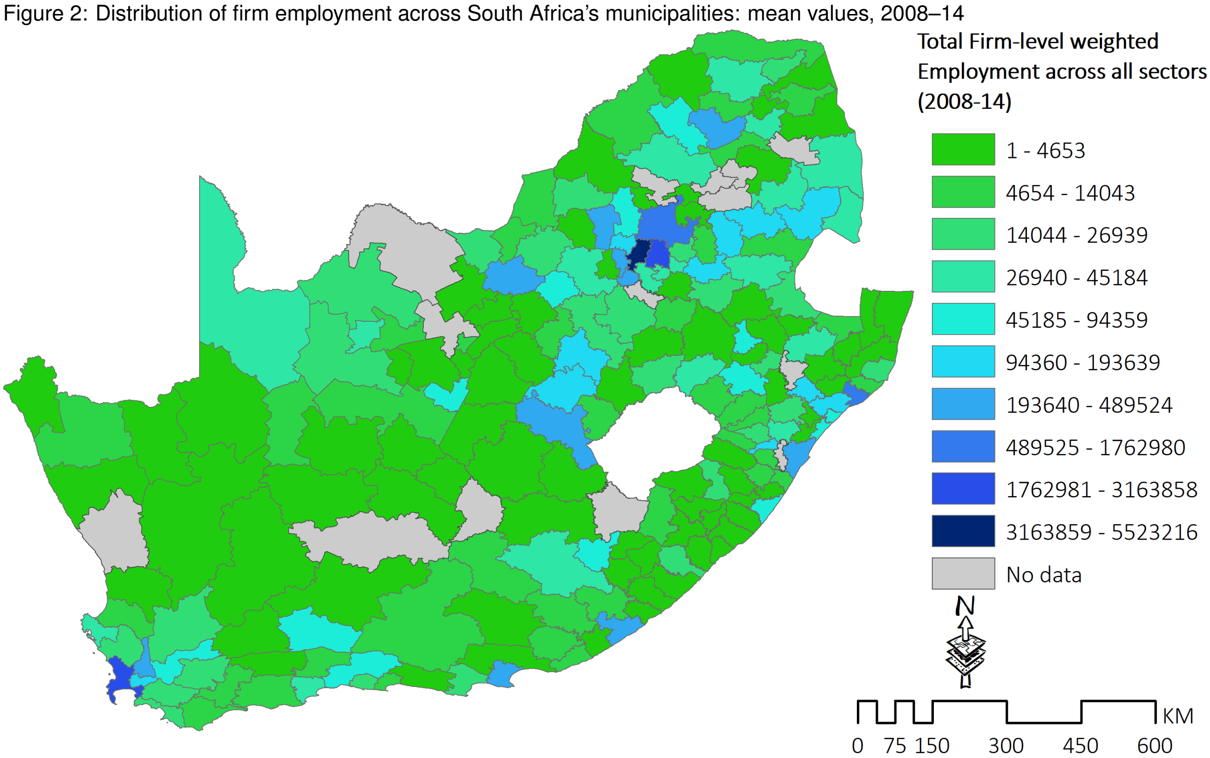
Source: authors' calculations using data from the SARS–NT panel dataset.

[5] The LP approach does not allow for fixed effects and assumes that, owing to zero costs, firms are able to adjust some inputs instantly when subjected to productivity shocks.

From Table 2, the elasticities of output with respect to capital, labour, and intermediate inputs are positive and statistically significant across all four estimation approaches. The average sum of output elasticities—that is, the returns to scale—is 0.5, indicating that outputs across South Africa's industries are subject to decreasing returns to scale (DRS).[6] The results indicate that firm age and labour remuneration (wages) affect output positively and the effects are statistically significant. We include wages as one of the explanatory variables in the specification, leading to the results in Table 2. Syverson (2010) notes that along with environmental factors such as productivity spillovers, inter-firm differences in productivity are driven by features of a firm's operations, including the quality and costs of labour (and capital). With the functional form of the LP approach allowing for the inclusion of control variables, we include wages to account for the fact that in the South African context, the costs of labour and related wage bargaining dynamics are key issues impacting firm output and productivity.

Next, we evaluate the spatial dimensions of the TFP estimates across South African local municipalities, including the metropolitan municipalities. As a first step, using firm-level employment over the period 2008–14, we plot the distribution of employment by South Africa's firms operating across the country's municipal sphere. Figure 2 shows a distinct concentration of employment in municipalities located in South Africa's major economic clusters of Gauteng, Western Cape, KwaZulu–Natal and parts of the Eastern Cape.

Figure 2: Distribution of firm employment across South Africa's municipalities: mean values, 2008–14

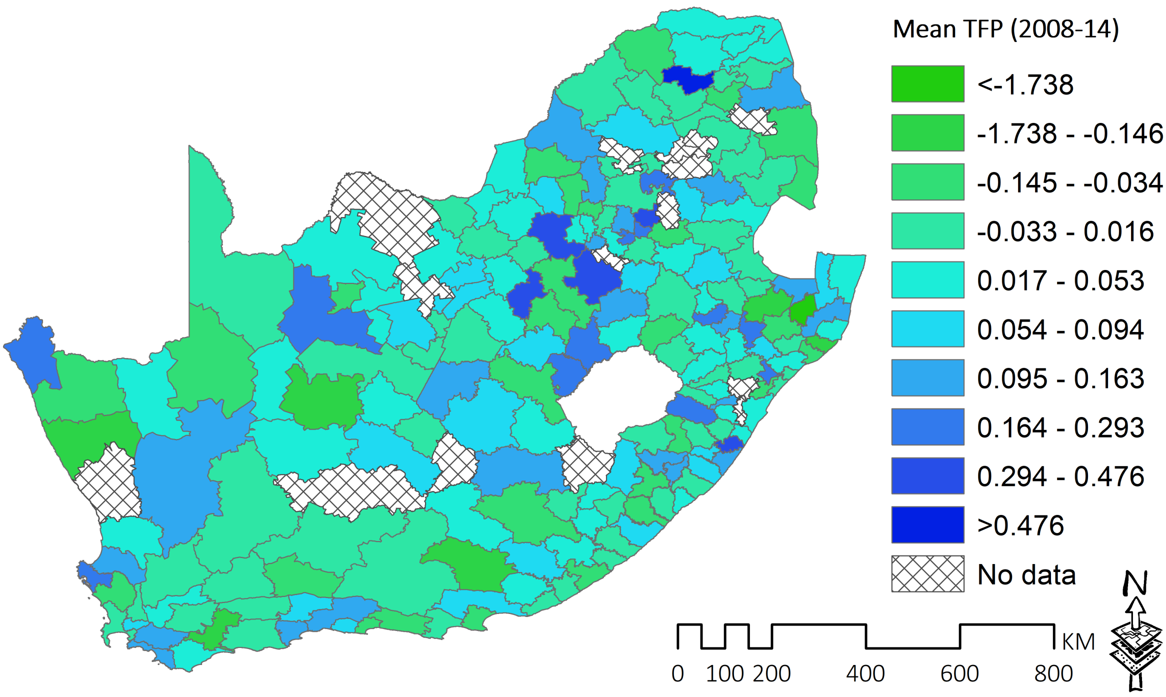


Source: authors' creation, based on data from the SARS–NT panel datset and Statistics South Africa Geography Metadata.

We plot the average (predicted) TFP distributions across municipalities in Figure 3. Ordering the predicted TFP values from highest–to–lowest, we note that, unlike Figure 2, there is some relative dispersion in productivity away from the major economic clusters. In particular, we notice that average TFP is highest in parts of South Africa's north–west, north-east, and central regions—municipalities located in parts of the Northern Cape, Free State and North-West, and Mpumalanga provinces, where agriculture and mining are the dominant economic activities.

[6] The finding of returns to scale less than 1 is similar to results reported by Krugell and Rankin (2012).

Figure 3: Geo-referenced distribution of TFP in South Africa: mean values, 2008–14



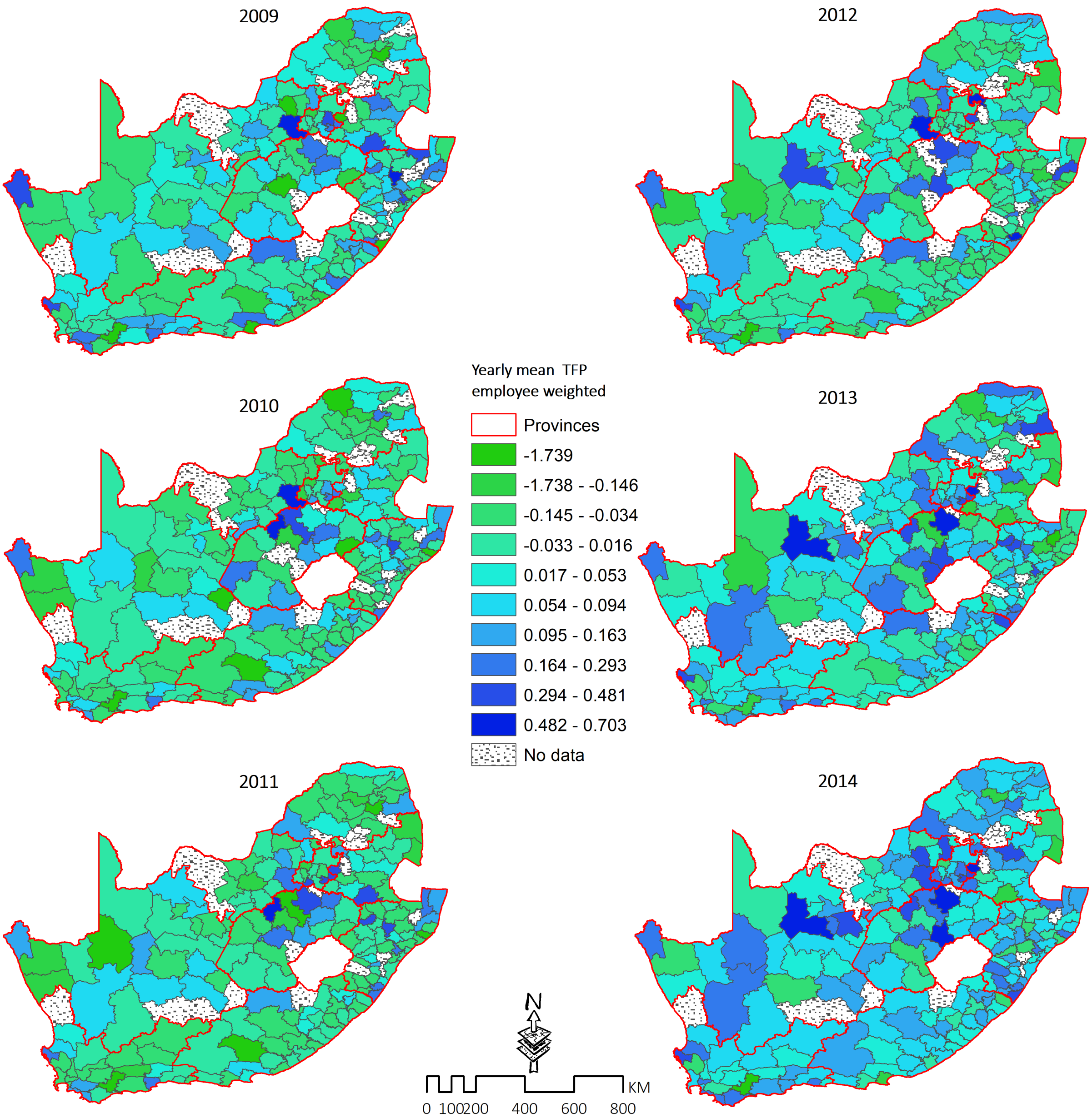
Source: authors' creation, based on data from the SARS–NT panel datset and Statistics South Africa Geography Metadata.

The productivity distribution in Figure 3 does not reveal the complete picture of the evolution of TFP over the sample period. We thus plot Figure 4 to compare, for each year in the sample period, the mean TFP estimates for all firms across all sectors between municipalities. We observe that distributions are consistent with those in Figure 3, with the distribution of municipal TFP data following a certain spatial pattern: first, a large number of high TFPs are concentrated in a small number of municipalities (deep blue colour); and second, in later years the TFP measures are much higher and municipalities with higher TFPs are clustered in geographic space. This pattern indicates that, in relative terms, firms located in regions of the North-West, Northern Cape, Free State, and Mpumalanga provinces have experienced a gradual increase in productivity. On the other hand, average productivity in the key economic hubs around Johannesburg, Cape Town, Durban, Port Elizabeth, and East London have remained relatively constant over the sample period.

We extend the analysis by examining productivity dynamics across South Africa's subsectors. Using the SARS International Standard Industrial Classification of All Economic Activities (ISIC) codes, we group firms into the 10 broad sectors/industries. From the groups, we select a major subcategory within the primary, secondary, and tertiary sectors and conduct a sectoral composition of the dataset to derive TFP estimates for the manufacturing, agriculture, forestry and fisheries, and financial and real estate subsectors.[7] Similar to estimates obtained in the aggregated case, the coefficients on capital and labour inputs identify a regime of decreasing returns to scale for the different branches of economic activity. Interestingly, the results show that for all three subsectors considered in Table 3, intermediate inputs have the most significant impact on firm-level output.

[7] The analysis at sectoral level is also informed by the need to provide a proof of concept to exemplify how spatial econometrics can help in expanding the analysis of the regional spatial dynamics of economic activities classified as productive sectors in South Africa. The choice of the specific sectors for analysis is informed by the availability of observations for the required variables needed to estimate the production function in Equation 4.

Figure 4: The space–time pattern of firms' TFP: South African municipalities, 2009–14.



Source: authors' creation, based on data from the SARS–NT panel datset and Statistics South Africa Geography Metadata.

Table 3: Sectoral estimates of production function using OLS, FE, LP, and Wooldridge approaches

Manufacturing				
	(1)	(2)	(3)	(4)
Coefficient (variable)	OLS	FE	LP	WLDRG
β_o	2.259 (0.01)***	4.960 (0.02)***		
β_k (k)	0.02 (0.000)***	0.01 (0.000)***	0.011 (0.001)***	0.015 (0.001)***
β_l (l)	0.049 (0.001)***	0.067 (0.001)***	0.04 (0.001)***	0.152 (0.001)***
β_m (m)	0.637 (0.000)***	0.563 (0.000)***	0.557 (0.04)***	0.185 (0.001)***
β_{wages} (Labour costs)	0.255 (0.000)***	0.148 (0.001)***	0.181 (0.02)***	0.692 (0.009)***
Observations	148330	148330	148330	95503
$R2$	0.951	0.677		
Agriculture, fisheries, and forestry				
β_o	2.940 (0.05)***	7.356 (0.115)***		
β_k (k)	0.02 (0.002)***	0.02 (0.002)***	0.03 (0.01)***	0.02 (0.003)***
β_l (l)	0.052 (0.01)***	0.105 (0.007)***	0.059 (0.01)***	0.181 (0.01)***
β_m (m)	0.521 (0.002)***	0.350 (0.005)***	0.404 (0.02)***	0.128 (0.02)***
β_{wages} (labour costs)	0.337 (0.004)***	0.192 (0.006)***	0.380 (0.05)***	0.585 (0.03)***
Observations	14650	14650	14650	8603
$R2$	0.870	0.436		
Finance and real estate				
β_o	3.563 (0.03)***	7.247 (0.06)***		
β_k (k)	0.03 (0.001)***	0.014 (0.001)***	0.022 (0.003)***	0.03 (0.002)***
β_l (l)	0.086 (0.002)***	0.126 (0.004)***	0.058 (0.003)***	0.216 (0.004)***
β_m (m)	0.428 (0.001)***	0.368 (0.002)***	0.423 (0.02)***	0.124 (0.013)***
β_{wages} (Labour costs)	0.343 (0.000)***	0.182 (0.000)***	0.241 (0.03)***	0.666 (0.005)***
Observations	46,109	46,109	46,109	24,890
R^2	0.853	0.487		

Notes: † Standard errors in parentheses. All variables in the estimations in columns 1–4 are in logarithms. The LP approach (column 3) reports results with the proxy variable (i.e. *intermediate inputs*) and the cost of labour (i.e. wages) as freely variable inputs. ***, **, and * indicate significance at the 1, 5, and 10 per cent levels, respectively.

Source: authors' compilation based on the SARS–NT panel dataset.

5.2 Estimates of the SAR model

The analysis of the space-–time dynamics of firm-level TFP in South Africa from a geographical perspective is carried out in two steps. First, we test for the existence of significant spatial patterns and associations between the TFP and its determinants. For the second stage, we integrate the knowledge production function with the spatial panel Durbin model to estimate the space-–time relationship between TFP and its covariates. In estimating the space–time relationship, we account for each variable's local effects and spatial externalities, and on this basis investigate the existence and extent of direct and indirect spillover effects.

For the explanatory spatial analysis, indicators of spatial autocorrelation of Moran's I type was applied to our productivity data, to detect forms of spatial heterogeneity (Haining 2003). To estimate Moran's I index, we applied two techniques: global and local indices of spatial autocorrelation.

The global Moran's I statistic gives a formal indication of the degree of linear association between the spatial units and their neighbours (Yu and Weie 2008). This statistic is specified as:

$$I_t = \frac{N\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} Z_{i,t} Z_{j,t}}{S_0 \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} Z_{i,t} Z_{j,t}} \tag{17}$$

where w_{ij} is an element of a spatial weight's matrix $\mathbf{W}$, which describes the spatial arrangement of all spatial units in the sample. The observations $Z_{i,t}$ and $Z_{i,j}$ are deviations of an attribute for feature i (or j) from its mean $(x_i - \hat{X})$ at time t. Finally, N is the number of observations in the sample and S_0 is a scaling factor to sum all the elements of W. The estimated Moran's score is used to test the null hypothesis which claims that feature values are randomly distributed across the study area. The global Moran's I value ranges from -1, indicating perfect dispersion, to 1, indicating perfect autocorrelation. Value of Moran's score close to 0 suggests a random spatial dispersion (Stankov et al. 2017). With highly robust significance (p-values of 0.000), the empirical values of the Moran's I statistic averaged 0.60 over the sample period (see Table 4). We thus reject the null hypothesis and conclude that over the entire sample period, there is positive spatial autocorrelation and this indicates that highly clustered patterns of economic activity persists across South Africa's municipalities.

Table 4: Global Moran's I, z-scores, and p-values from 2001 to 2013

Year	Global Moran's I	z-score	p-value
2008	0,611	12.275	0.00
2009	0.613	30.795	0.00
2010	0.613	31.08	0.00
2011	0.609	31.08	0.00
2012	0.606	30.89	0.00
2013	0.699	31.051	0.00
2014	0.596	30.973	0.00

Source: authors' compilation based on the SARS–NT panel dataset.

In examining spatial patterns of economic phenomena, it is important to also identify local hotspots or areas where economic activity is extremely pronounced across localities, as well as spatial outliers (Oliveau and Guilmoto 2005; Ord and Getis 1995). While the advantage of the global Moran's I lies in its simplicity, a major limitation is that the estimation in Equation 18 tends to average local variations in the strength of spatial autocorrelation. To address this drawback, we apply the local index of spatial association as a mechanism to identify areas where values of a variable of interest are both extreme and geographically homogeneous. This approach is most useful when, in addition to global trends in the entire sample of observations, there exist also pockets of localities displaying homogeneous values that do not follow the global trend. The standard tool to examine local autocorrelation is Anselin's (1995) local indicator of spatial association (LISA). Considered the local equivalent of Moran's I, the value of a LISA—that is, the local Moran's statistic for each observation i—is written as:

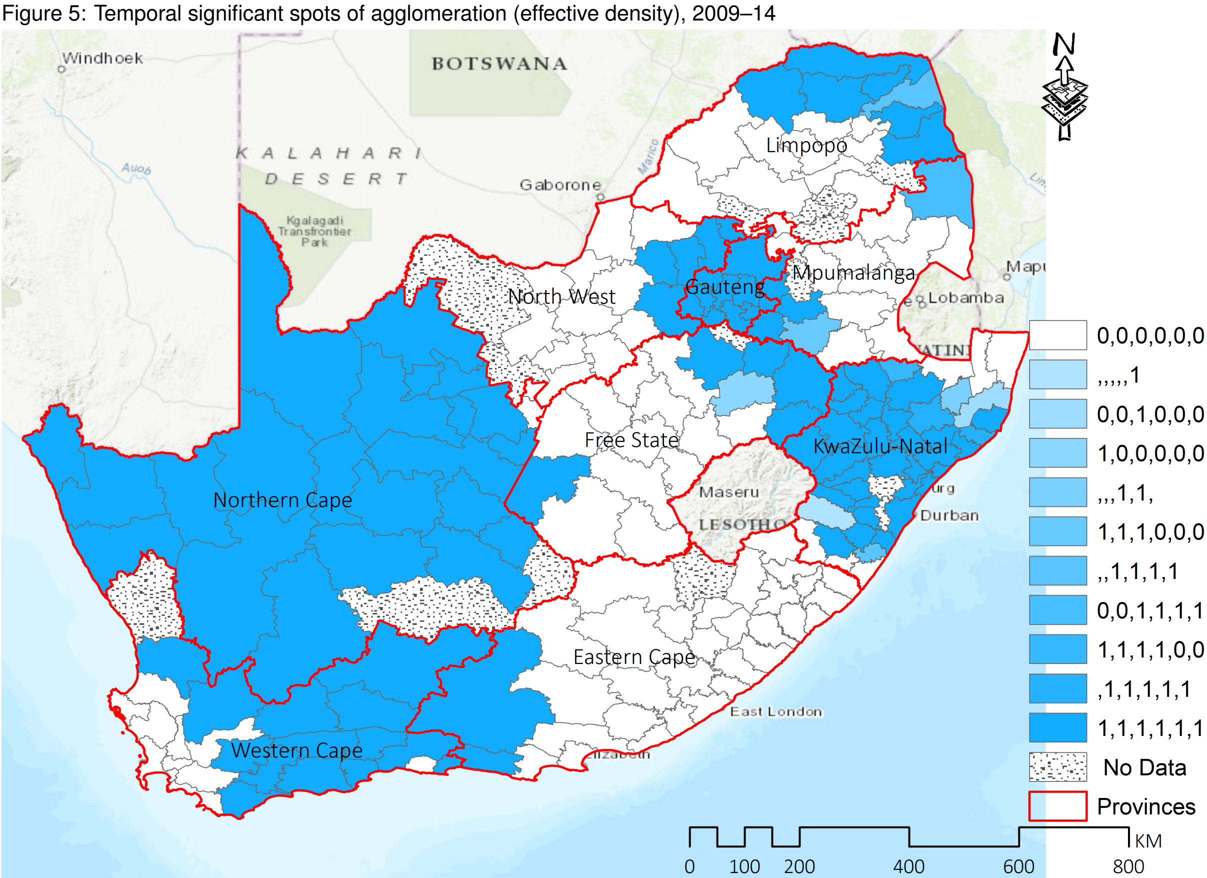
$$I_t = \frac{Z_i \sum_{i=1}^{N} w_{ij} z_j}{\sum_j^N \frac{Z_j^2}{N}} \tag{18}$$

where z_i and z_j are deviations from the mean. Under the null hypothesis of no spatial autocorrelation, the expected value of I_i is $E(I_i) = \dfrac{-\sum_j^N \mathbf{W}_{ij}}{N-1}$. According to Wang et al. (2016), when $I_i > E(I_i)$, the observation i of interest is characterized by positive spatial correlation (and vice versa when $I_i < E(I_i)$). Estimates from Equation 19 can be used to plot a significance map that depicts locations with significant local Moran statistics and allows for influential observations to be identified. However, the hotspots identified using the LISA method only reflect a particular type of spatial associations, namely high values surrounded by high values (high-–high). Thus, one drawback of the LISA significance map is its failure to denote other spatial association types (e.g. low-–high, low—low, and high—low). Noting that the visualization of spatial patterns cannot be solely based on hotspots, (Ye and Wu 2011) suggest a temporal stability mapping of hot spots. Such mapping is achieved by overlaying significant spatial associations at different years, thus allowing for the incorporation of the temporal nature of economic development and activities. Given the focus of this paper in assessing the spatial dimensions of the impact of agglomeration on productivity, we conduct a temporal stability mapping for TFP and the measure of agglomeration. The results are presented in Figures 5 and 6.

From the visualization of agglomeration, we observe that 11 categories of temporal stability are identified in Figure 5. If a municipal-level unit remains a significant agglomeration spot over the 2009–14 period, it is identified as '1, 1, 1, 1, 1, 1'. A municipality is labelled '1, 1, 1, 0, 0, 0' if it is a significant spot except in the first three years of the sample period, but not after 2011. Hence, the label '0, 0, 0, 0, 0, 0' suggests that the corresponding region has never been a significant spot in any of the six years analysed in the study. Figure 5 indicates that 116 municipalities are identified in 10 types of significant spots (blue to dark blue), mostly in the provinces of Gauteng, Free State, KwaZulu-Natal, Limpopo, Northern Cape, and Western Cape, creating three clear belts of economic agglomerations in the country.

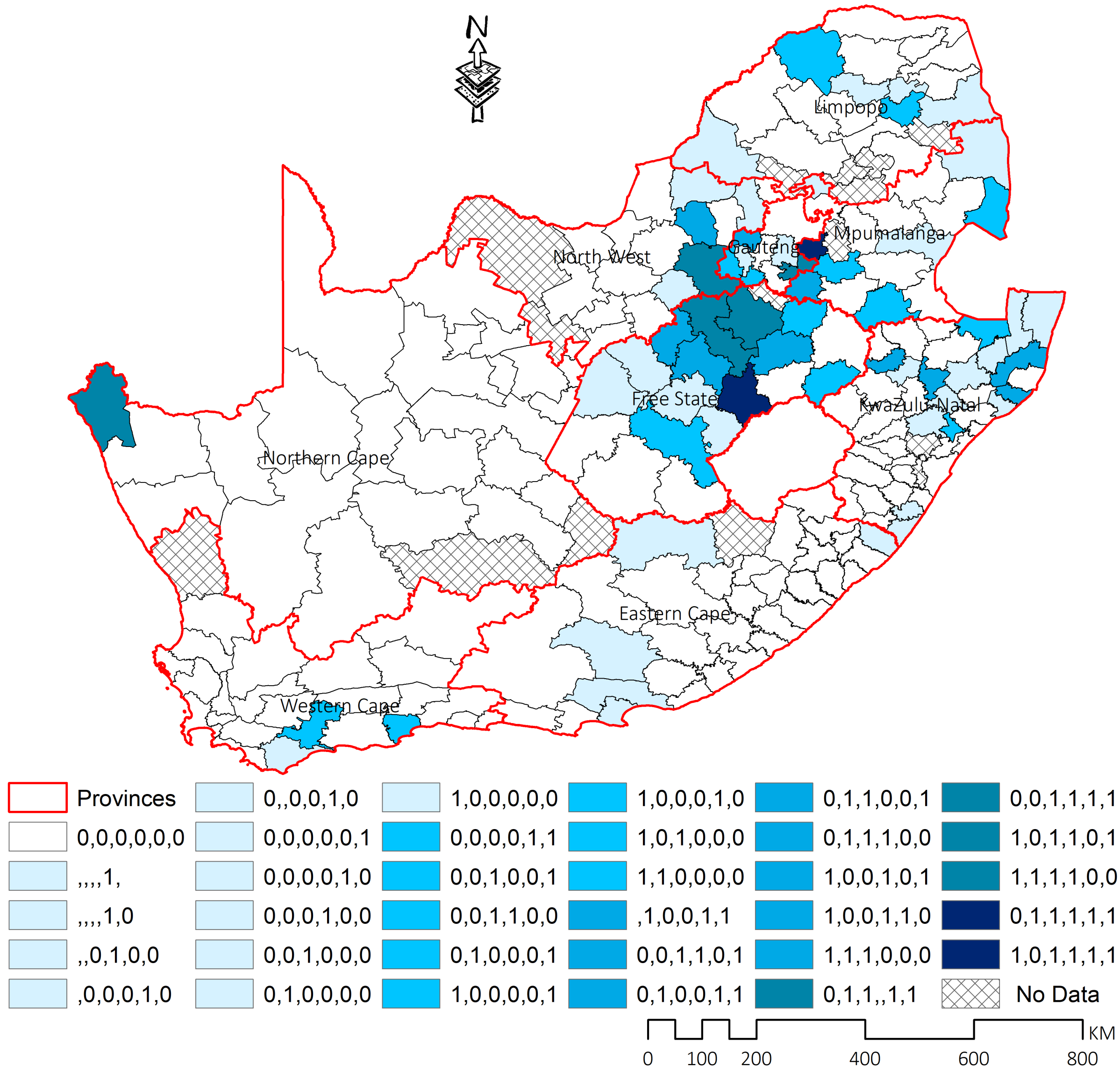
Figure 6 shows the spatial extent of significant TFP spots (hotspots) over time. Overall, 66 municipalities were identified in 34 types of hotspots (light blue to dark blue). A sizeable number of municipalities in provinces of Gauteng, Free State, KwaZulu-Natal, and Limpopo had high TFP returns in at least three of the years under study (light to dark blue),creating concentrated centres with high TFP returns.

Figure 5: Temporal significant spots of agglomeration (effective density), 2009–14



Source: authors' creation, based on data from the SARS–NT panel datset and Statistics South Africa Geography Metadata.

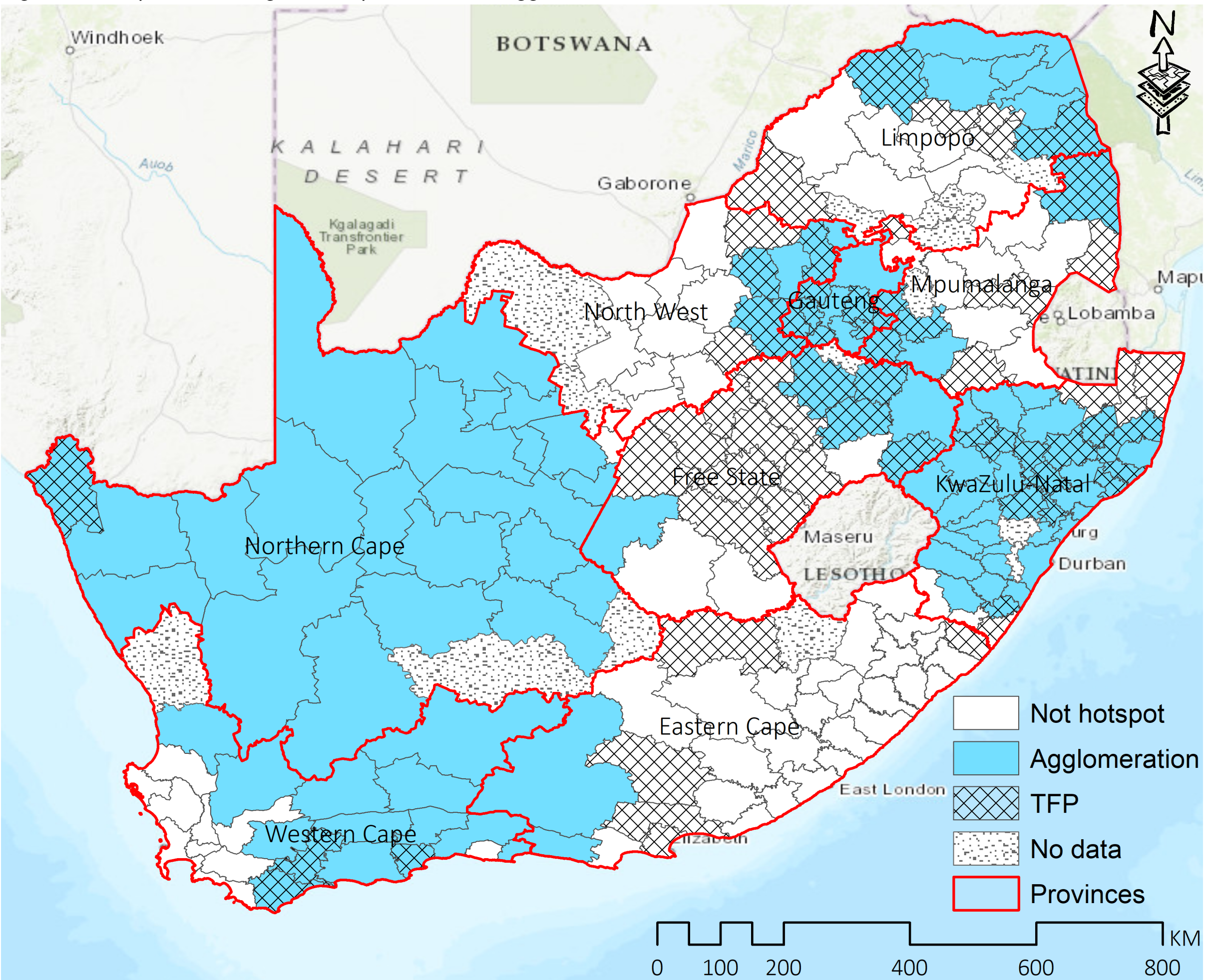
Figure 6: Temporal significant spots of TFP 2009–14



Source: authors' creation, based on data from the SARS–NT panel datset and Statistics South Africa Geography Metadata.

By extending the stability temporal mapping and overlaying significant spots for TFP and the measure of agglomeration economies, we obtain a map of comprehensive 'significant' spots. Figure 7 shows that with this overlap, three provinces—KwaZulu-Natal, Gauteng, and Free State—have a relatively higher number of municipalities with significant overlapping hotspots over the years. We also observe that the agglomeration measure has a stronger influence on the space-–time distribution of TFP in the municipalities where the two variables (TFP and agglomeration) have more coupled spatial units. Figure 7 also shows that some municipalities, especially in Northern Cape, have 'significant' agglomeration spots but not for TFP. The opposite is the case in Free State, where we have significant spots of TFP but not agglomeration. This could be mainly due to the sectors located in these regions, and their contribution to the overall TFP.

Figure 7: Comprehensive significant spots: TFP and agglomeration, 2009–14



Source: authors' creation, based on data from the SARS–NT panel datset and Statistics South Africa Geography Metadata.

The preliminary results obtained from the global and local indices of spatial autocorrelation suggests the presence of productivity spillovers originating from firms' agglomeration or clustering. This allows us to adopt a spatial panel model in estimating the space–time dynamics of the determinants of TFP in South Africa. As the spatial framework allows firm-specific characteristics (firm age) and local market conditions (concentration (i.e. HHI) and agglomeration (i.e. U_i)), to influence firms in neighbouring regions, their spatial lags are also included as regressors. With this inclusion, the version of Equation 14 to be estimated becomes:

$$\begin{aligned} \ln(tfp)_{it} = \lambda \mathbf{W}_{ij} ln(tfp)_{it} + \beta_1 Age_{it} + \beta_2 HHI_{it} + \beta_3 U_{it} \\ + \mathbf{W}_{ij} Age_{it} + \mathbf{W}_{ij} HHI_{it} + \mathbf{W}_{ij} U_{it} + \varepsilon_{it} \end{aligned} \tag{19}$$

where, as before, $\ln(tfp)_{it}$ represents estimates of aggregate TFP (for a generic firm i at time t), λ is the spatial autoregressive parameter associated with the spatial lag of TFP (i.e. $\mathbf{W}\ln(tfp)_{it}$), with $\mathbf{W}$ accounting for the weighted contributions of the productivity levels, firm age, market concentration, and agglomeration pertaining to firms in a neighbouring region j. ε_t is a pure idiosyncratic error term. The fitted version of Equation 19 is difficult to interpret as the estimated coefficients are a combination of direct and indirect effects. Direct effects provide the influence of the spatial unit on itself while indirect effects are spillover effects that capture the effects that spatial units have on other spatial units. From the fitted model, we split out these effects and summarize the results in Table 5.

Table 5: Estimates of the spatial model: all industries/sectors

Variable	(1) RE	(2) RE-II	(3) FE	(4) FE-II
Direct effect				
Agglomeration	0.04***	0.02	0.67	0.144
	(0.02)	(0.02)	(0.51)	(0.84)
HHI	0.09***	0.09***	0.08***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)
Firm age	0.06**	0.07**	0.04	0.05
	(0.03)	(0.03)	(0.03)	(0.03)
Spatial spillover (indirect) effects				
Agglomeration	–0.002	0.02	–0.06	0.49
	(0.002)	(0.01)	(0.05)	(0.43)
HHI	–0.004	0.04	–0.01*	0.01
	(0.003)	(0.06)	(0.01)	(0.02)
Firm age	–0.003	–0.02	–0.04	0.01
	(0.002)	(0.06)	(0.01)	(0.05)
Total effect				
Agglomeration	0.035***	0.035*	0.70	0.64
	(0.02)	(0.02)	(0.51)	(0.50)
HHI	0.08***	0.135***	0.11***	0.09***
	(0.01)	(0.03)	(0.03)	(0.02)
Firm age	0.06***	0.04	0.04	0.06
	(0.02)	(0.06)	(0.05)	(0.06)
Observations	1,128	1,128	1,128	1,128
Number of groups (spatial units)	188	188	188	188

Notes: standard errors in parentheses. RE denotes spatial random effects model while RE-II is a spatial random effects model that includes spatial lags for covariates and error term. FE denotes spatial fixed effects model while FE-II is a spatial fixed effects model that includes spatial lags for covariates and error term. ***, **, and * indicate significance at the 1, 5, and 10 per cent levels, respectively.

Source: authors' compilation based on the SARS–NT panel dataset.

From the estimates in Table 5, we note that direct effects are more pronounced than spillover effects. Across all model specifications, the direct effects of industrial clustering, market power, and firm experience in an industry or sector are positive. Specifically, across firms for which comprehensive data is available, a 1 per cent increase in a firm's market power will cause its productivity to increase by almost 0.1 per cent. The coefficient of market power (the HHI variable) is also statistically significant. The results also show that in the case of the spatial random effects model, a 1 per cent increase in firm age causes a statistically significant 0.07 per cent increase in firm productivity. This suggests that as firms mature in a particular industry/sector, they benefit from their greater business experience, established contacts with customers, and easier access to resources—that is, positive 'learning-by-doing effects'. The spillover effects of the independent covariates are mixed and largely insignificant. This suggests that market conditions and firm-specific characteristics in neighbouring spatial units do not impact productivity levels of firms. Combining both effects (i.e. the total effects case), we find that increases in firm clustering, the degree of a firm's market concentration, and duration of firm operation, have positive effects on productivity. These positive effects are statistically significant in the spatial random effects model where a 1 per cent increase in the clustering or agglomeration of a sector/industry a firm belongs to and its market share raises productivity by approximately 0.04 and 0.14 per cent, respectively.

As far as the sectoral composition of the dataset is concerned, we extend the analysis and present results for the manufacturing sector, which makes up over 30 per cent of firm classifications included in the tax administration database. Coefficient estimates presented in Table 6 indicate that for firms operating in the manufacturing sector, firm age is a critical driver of productivity. For the different model specifications, a 1 per cent increase in firm age results in productivity gains ranging between 0.12 and 0.15 per cent. In the spatial fixed effects model, increase in age of firms located in a neighbouring region generates a statistically significant and positive spillover effect (of 0.21 per cent). Across the different model specifications, the results show that when both direct and indirect effects are combined, a 1 per cent increase in firm age will cause firm productivity to increase by values ranging from 0.12 to 0.34 per cent.

Table 6: Estimates of the spatial model: manufacturing sector

Variable	(1) RE	(2) RE-II	(3) FE	(4) FE-II
Direct effect				
Agglomeration	0.01	0.01	0.63	0.35
	(0.02)	(0.02)	(0.69)	(0.51)
HHI	0.26	0.28	–0.01	0.01
	(0.20)	(0.21)	(12.54)	(15.53)
Firm age	0.12**	0.12***	0.15***	0.15***
	(0.03)	(0.03)	(0.03)	(0.03)
Spatial spillover (indirect) effects				
Agglomeration	0.0003	–0.01	0.03	0.18
	(0.0009)	(0.01)	(0.45)	(0.49)
HHI	0.01	0.05	–0.01	–0.03
	(0.01)	(0.09)	(12.26)	(15.89)
Firm age	0.004	0.07	0.01	0.21**
	(0.005)	(0.08)	(0.01)	(0.01)
Total effect				
Agglomeration	0.01	0.001	0.56	0.53
	(0.02)	(0.03)	(0.89)	(0.91)
HHI	0.266	0.33	–0.01	–0.01
	(0.21)	(0.28)	(24.80)	(31.42)
Firm age	0.12***	0.19**	0.34**	0.35***
	(0.03)	(0.08)	(0.06)	(0.09)
Observations	912	912	912	912
Number of groups (spatial units)	152	152	152	152

Note: standard errors in parentheses. RE denotes spatial random effects model while RE-II is a spatial random effects model that includes spatial lags for covariates and error term. FE denotes the spatial fixed effects model while FE-II is a spatial fixed effects model that includes spatial lags for covariates and error term. ***, **, and * indicate significance at the 1, 5, and 10 per cent levels, respectively.

Source: authors' compilation based on the SARS–NT panel dataset.

6 Conclusion

Using the comprehensive SARS–NT tax administration panel, this paper explores the space–time dynamics of TFP at the municipal level in South Africa from 2009 to 2014. The quantitative approach implemented in the paper consists of two independent components: estimating firm-level productivity and estimating the determinants of productivity using a SAR model. In brief, estimates of the production function using the Levinsohn and Petrin (2003) method indicate that intermediate inputs have the highest impact on firm productivity, a finding that is robust to estimations of production functions for subsectors of economic activity. In terms of distribution, productivity is relatively higher in firms located in regions/municipalities outside key economic clusters in Gauteng, KwaZulu–Natal, Western Cape, and parts of the Eastern Cape.

The geo-locational information contained in the SARS–NT panel allows us to analyse the determinants of firm-level productivity using a framework that accounts for the effects of firm-specific characteristics and local market conditions. Temporal stability mapping shows the distribution of municipal TFP data tends to follow a spatial pattern, with the agglomeration measure having a strong influence on the space-–time distribution of TFP with coupled spatial units or municipalities. With geographical dependencies confirmed by both local and global spatial indices of spatial autocorrelation, a SAR model is used to investigate the direct and indirect effects of agglomeration, firm age, and market concentration on firm-level productivity. Our results indicate that the direct effects of industry clustering, duration of firm existence, and share of market are more pronounced than spillover effects. This suggests that clustering with other firms, having increased market power, and extended period of existence/operations in a particular region have a greater impact on a firm's productivity than market conditions and firm-specific characteristics associated with firms located in neighbouring regions or municipalities.

Given the findings and ongoing revisions to the tax administrative database, future work will benefit from further geographical analysis focusing on specific sectors that are of key interest to industrial development policy and the evaluation of government policies around economic zoning.

References

Amusa, H., and T. Ngozo (2017). 'Role of Cities in Economic Growth: Industrial Diversity and Determinants of Economic Growth at Subnational Level'. In R. Mabugu (ed.), *The Financial and Fiscal Commission Technical Report on the Submission for the 2018/19 Division of Revenue*. Midrand: Financial and Fiscal Commission.

Anselin, L. (1995). 'Local Indicators of Spatial Association: LISA'. *Geographical Analysis*, 27(2): 93–115.

Arrow, K. (1962). 'The Economic Implications of Learning by Doing'. *The Review of Economic Studies*, 29(3): 155–173.

Beaudry, C., and A. Schiffauerova (2009). 'Who's Right, Marshall or Jacobs? The Localization Versus Urbanization Debate'. *Research Policy*, 38: 318–87.

Beveren, I. (2012). 'Total Factor Productivity Estimation: A Practical Review'. *Journal of Economic Surveys*, 26(1): 98–1282.

Boschma, R., and S. Iammarino (2009). 'Related Variety, Trade Linkages, and Regional Growth in Italy'. *Economic Geography*, 85(3): 289–311.

Brand, A. (2017). 'Corridor Development in Gauteng, South Africa'. *GeoJournal*, 82(2): 311–27.

Braunerhjelm, P., and B. Borgman (2004). 'Geographical Concentration, Entrepreneurship and Regional Growth: Evidence from Regional Data in Sweden, 1975–99'. *Regional Studies*, 38(8): 929–47.

Cody, D. (2006). 'The Adaptation and Advancement of the International Industrial District Model in Wenzhou: A Case Study of the Footwear Industry'. Master's thesis, Centre for East and South–East Asian Studies, Lund University. Available at: https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=1324976&fileOId=1324977.

Combes, P., G. Duranton, L. Gobbillon, D. Puga, and S. Roux (2012). 'The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection'. *Econometrica*, 80: 2543–94.

Davi, M., and I. Barbaccia (2015). 'Measuring Agglomeration by Spatial Effects: a Proposal'. *Ivista di Economia e Statistica Del Territorio*, 1: 44–70.

Ding, S., A. Guariglia, and R. Harris (2016). 'The Determinants of Productivity in Chinese Large and Medium–Sized Industrial Firms, 1998–2007'. *Journal of Productivity Analysis*, 45(2): 131–55.

Fedderke, J., and A. Wollnik (2007). 'The Spatial Distribution of Manufacturing in South Africa 1970–1996, Its Determinants and Policy Implications'. Economic Research Southern Africa Working Paper 53. Cape Town: Economic Research Southern Africa. Available at: https://www.econrsa.org/papers/w_papers/wp53.pdf.

Feng, C., and M. Wang (2019). 'Journey for Green Development Transformation of China's Metal Industry: A Spatial Econometric Analysis'. *Journal of Cleaner Production*, 225(10): 1105–17.

Gjeslvik, M. (2014). 'Capabilities for Innovation in a Globalizing World: from Nearby or a Distance?' *Entrepreneurial Business and Economic Review*, 2(2): 7–19.

Glaeser, E., H. Kallal, J. Scheinkman, and A. Shleifer (1992). 'Growth in Cities'. *The Journal of Political Economy*, 100(6): 1126–52.

Graham, D. (2007). 'Agglomeration, Productivity and Transport Investment'. *Journal of Transport Economics and Policy*, 41(3): 935–51.

Graham, D., P. Melo, P. Jiwattanakulpaisarn, and R. Noland (2010). 'Testing for Causality Between Productivity and Agglomeration Economies'. *Journal of Regional Science*, 50(5): 317–43.

Hafner, K. (2013). 'Agglomeration Economies and Clustering: Evidence from German and European Firms'. *Applied Economics*, 45(20): 2938–53.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Harris, R., and J. Moffat (2015). Plant-level determinants of total factor productivity in Great Britain, 1997–2006. *Journal of Productivity Analysis*, 44(1): 1–20.

He, M., Y. Chen, and R. Schramm (2017). 'Technological Spillovers in Space and Firm Productivity: Evidence from China's Electric Apparatus Industry'. *Urban Studies*, 55(11): 2522–41.

Hoover, E. (1937). *The Location Theory and the Shoe and Leather Industries*. Cambridge, MA: Harvard University Press.

Hoover, E. (1948). *The Location of Economic Activity*. New York: McGraw-Hill.

Jacobs, J. (1969). *The Economies of Cities*. New York: Random House.

Karlsson, C., B. Johansson, and R. Stough (2005). 'Industrial Clusters and Inter-Firm Networks: An Introduction'. In C. Karlsson, B. Johansson, and R. Stough (eds), *Industrial Clusters and Inter-Firm Networks*.Cheltenham: Edward Elgar.

Ke, S. (2010). 'Agglomeration, Productivity, and Spatial Spillovers across Chinese Cities'. *The Annals of Regional Science*, 45(1): 157–79.

Kelejian, H., and I. Prucha (2010). 'Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances'. *Journal of Econometrics*, 157(1): 53–67.

Krugell, W., and N. Rankin (2012). 'Agglomeration and Firm-Level Efficiency in South Africa'. *Urban Forum*, 23(3): 299–318.

Krugman, P. (1991). 'Increasing Returns and Economic Geography'. *Journal of Political Economy*, 99(3): 483–99.

Levinsohn, J., and A. Petrin (2003). 'Estimating Production Functions Using Inputs to Control for Unobservables'. *Review of Economic Studies*, 70(2): 317–42.

Lewis, D., and R. Bloch (1998). 'SDIs: Infrastructure, Agglomeration and the Region in Industrial Policy'. *Development Southern Africa*, 15(5): 727–55.

Louw, J. (2003). 'Rural–Urban Migration in South Africa'. In T. Falola and S. Salm (eds), *Globalization and Urbanization in Africa*. Asmara: Africa World Press Incorporated.

Mahofa, G. (2017). 'The Spatial Dimensions of Economic Activity in South Africa: The Role of Regional Policy, Crime and the Business Environment'. PhD thesis, School of Economics, University of Cape Town, Cape Town. Available at: https://open.uct.ac.za/bitstream/handle/11427/27835/thesis_com_2018_mahofa_godfrey.pdf?sequence=1.

Malmberg, A., B. Malmberg, and P. Lundequist (2000). 'Agglomeration and Firm Performance: Economies of Scale, Localisation, and Urbanisation among Swedish Export Firms'. *Environment and Planning A*, 32: 305–21.

Marshack, J. and W. Andrews (1944). 'Random Simultaneous Equations and the Theory of Production'. *Econometrica*, 12(3–4): 143–205.

Marshall, A. (1890 [1920]). *Principles of Economics*, 8th edition. London: Macmillan.

Martin, P., T. Mayer, and F. Mayneris (2011). 'Spatial Concentration and Plant-Level Productivity in France'. *Journal of Urban Economics*, 69(2): 182–95.

Matthee, M., N. Rankin, T. Webb, and C. Bezuidenhout (2018). 'Understanding Manufactured Exporters at the Firm–Level: New Insights from Using SARS Administrative Data'. *South African Journal of Economics*, 86(S1): 96–119.

Naudé, C. (2003). 'Industry Concentration in South Africa'. PhD thesis, Department of Economics, University of Pretoria, Pretoria. Available at: https://repository.up.ac.za/bitstream/handle/2263/27836/thesis.pdf?sequence=1&isAllowed=y.

Naudé, W., and W. Krugell (2006a). 'Are South Africa's Cities too Small'. *Cities: The International Journal of Urban Policy and Planning*, 20(3): 175–80.

Naudé, W., and W. Krugell (2006b). 'Economic Geography and Growth in Africa: The Determinants of Sub–national Growth in South Africa'. *Papers in Regional Science*, 85(3): 443–57.

Ohlin, B. (1933). *Interregional and International Trade*. Cambridge, MA: Harvard University Press.

Oliveau, S., and C. Guilmoto (2005). 'Spatial Correlation and Demography: Exploring India's Demographic Patterns'. Paper presented at the 25th IUSSP International conference, Tours, 18–23 July 2005. Available at: http://iussp2005.princeton.edu/download.aspx?submissionId=51529.

Olley, S., and A. Pikes (1996). 'The Dynamics of Productivity in the Telecommunications Equipment Industry'. *Econometrica*, 64(6): 1263–97.

Ord, J., and A. Getis (1995). 'Local Spatial Autocorrelation Statistics: Distributional Issues and an Application'. *Geographical Analysis*, 27(4): 286–306.

Parr, J. (2000). 'Agglomeration Economies: Ambiguities and Confusions'. *Environment and Planning A*, 34(4): 717–31.

Petrin, A., B. Poi, and J. Levinsohn (2004). 'Production Function Estimation in Stata Using Inputs to Control for Unobservables'. *The Stata Journal*, 4(2): 113–23.

Pieterse, D., E. Gavin, and C. Friedrich Kreuser (2018). 'Introduction to the South African Revenue Service and National Treasury Firm-Level Panel'. *South African Journal of Economics*, 86(S1): 6–39.

Porter, E. (1990). *The Competitive Advantage of Nations*. New York: The Free Press.

Porter, E. (1998). *On Competition*. Cambridge, MA: Harvard Business Review Books.

Rankin, N. (2001). 'The Export Behaviour of South African Manufacturing Firms'. Paper presented at Trade and Industrial Policy Strategies Forum 2001: New Directions in the South African Economy, 10–12 September 2001, Johannesburg. Available at: http://www.tips.org.za/files/The_Export_Behaviour_of_South_African_Manufacturing_Firms.pdf.

Romer, P. (1986). 'Increasing Returns and Long-Run Growth'. *Journal of Political Economy*, 94(5): 1002–1037.

Sangalli, I., and M. Lamieri (2015). 'A Spatial Econometric Model for Productivity and Innovation: The Role Played by Geographical and Sectorial Distances Between Firms'. Paper presented at the Sixth Italian Congress of Econometrics and Empirical Economics (ICEEE 2015), 21–23 January 2015, Salerno, Italy. Available at: http://paneldataconference2015.ceu.hu/Program/Sangalli-Lamieri.pdf.

Schmit, T. and J. Hall (2013). ‘Implications of Agglomeration Economies and Market Access for Firm Growth in Food Manufacturing’. *Agribusiness*, 29(3): 306–24.

Sheng, Y. and L. Song (2013). Re-estimation of Firms’ Total Factor Productivity in China’s Iron and Steel Industry. *China Economic Review*, 24: 177–88.

Simonis, D. (2002). ‘The New Economic Geography: A Survey of the Literature’. Belgian Federal Planning Bureau Working Paper 16-02. Brussels: Belgian Federal Planning Bureau. Available at: https://www.plan.be/admin/uploaded/200605091448041.WP0216en.pdf.

Stankov, U., T. Armenski, M. Klauco, V. Pavluković, M. Cimbaljević, and N. Drakulić-Kovačević (2017). ‘Spatial Autocorrelation Analysis of Tourist Arrivals Using Municipal Data: A Serbian Example’. *Geographica Pannonica*, 21(2): 106–14.

Syverson, C. (2010). ‘What Determines Productivity’. National Bureau of Economic Research (NBER) Working Paper 15712. Cambridge, MA: NBER. Available at: http://www.nber.org/papers/w15712.

Turok, I. (2012). ‘Urbanization and Development in South Africa: Economic Imperatives, Spatial Distortions and Strategic Responses’. International Institute for Environment and Development, United Nations Population Fund. Urbanization and Emerging Population Issues Working Paper 8. London: IIED. Available at: http://pubs.iied.org/10621IIED.html.

Turok, I., and J. Borel-Saladin (2013). The Spatial Economy: Background Research Report for the Integrated Urban Development Framework, August 2013. Available at: http://www.cogta.gov.za/cgta_2016/wp-content/uploads/2016/05/TUROK-SPATIAL-ECONOMY-DRAFT-FINAL.pdf.

von Thünen, J. (1826). *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Hamburg: Perthes. English translation by C.M. Wartenberg (1996). *von Thünen’s Isolated State*. Oxford: Pergammon Press.

Wang, Z., Y. Cheng, Y. Xinyue, and Y. Dennis Wei (2016). ‘Analyzing the Space–Time Dynamics of Innovation in China: ESDA and Spatial Panel Approaches’. *Growth and Change*, 47(1): 111–29.

Wooldridge, J. (2009). ‘On Estimating Firm–Level Production Functions Using Proxy Variables to Control for Unobservables’. *Economic Letters*, 104(3): 112–14.

Ye, Y., and L. Wu (2011). ‘Analyzing the Dynamics of Homicide Patterns in Chicago: ESDA and Spatial Panel Approaches’. *Applied Geography*, 31(2): 800–07.

Yu, D., and Y. Weie (2008). ‘Spatial Data Analyses of Regional Development in Greater Beijing, China, in a GIS Environment’. *Papers in Regional Science*, 87(1): 97–117.