WIDER Working Paper 2020/3

Turnin' it up a notch: how spillovers from foreign direct investment boost the complexity of South Africa's exports

Bjørn Bo Sørensen*

January 2020

United Nations University World Institute for Development Economics Research

wider.unu.edu

Abstract: Countries' economic complexity, and the associated diversification and sophistication of their exports, is a key determinant of economic growth. Understanding how South African firms learn to export more sophisticated products is, therefore, an important policy issue. Using administrative data covering the entire tax-paying population of firms in South Africa, we argue that foreign direct investment can stimulate export upgrading in manufacturing firms. We find that the level of sophistication of the most complex product exported by local firms increases in tandem with the presence of multinational enterprises located in upstream, supplying sectors within the same province. The study is the first within the associated literature to (i) provide firm-level evidence of export upgrading induced by foreign direct investment in Africa, (ii) employ the fitness algorithm to measure export complexity, and (iii) detect spillover effects from foreign direct investment materializing at the top line of domestic firms' export basket.

Key words: economic complexity, foreign direct investment (FDI), spillovers, export upgrading, manufacturing, South Africa

JEL classification: C38, D22, F61, O31

Acknowledgements: I would like to extend a warm thank you to John Rand for his invaluable guidance throughout this project. I would also like to thank the UNU-WIDER and National Treasury team in Pretoria, South Africa, for compiling the tax administrative data set: Amina Ebrahim, Charl van Schoor, Junior Chiweza, Humphrey Zondi, Michael Kilumelume, Carl Friedrich Kreuser, Aalia Cassim, and everyone else who has been involved in the project. I thank Beata S. Javorcik, Alessia Lo Turco, and Daniela Maggioni for their comments and for sharing their work on Hidalgo and Hausmann's complexity algorithm. I have also received valuable feedback and comments from Søren Post, Mikkel Pedersen, Jonas Knudsen, Lars Olaf Søvndahl, Sofia Torreggiani, and Inês A.R. Ferreira. Support from UNU-WIDER is gratefully acknowledged.

Note: unchecked appendix.

---

* Development Economics Research Group (DERG), University of Copenhagen, Denmark; nlh494@econ.ku.dk.

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 978-92-9256-760-6

https://doi.org/10.35188/UNU-WIDER/2020/760-6

Typescript prepared by Luke Finley.

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.

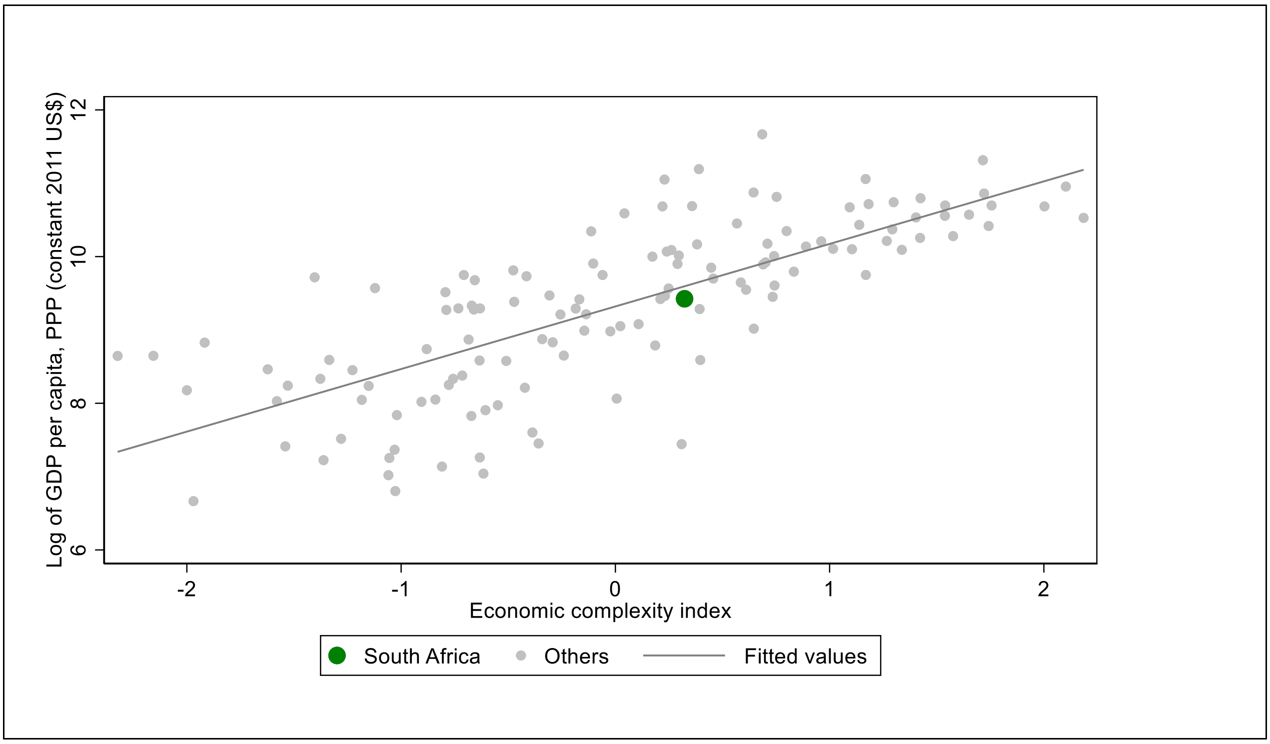
The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

In 1994, South Africa officially relinquished the chains of the Apartheid regime and embarked upon a new journey towards a brighter future. However, dismal economic performance has derailed South Africa's efforts to establish a society of inclusive growth and prosperity. The annual growth rate in GDP per capita has been far from impressive for an emerging economy, fluctuating around 2 per cent prior to the financial crisis in 2008/09 and plummeting to negative figures in recent years. Measured by the Gini coefficient, South Africa, at 0.63, is the world's most unequal society. Likewise, the unemployment rate of 27 per cent is among the highest in the world, and almost 60 per cent of the population live on less than US$5.5 a day (World Bank 2018).[1] Understanding how South Africa can break free from these worrying statistics is therefore a pressing policy issue.

Figure 1: Relationship between GDP per capita and economic complexity (2014)



Notes: economic complexity scores are calculated by applying Hidalgo and Hausmann's (2009) complexity algorithm to BACI world trade data at the HS6 level.

Source: author's illustration based on World Development Indicators (World Bank 2018) and BACI world trade data (Gualier and Zignago 2010).

The emerging academic field of economic complexity (EC) may hold part of the answer to South Africa's growth predicament. In brief, the complexity of a country's economy is defined by the pool of knowledge it holds and is able to combine for productive purposes. Therefore, it is possible to infer a country's EC

[1] The statistics are based on the World Bank's World Development Indicators from 2014.

from its ability to export a diverse set of sophisticated products (Hidalgo and Hausmann 2009).[2] Figure 1 presents the EC score for all countries of the world calculated using world trade data (the exact calculations will be explained later) plotted against GDP per capita. The correlation is remarkably clear. In fact, EC has also been shown to be a key determinant of economic growth—far superior to traditional growth predictors such as institutional quality, human capital, competitiveness, and financial depth (Hausmann et al. 2013).[3]

Seen in this light, South Africa's poor economic track record is intrinsically connected to the country's inability to diversify and upgrade its export basket and move away from trade in natural-resource-based primary commodities. As depicted in Figure 1, South Africa is relatively well positioned among the countries of the world—in terms of both EC and GDP per capita. The country's predicament lies in the fact that it has not been able to substantially advance its EC since the end of Apartheid. South Africa is stuck. In 1995, the country's total export volume was dominated by a few unsophisticated products: gold, platinum, and diamonds (19.6 per cent); iron ores and concentrates, coal, and refined petroleum oils (10.3 per cent); and various agricultural products (13.9 per cent) (Atlas of Economic Complexity 2019).[4] Nothing has changed since then. South Africa's export impasse is clearly illustrated in Figure 2, which shows the split between the country's exports in primary product sectors over time.

It has been forcibly argued elsewhere that an expansion of manufactured exports constitutes South Africa's main escape route from sluggish growth and high unemployment (Hausmann and Klinger 2006a; Rodrik 2008). But the manufacturing sector is shrinking (Kreuser and Newman 2018), and it is continuously focused on the export of low-skill, resource-based manufactured goods. The problem for South Africa, and many countries like it, is that surprisingly little is known about how EC evolves at the micro level, and what factors allow firms to upgrade their export basket (Javorcik et al. 2018). This paper argues that foreign direct investments (FDI) may play an important role by boosting export upgrading in South African manufacturing firms.

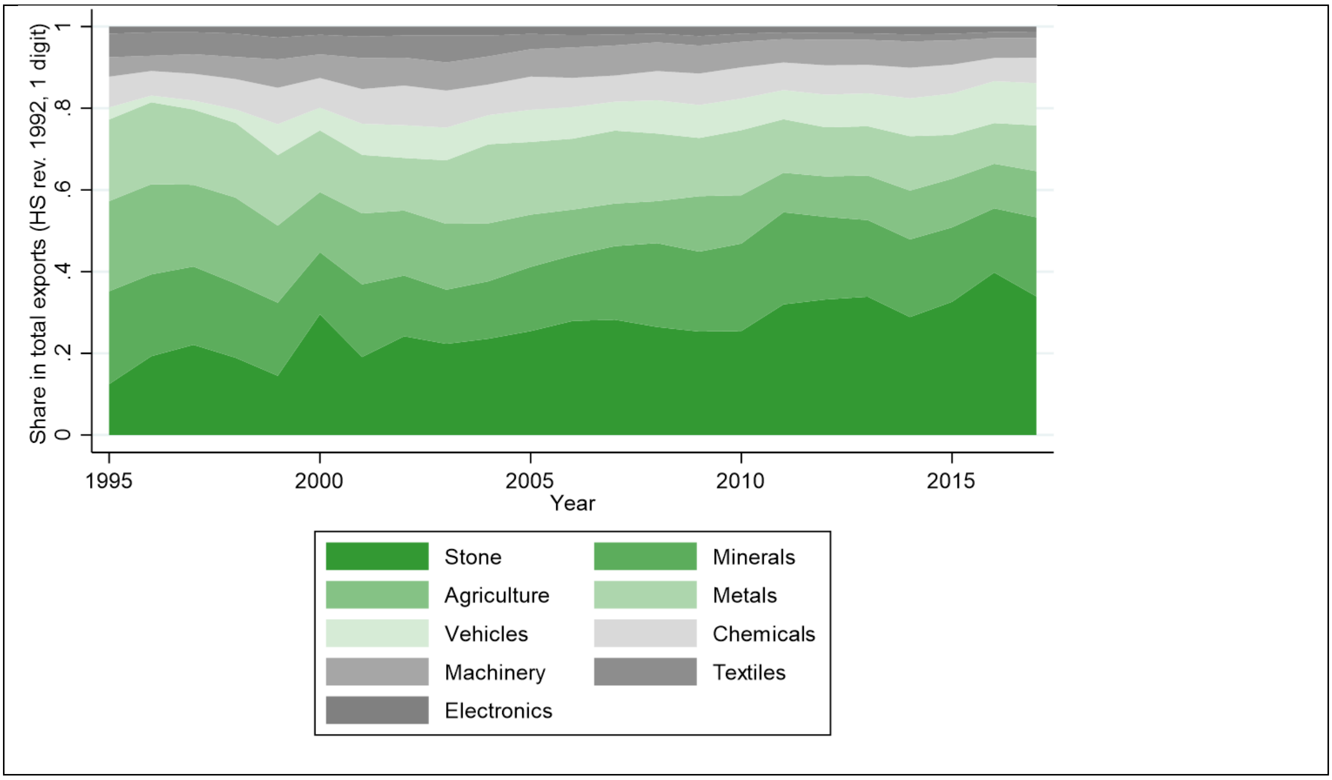
---

[2] In the same way, the sophistication of a product is understood as a function of how hard it is to produce. The complexity of a product can therefore be inferred from the number of countries able to produce it, weighted by the complexity of those countries (Hidalgo and Hausmann 2009).

[3] For example, the residuals depicted in Figure 1 are not random noise. Countries lying below the regression line are shown to be 'underperforming', comparing their low GDP per capita with the many capabilities they possess. Indeed, these countries tend to grow faster over time. Likewise, the countries located above the regression line are 'overperforming', and tend to grow at a slower pace.

[4] These numbers are calculated excluding services.

Figure 2: Split between South Africa's total natural-resource-based and manufactured exports (1995–2016)



Notes: products are grouped in accordance with the approach outlined in Harvard's online Atlas of Economic Complexity (2019). Product group 'Other' is left out of the figure. Split is calculated based on total export volume.

Source: author's illustration based on BACI world trade data (Gualier and Zignago 2010).

The argument draws on several contributions in the literature arguing that superior technology and knowledge brought by multinational enterprises (MNEs) can 'spill over' to indigenous firms in developing countries, thereby improving their economic performance (Caves 1974; Newman et al. 2015; Romer 1992). One strand of literature has already established a link between spillovers from MNEs and the export behaviour of local firms, by focusing on their *decision to export* and the *intensity* by which they do so (see for instance Aitken et al. 1997; Barrios et al. 2003; Benli 2016; Greenaway et al. 2004; Kneller and Pisu 2007; Kokko et al. 2001). However, it is a well-known fact in the empirical trade literature that 'the amount of exports matters, but the composition of exports matters more' (Moran 2010: 84), and it is therefore natural to take the export-focused FDI spillover literature beyond its emphasis on export entry and intensity. A small number of recent studies have thus begun to explore whether foreign firms can act as agents of structural change, boosting indigenous firms' ability to upgrade and diversify their production and export capabilities.[5] In a much-cited paper, Javorcik et al. (2018) find evidence that FDI flowing into downstream sectors in Turkey has enabled local manufacturing firms to increase the complexity of their newly introduced products. Likewise, Lo Turco and Maggioni (2019) show that regional discoveries of complex new goods in Turkey are driven by extra-regional knowledge introduced by MNEs. These studies focus, however, on product upgrading and diversification in general, and it is far from obvious that their conclusions can be extrapolated to the realm of exports. Yet a few studies have explored this issue and found FDI to be positively correlated with the introduction of new export varieties (Mayneris and Poncet 2015), improved export quality (Bajgar and Javorcik 2019), and higher

[5] An overview of these studies is provided in Table A1 in the appendix.

export sophistication (Eck and Huber 2016) among local manufacturing firms. This paper adds, to the best of our knowledge, several novel contributions to this literature.

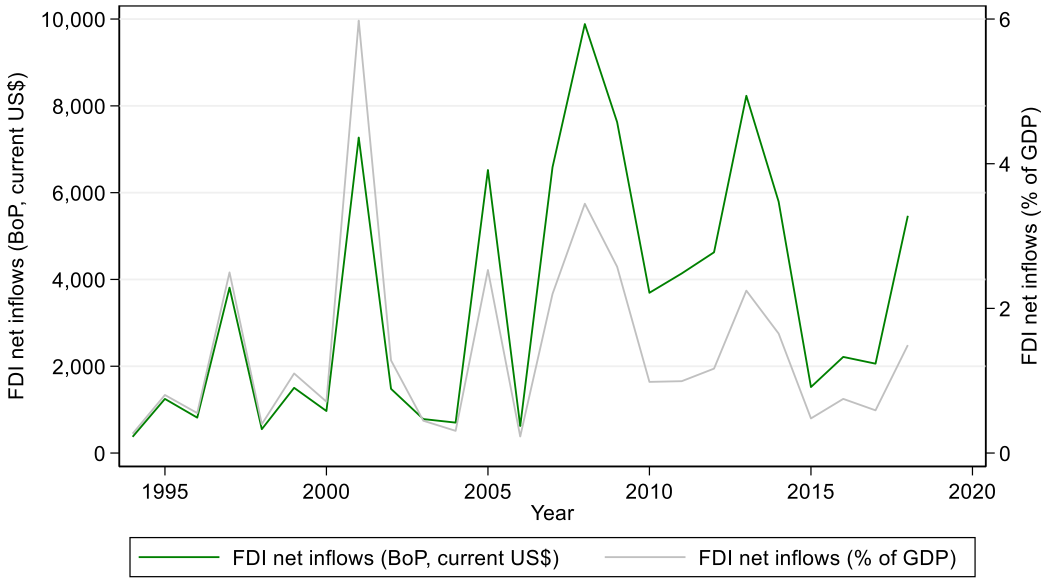
First, the study constitutes the first within the literature to be anchored in tax administrative data. Specifically, we rely on tax administrative panel data covering the entire population of tax-paying firms in South Africa, allowing us to identify foreign firms operating in the country from 2013 to 2016. We further link the data with South African customs data, national input–output (I–O) tables, and the BACI world trade database.[6]

Second, our empirical strategy follows and refines the novel approach introduced in Javorcik et al. (2018). We combine data from I–O tables with firm-level output information to generate proxies for the share of foreign firms operating in the same industry and region as domestic firms as well as in input-sourcing (downstream) and input-supplying (upstream) industries. Then we apply algorithms from the EC literature to world trade data in order to derive non-monetary measures of product complexity. While Javorcik et al. (2018) employ 'the complexity algorithm' developed by Hidalgo and Hausmann (2009), we refine their method with the application of 'the fitness algorithm' developed in Tacchella et al. (2012). The latter incorporates non-linearity in its foundational equation and is better aligned with the core assumptions in the complexity framework (Cristelli et al. 2013).

Finally, we find a positive and robust correlation between the level of sophistication of the most complex product exported by local firms (which we will henceforth call top-line complexity) and the presence of MNEs located in upstream, supplying sectors. Thereby, this study provides the first evidence of FDI-induced export complexity upgrading at the firm level on the African continent, and it is the first of its kind to detect a learning effect at the top line of domestic firms' exports. The finding suggests that FDI may constitute an important catalyst in the evolution of EC in South Africa, and it implies that the characteristics of the yearly FDI inflows depicted in Figure 3—the lack of a clear upward trend and high volatility—ought to be a policy concern. We further test for horizontal and backward spillovers flowing from foreign firms located in the same or downstream sectors, but find no evidence suggesting that these mechanisms matter in the case of South Africa. Nor do we find any indication that the absorptive capacity of local firms mediates their ability to exploit the benefits when allowing for heterogeneous effects of FDI.

[6] For further details, see Section 3.1. The CIT-IRP5 Panel of tax administrative data and South African customs data were accessed in a secure data lab in Pretoria; the data are described by Pieterse et al. (2018). For the national I–O tables, see Quantec EasyData (2019). For the BACI world trade data, see Gualier and Zignago (2010).

Figure 3: South African inward FDI flows (1994–2018)



Note: BoP = balance of payments.

Source: author's illustration based on World Bank (2018).

The remainder of the paper proceeds as follows. The next section outlines and reconciles the relevant theoretical and empirical literature in order to derive a conceptual framework and testable hypotheses. Section 3 describes our data and methodology in detail. We outline the main results in Section 4 and subject them to a plethora of robustness checks. In Section 5, we discuss the main findings, draw policy lessons, and conclude.

2 Conceptual framework

A prerequisite for testing the effect of FDI on export upgrading in local firms is to establish a theoretical link between these two components. In the following paragraphs, we develop a conceptual framework by integrating two models from the EC literature with the commonly used framework employed in the empirical FDI spillover literature. Our conceptual framework, depicted in Table 1, highlights the different mechanisms and channels through which export spillovers occur from FDI flow, and it allows us to develop a set of testable hypotheses.

2.1 Models from economic complexity

The EC of a country is defined as the pool of knowledge it holds and is able to combine for productive purposes. Hidalgo and Hausmann (2009) call this productive knowledge 'capabilities'. The idea is simple: having a larger number of capabilities allows a country to export a more diverse set of complex products. It follows that the answer to the riddle of economic growth lies in accumulating productive capabilities. The EC framework does not, however, provide any theory as to how productive capabilities emerge in the first place (Hidalgo and Hausmann 2009). Yet two models are commonly used to explain why developing countries find it difficult to accumulate the new capabilities necessary to foster structural transformation. The key, it is argued, lies in appropriation issues caused by product-related externalities.

The product space model

In the first model, Hausmann and Klinger (2006b) take as their point of departure a capability-enhancing spillover—an externality that improves the productive capacity of affected firms.[7] Their model builds on the idea of 'the product space'—a network connecting products through the complementary capabilities it takes to produce them. Products in close proximity rely on similar capabilities for their production. When an entrepreneurial firm makes an investment in the new capabilities necessary to introduce a new product, capability-enhancing *intra*-industry spillovers may simultaneously reduce the cost of 'jumping' into the same product for other firms. The model also allows for *inter*-industry spillovers: as the entrepreneurial firm moves into a new product, it may also be easier for emulators to move into other not-yet-produced products that share similar capabilities. Appropriation problems arise because the entrepreneurial firm is not able to internalize the societal value generated by its investment. The result is under-investment in new economic activities.

Within this product space framework, one can think of FDI as constituting a particular form of capability-enhancing investment. In fact, spillovers from FDI are likely to be of greater importance to structural transformation than the ones flowing from domestic investments, since foreign firms are commonly argued to bring superior productive knowledge and technologies to the host economy (Newman et al. 2015; UNCTAD 2005). It is not just that foreign firms can show domestic firms how to more efficiently do the things they already do. Rather, foreign firms may supply entirely new capabilities to an economy, enabling domestic firms to populate the product space as they learn to produce new and more complex goods (see for instance Moran 2018; Romer 1992).

The cost discovery model

In two seminal papers, Hausmann and Rodrik (2003) and Hausmann et al. (2007) develop the concept of cost discovery—an information spillover that reduces investment uncertainty and thereby improves the incentives for emulators to acquire new capabilities themselves.[8] The authors start from the observation that most innovation in developing countries occurs through the adaptation and adoption of existing products. Like the innovation of new products, this adaptation process is fraught with uncertainty because the underlying 'cost structure' of the local economy is unknown *a priori*. Hence, when an entrepreneurial firm tries to introduce a new product, it effectively engages in cost discovery by exploring the feasibility of local production. If the investment succeeds, the gains are socialized because information spillovers reduce investment uncertainty for emulators. If the investment fails, the loss remains private to the entrepreneur.

It is also possible to imagine an important role for foreign firms within the cost discovery framework. As argued by Javorcik et al. (2018), MNEs can be expected to have a higher propensity to succeed in cost discovery activities. First, experience of global production networks places foreign firms in a better position to evaluate the feasibility of the host country as a production site for a particular product. Second, large multinationals are financial powerhouses with the capability to venture into new industries

[7] Similar to what is commonly described as a 'true knowledge spillover' in the FDI literature.

[8] The model of cost discovery is, strictly speaking, not part of the complexity framework, but it is conceptually very close to the product proximity externality (Hidalgo 2011: 1) and it is applied in numerous papers linking FDI and product upgrading (Eck and Huber 2016; Harding and Javorcik 2012; Javorcik et al. 2018; Moran 2010).

despite the high fixed costs needed to develop multiple new capabilities simultaneously. Third, foreign affiliates may simply face lower costs in moving into new products because they have access to capabilities residing in their global value chain (GVC) that are not easily acquired through arms-length transactions by local firms. These propositions are supported by evidence suggesting that MNEs are more likely than their local competitors to introduce new products (Brambilla 2009).

2.2 Reconciling the EC models with the FDI spillover framework

The conceptual framework commonly adopted in empirical FDI spillover studies substantiates how externalities from foreign firms work at the micro level.[9] It builds on the assumption that FDI spillovers occur horizontally (*intra*-industry) and vertically (*inter*-industry) along the value chain through backward (downstream) and forward (upstream) linkages. It is further possible to categorize the spillovers discussed in the FDI spillover framework as (i) capability-enhancing, as discussed in the product space model; (ii) incentivizing, as in the cost discovery model; or (iii) outright negative. This makes it possible to reconcile the FDI spillover framework with the models discussed above and arrive at the conceptual framework depicted in Table 1.

Table 1: Conceptual framework: How FDI can impact export upgrading in domestic firms

Channels	**Export spillover mechanisms**		
	(1) Capability-enhancing	**(2) Incentivizing**	**(3) Negative**
Horizontal spillovers	Labour mobility Demonstration effect	Cost discovery Competition effect	Brain drain Crowd-out effect
Backward spillovers	Knowledge and technology transfer	Quality standards Demand for new intermediaries	Monopsonistic foreign customers (lock-in effect)
Forward spillovers	Embodied technologies Accompanying services	Supply of new and/or cheaper intermediaries	Monopolistic foreign suppliers (higher prices, lower quality)

Notes: the list of spillovers is not exhaustive, but it captures those most commonly highlighted in the literature.

Source: author's illustration.

First, the framework tells us that the capability-enhancing spillovers discussed in the product space model can occur in multiple ways (see Column 1 in Table 1). One commonly identified *horizontal spillover* mechanism is the demonstration effect. By observing their multinational competitors, domestic firms can learn how to improve the quality and standardization of their products and imitate effective marketing efforts. The diffusion of such knowledge can also be enhanced through labour mobility, as local workers hired by MNEs gain valuable on-the-job training and bring their new knowledge along when they return to local firms.[10] Capability-enhancing spillover may also occur through *backward linkages*, as foreign firms have an incentive to engage in direct knowledge and technology transfers to their local suppliers

[9] This widely adopted framework has been developed on the basis of a mix between predictions from theoretical work (Markusen and Venables 1999; Rodríguez-Clare 1996) and empirical evidence from case studies and surveys (see for instance Javorcik 2008; Moran 2001). Although the framework is most commonly applied to explain efficiency-enhancing spillovers, it has recently been adopted in studies investigating FDI's impact on product and export upgrading (Bajgar and Javorcik 2019; Eck and Huber 2016; Javorcik et al. 2018).

[10] A number of recent studies with access to matched employer–employee data have successfully shown that knowledge spillovers occur through this channel (Balsvik 2011; Görg and Strobl 2005; Poole 2013).

(Newman et al. 2015).[11] Finally, capability-enhancing *forward spillovers* occur when domestic firms learn by sourcing from multinational suppliers through embodied technologies and accompanying services.[12]

Second, the presence of MNEs can incentivize local firms to invest in the capabilities necessary to upgrade their export portfolio (see Column 2 in Table 1). As discussed above, foreign firms are more likely to introduce new products than their indigenous counterparts. This cost discovery process generates incentives for emulators to move into the same product and can therefore be conceptualized as a *horizontal externality*. The competition effect constitutes another often lauded horizontal externality of FDI. It occurs when foreign firms push their domestic competitors to upgrade their product portfolio. In terms of *backward spillovers*, multinationals may also incentivize local firms to invest in upgrading, as they impose higher quality standards on suppliers.[13] Additionally, as downstream MNEs introduce new complex final goods into the economy, they also generate demand for sophisticated new inputs, as described in the seminal model put forth by Rodríguez-Clare (1996). Similarly, the upstream presence of foreign firms can change incentive structures in downstream industries through *forward externalities*. For instance, supplying MNEs can make entirely new and better-quality inputs locally available, creating stronger incentives for domestic customers to exploit this new possibility for export diversification.[14]

Finally, the FDI spillover framework emphasizes that the impact of FDI on local firms need not be positive (see Column 3 in Table 1). *Inter-industry* labour mobility can cause 'brain drain' if local workers hired by multinationals never return to the local economy. Increased competition can crowd out domestic firms in the home market and hamper their ability to undertake the investments required to upgrade their export portfolio (Markusen and Venables 1999). *Backward externalities* can also be negative. A rich literature from the field of economic geography concerned with upgrading in GVCs highlights how unequal power relationships can enable monopsonistic multinationals to 'lock in' domestic suppliers, preventing functional upgrading (Gereffi et al. 2005). Finally, negative *forward externalities* arise in cases where foreign suppliers gain high market shares, exploiting their monopolistic position to increase prices and slacken on quality (Newman et al. 2015).[15]

2.3 Hypotheses

As is clear from Table 1, one should be careful making strong predictions about the sign and size of the effect of FDI on export upgrading *a priori*. However, it can be argued that positive spillovers are more likely to occur vertically than horizontally. On one hand, MNEs have an incentive to prevent valuable information from 'leaking' to competitors in their industry. On the other hand, multinationals face incentives to actively co-operate with, and transfer knowledge and technology to, their domestic

---

[11] For instance, a survey of Czech manufacturing firms indicates that 40 per cent of the domestic firms received assistance from their multinational customers, including personnel training (19 per cent) and help with quality assurance (10 per cent) and export opportunities (7 per cent) (Javorcik 2008).

[12] A range of studies document that access to a diverse set of high-quality inputs facilitates productivity increases, product innovation, and export quality upgrading. For example, Goldberg et al. (2013) find that increased access to intermediary imports has allowed Indian firms to expand their product portfolio domestically.

[13] Half of the Czech suppliers mentioned above reported that they had increased their product quality in order to supply to their MNE customers (Javorcik 2008).

[14] A more subtle but related point worth noting is that an inflow of FDI in upstream sectors also generates a competitive pressure in those industries, thus forcing domestic suppliers to upgrade their intermediary products and/or lower their price (Markusen and Venables 1999).

[15] For brevity we have only included the most important spillover mechanisms. Thus, the list highlighted here should not be taken as exhaustive.

customers and suppliers (Javorcik 2004). This observation has been empirically confirmed in the vast majority of spillover studies concerned with the effect on FDI on total factor productivity (TFP),[16] as well as in the small set of studies focusing on FDI-induced product and export upgrading among local firms (Bajgar and Javorcik 2019; Eck and Huber 2016; Javorcik et al. 2018). Based on these observations, we deduce two testable hypotheses in the context of South Africa:[17]

- **H_1**: Horizontal spillovers from foreign firms have no effect on domestic firms' ability to export more complex products.
- **H_2**: Vertical spillovers from foreign firms have a positive effect on domestic firms' ability to export more complex products.

3 Data and methodology

After first providing an overview of the data used, this section explains how we arrive at the key variables employed in the regression analysis. Specifically, it describes how we measure (i) FDI spillovers, (ii) product complexity, and (iii) firm-level export complexity. Finally, we put it all together in two regression models.

3.1 Data

The final data set used in the analysis is compiled from four different data sources: the CIT-IRP5 Panel of tax administrative data, South African customs data, world trade data from the BACI database, and South African I–O tables. A detailed description of the cleaning procedure for each data set is provided in Appendix Table A2.

First, tax administrative data from South Africa lie at the core of the analysis. They are collected on an annual basis by the South African Revenue Service (SARS) and cleaned at the South African National Treasury in collaboration with UNU-WIDER. The CIT-IRP5 Panel, covering the entire population of tax-paying firms in South Africa from 2008 to 2016, contains corporate income tax data (CIT) from the IT14 and ITR14 tax forms, with information on firms' income statements, balance sheets, and tax liabilities. The panel's employment figures come from IRP5 employee tax certificates. These figures are linked to the firm-level tax reference numbers in the CIT data using pay-as-you-earn (PAYE) codes.[18] Given the focus of this paper, we restrict the final sample to include only exporting South African manufacturing firms, leaving us with just under 5,500 firms over the period 2013–16.

[16] See Havranek and Irsova (2011) for a review of the literature.

[17] Table 1 presents 14 ways, grouped into nine conceptually distinct categories, in which such an impact can be theorized to materialize at the micro level. In an ideal world, one would test the effect of each of these causal mechanisms to understand whether spillovers actually occur and which mechanisms are important. However, operationalizing such a framework has proven econometrically difficult for economists because knowledge spillovers and economic incentives are tacit and difficult to quantify. As a consequence, it has become standard practice to treat the exact mechanisms through which spillovers occur—Columns 1, 2, and 3 in Table 1—as a 'black box'. Common econometric models simply adopt an outcome-based measure of the degree to which the effect of spillovers can be observed to occur through horizontal, backward, and/or forward channels (Görg and Strobl 2001). Adopting this approach, the set of hypotheses we are able to test boils down to a test for the presence of horizontal, backward, and forward spillovers.

[18] For a detailed discussion on the construction of the data set and the calculation of employment measures see Pieterse et al. (2018).

Second, we calculate product complexity scores using the BACI database compiled by the Centre d'Études Prospectives et d'Informations Internationales (CEPII; Gualier and Zignago 2010) and accessed through MIT's Observatory of Economic Complexity. The data contain product-level export information for all countries in the world at the HS6 level (2007 revision). On this data, we run the fitness algorithm of Tacchella et al. (2012) in order to calculate product complexity scores (for a detailed discussion, see Section 3.3).

Third, the product-level complexity scores are merged with South African customs data made available by SARS. This allows us to derive different firm-level export/import complexity scores for each exporting manufacturing firm that are linked to the variables in CIT-IRP5 Panel.

Finally, we use yearly South African I–O tables to construct the input–output coefficients necessary for the analysis of backward and forward spillovers (Quantec EasyData 2019; a detailed description of the procedure is provided below. The industry classification in the I–O tables does not align with the Standard Industrial Classification Rev. 7 (SIC7) in the CIT-IRP5 Panel, and no online correspondence table exists. The two data sets are therefore manually matched based on industry descriptions.[19] The match is not one-to-one, and some I–O industries match to two-digit SIC7 codes while others match to four-digit codes. Taking the industry classification in the I–O tables as the point of departure, we end up with a total of 90 industries, of which 39 belong to the manufacturing sector.[20]

3.2 Measuring FDI spillovers

In order to operationalize the conceptual framework depicted in Table 1 and capture the impact of the presence of foreign firms on the complexity of exports among South African manufacturing firms, we follow the standard approach introduced by Javorcik (2004) to proxy for horizontal, backward, and forward spillovers. The proxies are calculated at the province-sector-year level, thus exploiting variation in the presence of foreign firms over time and across South Africa's nine provinces and 39 manufacturing industries. The horizontal spillover measure is thereby given as the share of total sales accounted for by foreign firms in a given sector, province, and year:

$$Horizontal_{jpt} = \frac{\sum_{i=1}^{Njpt} Y_{it} \times ForeignFirm_{it}}{\sum_{i=1}^{Njpt} Y_{it}} \tag{1}$$

where, N_{jpt} is the number of firms in province p in industry j at time t. Y_{it} denotes the total sales of firm i at time t, whereas $ForeignFirm_{it}$ is a dummy variable indicating whether a firm is foreign. Here, it is important to note that there is no direct information on asset ownership in the CIT data, and it is therefore not straightforward to identify foreign firms in accordance with the Organisation for Economic Co-operation and Development's benchmark definition (OECD 2008). According to this definition, an FDI enterprise is a corporation (or quasi-corporation) at least 10 per cent of the operating assets of which are owned by a foreign investor. A detailed description of our approach to constructing the $ForeignFirm_{it}$ dummy and the potential consequences of measurement error are discussed in Appendix Table A3. The distribution of domestic and foreign firms across manufacturing industries is depicted in Appendix Table A4.

---

[19] The correspondence table is available upon request.

[20] Notice that the SARS tax year in, say, 2014 runs from 1 March 2013 to 28 February 2014. The tax year 2014 is thus best aligned with the calendar year 2013. Therefore, we lag all non-tax administrative data sources by one year.

To calculate the proxy for backward spillovers, we combine the measure of horizontal spillovers with coefficients calculated from the I–O data, indicating the share of each sector's total sales supplied to other sectors. The variable proxies for the presence of foreign firms in downstream sectors are given by:

$$Backward_{jpt} = \sum_{s=1, s\neq j}^{S} \alpha_{jst} \times Horizontal_{spt} \tag{2}$$

where α_{jst} denotes the share of output in manufacturing sector j supplied to sector s at time t.[21] The backward spillover proxy for industry j can be interpreted as a weighted average capturing the presence of foreign firms in all downstream sectors weighted by the share of industry j's output sold to each of these downstream sectors.

When calculating the proxy for forward spillovers, we adjust the measure of horizontal presence slightly, following Eck and Huber (2016). Specifically, we subtract exports from the total sales of foreign firms in a sector-province-year triad, $\sum_{i=1}^{Nspt}(Y_{it} - X_{it})$, to account for the fact that the exports of foreign firms cannot be purchased by domestic firms in sourcing sectors. The forward FDI proxy is then given by:

$$Forward_{jpt} = \sum_{s=1, s\neq j}^{S} \alpha_{jst} \times \frac{\sum_{i=1}^{Nspt}(Y_{it} - X_{it}) \times ForeignFirm_{it}}{\sum_{i=1}^{Nspt}(Y_{it} - X_{it})} \tag{3}$$

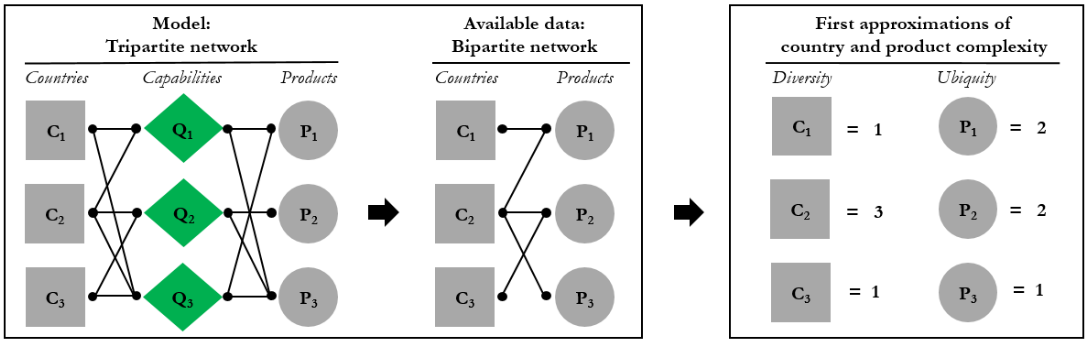
where α_{jst} denotes the share of total purchases of manufacturing sector j sourced from sector s. Thus, the forward-spillover proxy for sector j captures the presence of foreign firms in supplying sectors, weighted by the share of inputs sourced by sector j from other sectors.

3.3 Measuring product complexity

As already described, the theory of EC assumes that countries are linked to the capabilities with which they are endowed, and that these capabilities are linked in turn to the products that require them in production. Put in network lingo: countries, capabilities, and products are connected in a tripartite network as visualized in Figure 4. Yet the capabilities in the tripartite country-capability-product network are 'hidden'—it is only possible to observe a bipartite network linking countries to the products they produce. The econometric challenge is to extract information about the hidden capabilities from this bipartite network in order to infer the complexity of countries and products.

[21] Because the proxy for horizontal spillovers already captures intra-industry sourcing and supplying relationships, Equations 2 and 3 exclude these within-industry linkages.

Figure 4: The hidden capabilities layer



Source: author's illustration based on Cristelli et al. (2013).

Since Hidalgo and Hausmann (2009) first introduced the EC framework, the question of the best 'extraction procedure' has been the subject of heated debate. Located at the centre of this spectacle are Hidalgo and Hausmann's own complexity algorithm and the fitness algorithm developed in Tacchella et al. (2012), Cristelli et al. (2013), and Tacchella et al. (2013). On one hand, the overall idea is similar in both algorithms: one can measure the EC of a country by the range of products it is able to produce (the country's diversity) weighted by the sophistication of those products. The complexity of a product, on the other hand, can be estimated by the range of countries that are able to produce it (the product's ubiquity) weighted by the complexity of those countries (assuming that ubiquitous products are less complex). On the other hand, it has been repeatedly argued that the fitness algorithm is superior. First, it employs a superior weighting scheme that is mathematically better aligned to the theoretical assumptions of the complexity framework. Second, simulated toy models reveal that the fitness algorithm's complexity scores are better reflections of the underlying capabilities that they are intended to measure (Mariani et al. 2015; Tacchella et al. 2013).Therefore, we rely on the fitness algorithm for the primary analyses in this paper, but we also utilize the complexity algorithm for robustness checks.

To reiterate, we retrieve information on country–product links to set up the bipartite network from the BACI database. The data contain information on all countries' export values in all products defined at the HS6 level. In order to define whether a country is 'good' at producing and exporting a product, we employ Balassa's formula of revealed comparative advantage (RCA):

$$RCA_{cp} = \frac{X_{cp}}{\sum_c X_{cp}} \Big/ \frac{\sum_p X_{cp}}{\sum_{c,p} X_{cp}} \tag{4}$$

where X_{cp} is the total exports of country c in product p. Simply put, RCA_{cp} measures whether a country exports more than its 'fair share' of a product given the size of its export basket. We use the calculation of RCA_{cp} to construct a matrix, M_{cp}, connecting each country to the products it exports. The rows of M_{cp} represent different countries, and the columns different products. The cells in M_{cp} take the value 1 if a country exports a product with an RCA of above 1 and 0 otherwise:

$$M_{cp} = \begin{cases} 1 \;\; if \;\; RCA_{cp} > 1 \\ 0 \;\; if \;\; RCA_{cp} < 1 \end{cases} \tag{5}$$

In all its simplicity, the bipartite network structure encoded in M_{cp} contains the information needed to estimate the capabilities that different countries hold and that different products require. The fitness algorithm extracts information on the hidden-capabilities layer through a two-step procedure akin to the Google PageRank algorithm:

$$\begin{cases} \tilde{F}_{c,N} = \sum_p M_{cp} \times Q_{p,N-1} \\ \tilde{Q}_{p,N} = \dfrac{1}{\sum_c M_{cp} \dfrac{1}{F_{c,N-1}}} \end{cases} \rightarrow \begin{cases} F_{c,N} = \dfrac{\tilde{F}_{c,N}}{\langle \tilde{F}_{c,N} \rangle_c} \\ Q_{p,N} = \dfrac{\tilde{Q}_{p,N}}{\langle \tilde{Q}_{p,N} \rangle_p} \end{cases}, \tag{6}$$

where $F_{c,N}$ is the fitness of country c and $Q_{p,N}$ is the product complexity for product p after N iterations of the algorithm. $\tilde{F}_{c,N}$ and $\tilde{Q}_{p,N}$ are intermediate values of country fitness and product complexity, respectively.

In the first step, intermediate values of country fitness, $\tilde{F}_{c,N}$, and product complexity, $\tilde{Q}_{p,N}$, are calculated. $\tilde{F}_{c,N}$ is simply the export diversity of a country c weighted by the complexity of the products it produces, obtained from the previous iteration, $Q_{p,N-1}$. $\tilde{Q}_{p,N}$ is a weighted average of the ubiquity of product p, where the weight is the inverse of the complexity of countries producing that product. The non-linear weight, $\frac{1}{F_{c,N-1}}$, is larger for less complex economies. Thus, if many simple economies produce a given product, the algorithm assigns a lower complexity score to $\tilde{Q}_{p,N}$.

In the second step, the intermediate values are normalized as they are divided by the mean of country fitness, $\langle \tilde{F}_{c,N} \rangle_c$, and product complexity, $\langle \tilde{Q}_{p,N} \rangle_p$. The initial conditions to start the algorithm are $\tilde{Q}_{p,0} = 1 \, \forall \, p$ and $\tilde{F}_{c,0} = 1 \, \forall \, c$. We stop the algorithm after 50 iterations.

Despite its superiority, the fitness algorithm entails a few drawbacks compared with the complexity algorithm. First, the non-linear weighting scheme means that product complexity scores have 'an intrinsically higher degree of volatility' (Cristelli et al. 2013: 14). The product complexity scores both are less robust to flaws in the underlying trade data and respond strongly if a non-complex economy suddenly adds a new product to its export basket. We take this concern seriously, responding to it by using the average complexity score for each product over the entire sample period (2013–16).

3.4 Measuring firm-level export complexity

We calculate three measures of *firm*-level export complexity based on the *product*-level complexity scores derived above. First, it is likely that the sophistication of a firm's export basket increases solely due to the introduction of new, high-complexity exports. To measure such a phenomenon, we follow Javorcik et al. (2018) and calculate the weighted average of newly introduced exports, EC_{it}^{new}, by firm i at time t:

$$EC_{it}^{new} = \sum_{p=1}^{P_{it}^{new}} Q_{p,50} \times \frac{Y_{ipt}}{\sum_{p=1}^{P_{it}^{new}} Y_{ipt}} \tag{7}$$

where P_{it}^{new} denotes the number of new exports introduced by firm i at time t. Since we have customs data as far back as 2010 for domestic manufacturing firms, we define product p as new if it has not been exported by firm i in any previous period since 2010. $Q_{p,50}$ denotes the complexity level of product p after 50 iterations of the fitness algorithm. Y_{ipt} denotes the exported value of product p by firm i at time

t. Y_{ipt} is used to weight the complexity of each new product p by its monetary share in the entire basket of new exports introduced by firm i.

Export upgrading may also be observed as a shift in the average complexity of a firm's entire export basket (Eck and Huber 2016). Therefore, we replicate Equation 7 but sum over all products in the firm's export basket, to calculate EC_{it}^{all}:

$$EC_{it}^{all} = \sum_{p=1}^{P_{it}^{all}} Q_{p,50} \times \frac{Y_{ipt}}{\sum_{p=1}^{P_{it}^{all}} Y_{ipt}} \tag{8}$$

where P_{it}^{all} denotes the number of all products exported by firm i at time t.

Third, we generate a new estimate of firm-level export complexity that simply measures the complexity of the most sophisticated product exported by firm i at time t. Formally, we define P_{it} as the set of products exported by firm i at time t. Then, for each firm, we define:

$$EC_{it}^{topline} = max\{Q_{pit,50}\} \tag{9}$$

where $Q_{pit,50}$ is the complexity of product $p \in P_{it}$. The measure is built on the intuition that the productive capabilities a firm holds are likely to be revealed by the most complex product it is able to export. In other words, the learning effects from FDI are likely manifested at the tail of a firm's product complexity distribution. [22]

3.5 Empirical strategy

Model 1: OLS with fixed effects

When EC_{it}^{all} or $EC_{it}^{topline}$ constitute the dependent variables, we employ an OLS specification with industry, province, year, industry-year, and province-year fixed effects. Because the spillover proxies vary at the sector-province level, we use robust standard errors clustered at this level as well. The regression equation is then given by:

$$\begin{aligned} EC_{it} = \beta_0 + \beta_1 Horizontal_{jpt-1} + \beta_2 Backward_{jpt-1} + \beta_3 Forward_{jpt-1} + \\ \boldsymbol{\beta' Controls_{it-1}} + \alpha_j + \delta_p + \mu_t + \theta_{jt} + \tau_{pt} + \varepsilon_{it} \end{aligned} \tag{10}$$

where EC_{it} denotes the log of either EC_{it}^{all} or $EC_{it}^{topline}$, and measures the export complexity for firm i at time t. Because $Horizontal_{jpt-1}$, $Backward_{jpt-1}$, and $Forward_{jpt-1}$ are ratios between 0 and 1 and EC_{it} is in logarithmic form, their coefficients can be interpreted as semi-elasticities approximating

[22] Note that a few product complexity scores converge towards 0 as we iterate the fitness algorithm. Taking the log of these values, as we do in the regression analysis, creates very large negative numbers. As a consequence, we leave out the bottom 1 per cent of the $EC_{it}^{topline}$ scores. Robustness checks, not reported here, show that this procedure does not significantly change the results. As a further robustness check, we also recreated the regression analysis the bottom outliers removed from EC_{it}^{new} and EC_{it}^{all}. These results are not reported here either, since they do not change our results in any meaningful way.

the percentage change in export complexity given a one-percentage-point increase in the ratio of foreign firms operating in the same, downstream, or upstream sectors.

The set of control variables included in $\boldsymbol{Controls_{it-1}}$ can be separated into three groups. In line with Javorcik et al. (2018), we first include a set of standard variables controlling for a firm's ability to produce and export new products as well as its absorptive capacity. This set includes the size of the labour force ($Size_{it-1}$),[23] value added per employee ($LabourProductivity_{it-1}$),[24] R&D spending per employee ($R\&DIntensity_{it-1}$),[25] and labour cost per employee ($Wage_{it-1}$).[26] In the regression analysis, we use the logarithmic transformation of these variables, obtained as $log(var + 1)$. The second set of controls takes account of the potential for learning by exporting proxied through the number of countries a firm exported to in the previous period ($CountryDiversification_{it-1}$) and the number of products it exported ($ProductDiversification_{it-1}$).[27] The third set of controls comprises two complexity measures. We include the lagged value of the average complexity of each firm's export basket (EC_{it-1}^{all}) as an outcomes-based measure of firm capabilities. We also calculate the average complexity of each firm's import basket (IC_{it-1}^{all}) to control for the learning potential embodied in complex imported products.

We lag all explanatory variables by one year. First and foremost, this is done as a precautionary measure against any simultaneity issues. One could, for instance, imagine that the ability of domestic firms, expressed by the complexity of their export basket, would induce foreign customers to locate in downstream industries. Second, lagging the three spillover variables acknowledges the fact that it takes time for spillovers to be accommodated into the operations of domestic firms and translated into new export products. As a consequence, the explanatory variables are observed in the tax years 2013–15, while the dependent variables run from 2014 to 2016.

In addition to the firm-level controls, we include industry and province fixed effects, α_j and δ_p, to control for time-invariant unobservables that impact the export capabilities of local firms and the inflow of FDI in a particular industry or province. The inclusion of industry-year and sector-year fixed effects, θ_{jt} and τ_{pt}, extends the set of controls to account for time-varying shocks at the industry and province level, respectively. Finally, year fixed effects, μ_t, are included to account for any economy-wide shock driving both FDI inflows and export upgrading.

Despite the use of fixed effects and lagged control variables, the econometric specification presented in Equation 10 does not fully eliminate the risk of endogeneity and reverse causality. On one hand, the model would ideally include firm fixed effects to control for unobservable time-invariant factors that might simultaneously correlate with the three spillover proxies and export complexity. However, the fixed effects estimator has proved infeasible, since more than one-fourth of the data are lost when it is implemented. On the other hand, identification based on Equation 10 will suffer from reverse causality

23 The firm-level employment measure is based on the number of individuals reporting *any* source of income derived from a firm in the IRP5 tax forms.

24 We control for the average labour productivity in each firm based on the intuition that more productive firms find it easier to overcome the fixed costs of exporting (Melitz 2003).

25 The measure of R&D intensity is included to control for a firm's ability to develop new, or adopt and adapt existing, complex export goods.

26 The measure of a firm's average wage proxies for the average ability of its workforce.

27 For a discussion on the topic see Clerides et al. (1998).

if foreign firms are able to anticipate rising complexity among domestic firms and strategically locate in downstream or upstream sectors in order to exploit an associated future surge in the supply of complex domestic inputs or a rising domestic demand for MNE-produced inputs. In practice, however, such a scenario is unlikely, as it would require foreign firms to possess information at an implausible level of detail (Javorcik et al. 2018).

Model 2: Heckman selection model

While the model specified in Equation 10 is feasible with EC_{it}^{all} and $EC_{it}^{topline}$ as dependent variables, it fails to acknowledge the self-selection issue arising when EC_{it}^{new} constitutes the left-hand side of the equation. EC_{it}^{new} measures the complexity of new exports introduced by a firm, but not all exporters introduce a new product in every period; thus, we do not observe the dependent variable for all firms in the population of interest (all exporting manufacturing firms). In effect, this makes the sample non-random, because firms self-select into introducing new exports. We tackle this problem by estimating a two-step Heckman selection model.[28]

The first stage models an exporting firm's decision on whether to export a new product through a probit model. Here, we include the same variables as in Equation 10, but we also introduce two exclusion restrictions to circumvent the risk that collinearity between the correction term and the regressors in the second stage will inflate the standard errors. In our case, the exclusion restrictions should have an impact on firm i's decision to export new goods, but it should not be correlated with the complexity of the new goods that firm introduces. Our first exclusion restriction is the weighted average distance from South Africa to each country that firm i exports to ($Distance_{it}$). Each distance is weighted by the share of a firm's total exports sold in each country. Country distances are obtained from the online CEPII database (Mayer and Zignago 2011) and reported in kilometres. The intuition is as follows: it is likely to be less costly for a firm to introduce a new product in markets where it has already gained a foothold. Thus, switching to a new export destination involves fixed costs. Therefore, a firm's initial choice of export destination creates a lock-in effect. If the firm is locked in to destinations further away, it faces higher fixed costs to introduce a new product. Firms with 'high-distance export baskets' can therefore be expected to introduce fewer new exports, *ceteris paribus*. The second exclusion restriction ($DestinationGDP_{it}$) models an exogenous demand shock in the countries to which firm i exports. If the firm exports to destinations where the country's GDP is increasing, it should be more likely to introduce a new product in that market in order to capitalize on the rising purchasing power.[29]

In the second stage, we repeat the regression from Equation 10, but with EC_{it}^{new} as the dependent variable and with the inverse Mills ratio, $\widehat{\lambda}_{it}$, included on the right-hand side to control self-selection:

28 This approach makes it impossible to cluster standard errors, and the results should thus be interpreted with this reservation in mind.

29 While $Distance_{it}$ and $DestinationGDP_{it}$ can be expected to influence a firm's ability to introduce a new product, it is unlikely that the variables affect the complexity of a firm's new exports, i.e. they seem like feasible exclusion restrictions. However, even though we believe that $DestinationGDP_{it}$ and $Distance_{it}$ 'fit the bill', within-firm capabilities may influence firms' ability to both introduce complex export goods and ship them to distant countries. We explicitly tested this assumption by including $Distance_{it}$ and $DestinationGDP_{it}$ in a regression based on the model from Equation 10, but with EC_{it}^{new} as the dependent variable. Neither of the exclusion restrictions were significant, thus confirming our initial intuition.

$$EC_{it}^{new} = \beta_0 + \beta_1 Horizontal_{jpt-1} + \beta_2 Backward_{jpt-1} + \beta_3 Forward_{jpt-1} + \boldsymbol{\beta' Controls_{it-1}} + \widehat{\lambda}_{it} + \alpha_j + \delta_p + \mu_t + \theta_{jt} + \tau_{pt} + \varepsilon_{it} \quad (11)$$

4 Main results

4.1 Descriptive statistics

Table 2, Column 1, reports the mean values of the variables included in Equations 10 and 11 for domestic exporters. These are compared with the mean values for foreign firms, reported in Column 2, through a t-test of the mean differences reported in Column 3.

Table 2: t-test of mean differences between key variables for domestic exporters and foreign firms

	Mean Domestic exporters	**Mean** Foreign firms	**Difference**
Dependent variables			
EC_{it}^{new}	−0.7084	−0.6503	−0.0581**
EC_{it}^{all}	−0.9036	−0.8363	−0.0673***
$EC_{it}^{topline}$	−0.0804	0.0150	−0.0954***
Spillover proxies			
$Horizontal_{jpt-1}$	0.3094	0.3301	−0.0207***
$Backward_{jpt-1}$	0.0340	0.0339	0.0001
$Forward_{jpt-1}$	0.0319	0.0319	−0.0000
Controls			
$Size_{it-1}$	3.6491	2.9164	0.7327***
$LabourProductivity_{it-1}$	12.4729	12.2747	0.1983***
$R\&DIntensity_{it-1}$	0.5150	0.1517	0.3634***
$Wage_{it-1}$	11.5693	11.4232	0.1461***
$CountryDiversification_{it-1}$	5.1369	7.0807	−1.9439***
$ProductDiversification_{it-1}$	8.2338	10.7941	−2.5603***
EC_{it-1}^{all}	−0.9141	−0.8355	−0.0786***
IC_{it-1}^{all}	0.8590	0.8800	−0.0210

Notes: all variables except spillover proxies, $CountryDiversification_{it-1}$ and $ProductDiversification_{it-1}$ are reported in logs. *** p<0.01, ** p<0.05, * p<0.1

Source: author's calculations based on SARS data described in Pieterse et al. (2018).

It is clear from Table 2 that foreign firms, given that they export, export more complex goods than their domestic counterparts—in all three ways of measuring this. Additionally, the complexity of imported products IC_{it-1}^{all} is much higher than the complexity of exported products EC_{it-1}^{all}—for both domestic and foreign firms. This confirms that the manufacturing firms in South Africa are reliant on complex imports. Furthermore, it can be seen that foreign exporters are more diversified than domestic exporters, in terms of both the number of destinations they export to and the number of products they export.[30] The table also reports spillover values for foreign and domestic firms. Here, the only statistically significant difference in mean values between the two groups is found in the horizontal proxy indicating that foreign firms tend to cluster together within the same industry. Finally, and perhaps surprisingly, domestic exporters tend to be bigger, to be more productive and R&D-intensive, and to pay higher wages than foreign firms.

4.2 Regression results from Model 1

The regression results from Equation 10 are presented in Table 3. Columns 1–4 display the regression coefficients with EC_{it}^{all} as the dependent variable, while Columns 5–8 display the results when $EC_{it}^{topline}$ constitutes the main variable of interest. The proxies for horizontal, downstream, and upstream FDI enter first one by one and then all together.

When EC_{it}^{all} is the dependent variable, we find no evidence of spillovers. All spillover coefficients are positive, but insignificant—both on their own and when included together. This result suggests, contrary to our expectations and the evidence found for India by Eck and Huber (2016), that the average sophistication of domestic firms' export basket is not significantly affected by the presence of foreign competitors, customers, or suppliers.

When $EC_{it}^{topline}$ constitutes the explained variable, the estimate of forward spillovers is positive and significant—both on its own and when it enters the regression along the two other spillover proxies. The coefficient of $Forward_{jpt-1}$ indicates that a one-percentage-point increase in the ratio of foreign firms operating in upstream sectors is correlated with a 5.4 per cent increase in the complexity of domestic manufacturing firms' most sophisticated export good, on average. Given the mean value of 0.031 for $Forward_{jpt-1}$ reported in Table 2, a one-percentage-point increase in the ratio of foreign firms in upstream industries corresponds to an increase of approximately 30 per cent. The effect size is large when compared with the estimated coefficients of the control variables in Column 8. For instance, doubling the size of your workforce is associated with only a 4.9 per cent increase in top-line complexity. Likewise, a 100 per cent increase in labour productivity correlates with only a 2 per cent increase in top-line complexity. According to these estimates, attracting FDI seems to be a very effective tool in boosting top-line product complexity of South African manufactured exports.

30 A few very large and diversified firms drive the high mean values of the product and country diversification variables.

Table 3: Main results, EC_{it}^{all} and $EC_{it}^{topline}$—OLS

	EC_{it}^{all}				$EC_{it}^{topline}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Horizontal$_{jpt-1}$	0.012			0.013	−0.067			−0.065
	(0.056)			(0.057)	(0.076)			(0.073)
Backward$_{jpt-1}$		1.963		1.304		1.561		0.366
		(1.545)		(1.657)		(2.100)		(2.183)
Forward$_{jpt-1}$			1.864	1.582			5.439***	5.434***
			(1.228)	(1.292)			(1.446)	(1.420)
Size$_{it-1}$	−0.023***	−0.023***	−0.023***	−0.023***	0.049***	0.051***	0.050***	0.049***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.011)	(0.011)	(0.011)	(0.011)
LabourProductivity$_{it-1}$	−0.012	−0.012	−0.012	−0.012	0.021*	0.022*	0.022*	0.020*
	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.011)	(0.011)	(0.011)
R&DIntensity$_{it-1}$	0.006	0.006	0.006	0.005	0.008	0.009*	0.009*	0.008
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Wage$_{it-1}$	0.007	0.007	0.007	0.007	0.022**	0.022**	0.022**	0.022**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.010)	(0.010)	(0.010)	(0.010)
CountryDiversification$_{it-1}$	−0.000	−0.000	−0.001	−0.000	0.019***	0.019***	0.019***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
ProductDiversification$_{it-1}$	0.000	0.000	0.000	0.000	0.017***	0.017***	0.017***	0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)
EC_{it-1}^{all}	0.695***	0.696***	0.696***	0.695***	0.378***	0.378***	0.379***	0.379***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.020)	(0.020)	(0.020)	(0.020)
IC_{it-1}^{all}	0.070***	0.070***	0.071***	0.070***	0.073***	0.075***	0.076***	0.073***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.026)	(0.026)	(0.026)	(0.026)
Fixed effects								
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,485	5,530	5,530	5,485	5,442	5,487	5,487	5,442
R-squared	0.641	0.644	0.644	0.642	0.417	0.415	0.416	0.418

Notes: all variables except spillover proxies, $CountryDiversification_{it-1}$ and $ProductDiversification_{it-1}$ are reported in logs. Robust standard errors, clustered at the province-industry level, in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Source: author's calculations based on SARS data described in Pieterse et al. (2018).

What might this effect size mean from a macroeconomic perspective? One way to interpret the result is by linking GDP per capita to country-level top-line complexity. In Table A5 in the Appendix, we regress the log of GDP per capita for all countries of the world on the log of the most complex product produced by each country (the country-level equivalent to $EC_{it}^{topline}$) for the year 2014.[31] As expected, the relationship is not nearly as strong as the one between EC and GDP per capita depicted in Figure 1 (R-

[31] The corresponding scatter plot to this simple correlation is shown in Figure A3 in the Appendix.

squared of 0.12 vs 0.57), but the regression does indicate a positive relationship. Specifically, a 10 per cent increase in country-level top-line complexity is correlated with a 3.7 per cent increase in GDP per capita. According to the results we have presented thus far, a 10 per cent increase in firm-level top-line complexity in the manufacturing sector can be achieved by less than a two-percentage-point increase in the presence of foreign firms in downstream industries.

Two additional points are worth noting with respect to the control variables. First, all regressors except $R\&DIntensity_{it-1}$ are positive and statistically significant in the second model—just as predicted.[32] Second, the results from estimating Equation 10 with $EC_{it}^{topline}$ as the dependent variable and adding the three different sets of control variables one by one show that the backward spillover proxy is significant when EC_{it-1}^{all} and IC_{it-1}^{all} are excluded from the regression (see Table A6 in the Appendix). This suggests that foreign firms tend to locate in industries whose domestic suppliers are capable of producing high-complexity products and have access to complex imports. Failing to control for these factors can thus lead to significant omitted-variable bias.

4.3 Regression results from Model 2

Table 4 presents the results from the first and second stage of the Heckman selection model. Overall, we find no evidence supporting the hypothesis that the presence of foreign firms in the same, downstream, or upstream industries makes it easier for domestic firms to export new products or upgrade the complexity of those products. This applies both when each spillover proxy enters the equation on its own (Columns 1–6) and when they are jointly included (Columns 7–8).

In the first stage, the two exclusion restrictions are both statistically significant at the 1 per cent level and their signs are as predicted. $DestinationGDP_{it}$ is positively correlated with the probability that firm i introduces a new product at time t, indicating that South African manufacturing firms, on average, respond positively to demand shocks in their current export markets. In contrast, $Distance_{it}$ is negatively correlated with the probability of new export introductions. The expected negative sign indicates that the distance to a firm's current markets increases the fixed cost of new product introductions.

[32] $Size_{it-1}$ has, contrary to our expectation, a negative coefficient in the first model. This somewhat puzzling finding may be explained by the fact that the largest firms export a wide range of products—both complex and unsophisticated. On average, then, small firms may specialize in a smaller set of core export goods with higher complexity. This hypothesis is somewhat confirmed by the results in Columns 5–8, Table 4, indicating that there is indeed a top-line complexity premium for large firms. In other words, the evidence suggests that large firms do indeed possess the capabilities necessary to export the most complex products, but at the same time the average complexity of their export basket is lower than that of specialized smaller firms.

Table 4: Main results, EC_{it}^{new}—Heckman selection model

	2nd step EC_{it}^{new}	**1st step** $Introduction_{it}$	**2nd step** EC_{it}^{new}	**1st step** $Introduction_{it}$	**2nd step** EC_{it}^{new}	**1st step** $Introduction_{it}$	**2nd step** EC_{it}^{new}	**1st step** $Introduction_{it}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Horizontal_{jpt-1}$	−0.078	−0.149					−0.078	−0.148
	(0.106)	(0.150)					(0.106)	(0.150)
$Backward_{jpt-1}$			0.683	2.589			0.965	0.601
			(3.191)	(4.340)			(3.322)	(4.540)
$Forward_{jpt-1}$					0.386	4.492	0.418	4.724
					(2.253)	(3.091)	(2.316)	(3.191)
$Size_{it-1}$	−0.003	0.042**	−0.003	0.040**	−0.003	0.040**	−0.003	0.042**
	(0.013)	(0.018)	(0.013)	(0.018)	(0.013)	(0.018)	(0.013)	(0.018)
$LabourProductivity_{it-1}$	−0.006	0.022	−0.006	0.025	−0.006	0.024	−0.006	0.022
	(0.013)	(0.017)	(0.013)	(0.016)	(0.013)	(0.016)	(0.013)	(0.017)
$R\&DIntensity_{it-1}$	0.007	−0.001	0.007	−0.001	0.007	−0.000	0.007	−0.001
	(0.007)	(0.011)	(0.007)	(0.011)	(0.007)	(0.011)	(0.007)	(0.011)
$Wage_{it-1}$	0.012	0.031*	0.012	0.030*	0.012	0.030*	0.012	0.030*
	(0.011)	(0.016)	(0.011)	(0.016)	(0.011)	(0.016)	(0.011)	(0.016)
$CountryDiversification_{it-1}$	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0.002)	(0.005)	(0.002)	(0.004)	(0.002)	(0.004)	(0.002)	(0.005)
$ProductDiversification_{it-1}$	0.002	0.071***	0.002	0.072***	0.002	0.072***	0.002	0.071***
	(0.001)	(0.005)	(0.001)	(0.005)	(0.001)	(0.005)	(0.001)	(0.005)
EC_{it-1}^{all}	0.255***	0.097***	0.250***	0.096***	0.250***	0.097***	0.255***	0.098***
	(0.018)	(0.020)	(0.018)	(0.020)	(0.018)	(0.020)	(0.018)	(0.020)
IC_{it-1}^{all}	0.100***	−0.010	0.103***	−0.011	0.103***	−0.010	0.100***	−0.010
	(0.026)	(0.032)	(0.026)	(0.032)	(0.026)	(0.032)	(0.026)	(0.032)
$Distance_{it}$		−0.000***		−0.000***		−0.000***		−0.000***
		(0.000)		(0.000)		(0.000)		(0.000)
$DestinationGDP_{it}$		0.066***		0.064***		0.065***		0.067***
		(0.025)		(0.025)		(0.025)		(0.025)
$\hat{\lambda}_{it}$	−0.195*		−0.217*		−0.209*		−0.198*	
	(0.114)		(0.114)		(0.115)		(0.115)	
Fixed effects								
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,164	5,476	4,197	5,521	4,197	5,521	4,164	5,476

Notes: all variables except spillover proxies, $CountryDiversification_{it-1}$ and $ProductDiversification_{it-1}$ are reported in logs. Standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Source: author's calculations based on SARS data described in Pieterse et al. (2018).

In the second stage, the inverse Mills ratio, $\hat{\lambda}_{it}$, is statistically significant at the 10 per cent level, implying self-selection into export introductions. The sign of lambda's coefficient signals a negative correlation between the error terms in the first- and second-stage regressions, suggesting that firms more likely to introduce new products are also more likely to introduce less sophisticated products.

At first this might seem counterintuitive, but it is in line with the idea that firms able to export very complex products are likely to be highly specialized. Such firms should be less likely to introduce new products.[33]

4.4 Robustness checks

Since the results in FDI spillover studies are sensitive to variations in the econometric specification applied (Görg and Strobl 2001), we test the robustness of our main finding that forward spillovers from FDI positively impact the top-line complexity of domestic firms' exports.

We start by investigating whether the finding is sensitive to three sets of technical robustness checks. The results, along with a detailed description of each check, are provided in Appendix Table A8. First, we change the dependent variable by estimating $EC_{it}^{topline}$ at the HS4 level and by employing Hidalgo and Haussmann's (2009) complexity algorithm to derive product complexity scores.[34] Second, we implement different employment-based FDI proxies instead of the output-based proxies employed in our preferred specification. Finally, we change two temporal aspects in the data. In one test, we extend the timespan of the panel to incorporate the tax year 2012 by extrapolating the foreign presence in each industry to this year based on foreign firm identification in 2013–16. In another test, we use a time-period indicator, which ensures a tighter alignment between firms' financial (CIT) and employment (IRP5) data. In all cases, the coefficient on $Forward_{jpt-1}$ remains positive, stable, and statistically significant at the 1 per cent level.

In Table A9 in the Appendix, we further investigate whether alternative explanations call the validity of our finding into question. In Column 1, we test whether firm liquidity (measured by firms' current ratio—that is, total current assets over total current liabilities) constitutes a source of omitted-variable bias. In Column 2, we re-run our model on a balanced panel to ensure that the results are not driven by FDI-induced export market exits of the least complex South African firms, in which case they could be attributed to spillover effects. In both specifications, our initial result remains significant at the 1 per cent level.

4.5 Heterogeneous effects

The absorptive capacity of domestic firms—that is, the 'ability to recognize the value of new information, assimilate it, and apply it to commercial ends' (Cohen and Levinthal 1990: 128)—is

---

[33] The results from the Heckman model, indicating no spillover effects, should be interpreted with caution. First, the validity of the results is dependent on the exogeneity of the exclusion restrictions. The results reported in Table 4 are sensitive to the exact specification of the model. In all variations, however, neither of three spillover proxies comes out as significant. As a further robustness check, we also conducted a one-step analysis using the OLS model with fixed effects from Equation 10, ignoring self-selection issues. This test supports the findings from the Heckman model, indicating no significant spillover effects. For brevity the results are not presented here, but they are made available in Table A7 in the Appendix.

[34] A detailed description of the complexity algorithm, and how we use it to calculate product complexity scores, is provided in Appendix Table A9.

often found to mediate the impacts of FDI. Table A10 in the Appendix presents the results when we test this proposition on the South African data following the approach in Javorcik et al. (2018). In Column 1, we use the complexity of a firm's export basket to proxy for its absorptive capacity by generating a dummy variable, $D_{it-1}^{complexity}$, taking the value 1 if EC_{it-1}^{all} lies above the sample mean, and interacting it with $Forward_{jpt-1}$. In Column 2, we proxy for absorptive capacity using firm size. We set the dummy variable D_{it-1}^{size} as equal to 1 if a firm employs more workers than the sample average and interact it with the forward-spillover variable. Surprisingly, we find no evidence that the effect of forward FDI spillovers are mediated by either of these firm characteristics.

5 Discussion and conclusion

The complexity of a country's economy, and the associated diversification and sophistication of its exports, has been shown to be a key determinant of economic growth. However, relatively little is known about how EC evolves at the micro level. This study exploits a unique administrative data set covering the entire tax-paying population of firms in South Africa from 2013 to 2016 in order to explore if, and how, the presence of foreign firms affects export upgrading in manufacturing firms. The case of South Africa is of particular importance, since the country is caught in an export impasse, with low growth, high unemployment, widespread poverty, and abject inequality. The main findings can be summarized in three steps.

First, we find significant, positive, and robust evidence that FDI boosts the top-line export complexity in South African manufacturing firms. The detection of upgrading at the top line constitutes a novel contribution to the literature. It implies that it might be necessary to study spillover effects at a granular level to detect learning effects from FDI. Given this discovery, the study contributes to the evidence that foreign firms can act as a catalyst in the process of structural transformation (Javorcik et al. 2018; Lo Turco and Maggioni 2019) and boost the export capabilities of local firms (Aitken et al. 1997; Bajgar and Javorcik 2019; Eck and Huber 2016; Greenaway et al. 2004). There are, however, reservations to this conclusion. Previous studies concerned with the FDI–complexity link have found an effect on the *average* complexity of new product innovations for local firms (Javorcik et al. 2018) and on the *average* level of sophistication of the entire export basket of firms (Eck and Huber 2016). In comparison with these studies, the effect we detect is arguably modest—the boost to top-line export complexity is not strong enough to translate into an average effect. One reason for this might be that South African exporters already know how to produce a lot of things (that is why they are exporters), implying that the knowledge gap between domestic and foreign firms, and therefore the learning potential, is situated in very sophisticated production tasks. It may also take time for the top-line learning effect to manifest itself, because it takes time for businesses to rationalize their product portfolio. Perhaps three years constitutes too short a timeframe over which to detect such a rationalization process.

Second, we find no evidence of horizontal spillovers, thus confirming Hypothesis $\mathbf{H}_1$. This finding adds another piece of evidence to the emerging consensus that, on average, horizontal spillovers from FDI are unlikely to profoundly change the export capacity of local firms (Bajgar and Javorcik 2019; Eck and Huber 2016; Javorcik et al. 2018). On one hand, the result confirms the theoretical prediction that spillovers are less likely to occur horizontally because foreign firms have an incentive to prevent their firm-specific knowledge from leaking to domestic competitors. On the other hand, one should be careful about drawing the conclusion that intra-industry spillovers do not occur at all. As pointed out by Moran (2007), it might be that the positive capability-enhancing (demonstration effect and labour mobility) and incentivizing (cost discovery and competition) externalities from FDI are cancelled out by negative effects (brain drain and crowding out).

Third, Hypothesis $\mathbf{H_2}$ is only partly confirmed given the mixed results on vertical spillovers. On one hand, the analysis has shown that FDI-engendered export upgrading in South Africa is driven by foreign firms located in supplying (upstream) sectors. The result is economically meaningful: a one-percentage-point increase in the ratio of foreign firms located in an upstream sector translates into a 5.4 per cent boost to the most complex product exported by domestic firms. It confirms a general conclusion drawn by studies outside the FDI literature that access to better inputs matters for firm performance (Goldberg et al. 2013; Newman et al. 2016). This is interesting because most empirical evidence on FDI spillovers, especially within the FDI-TFP literature, points towards a quantitatively limited effect of forward spillovers, which has focused researchers' and policymakers' attention on the importance of backward linkages. Yet the results presented here indicate that forward spillovers may have an important role to play when the outcome measured is export complexity. If this is true in general, then forward spillover deserves more attention as an important channel through which FDI can contribute to economic development.[35]

On the other hand, we find no statistically significant relationship between export upgrading and FDI in downstream sectors. This is somewhat surprising, since backward spillovers have been reported by all reviewed studies employing a method similar to the one implemented in this paper (Bajgar and Javorcik 2019; Eck and Huber 2016; Javorcik et al. 2018). This contradictory finding may be interpreted as indicative of a lack of local sourcing by foreign companies in the South African context, limiting the potential for capability-enhancing (technology and knowledge transfer) and incentivizing (demand for new and complex inputs) externalities to benefit local suppliers upstream. Our result may also differ from previous studies due to the data used. For instance, the studies by Bajgar and Javorcik (2019) and Javorcik et al. (2018) use census data covering only firms with more than 20 employers. If large domestic firms are deemed to be more capable suppliers by MNEs, the estimates of positive backward spillovers reported in these studies might be upward-biased due to selection issues.

The results presented here matter for policy. EC is pivotal in South Africa's efforts towards better economic conditions and a brighter future. At a general level, this paper has shown that FDI constitutes a viable policy tool in attempts to boost EC. More specifically, the results imply that policymakers should take a dual approach to FDI. On one hand, FDI promotion targeted at attracting foreign investors into *input-supplying* sectors constitutes an effective industrial policy tool that can spur structural transformation. On the other hand, foreign firms located in *input-sourcing* sectors do not currently stimulate domestic export upgrading. In these sectors, therefore, it is advisable that policymakers shift the focus from attracting FDI to incentivizing the foreign firms already present to source locally and transfer knowledge and technology to domestic suppliers. Such policy efforts may cultivate spillovers from input-sourcing industries.

[35] Our approach of using Javorcik's (2004) econometric strategy to operationalize the conceptual framework presented in Table 1 makes it impossible to directly observe the exact mechanisms driving forward spillovers in South Africa. The econometric method thus falls subject to a criticism Moran has previously levelled at other studies—that it leaves the causal mechanisms by which FDI engenders export spillovers 'totally opaque' (2007: 46). In other words, while the conceptual framework can give us confidence that the correlation detected in the analysis is likely to imply causation, the operationalization of the framework does not allow us to detect which causal mechanisms are at play.

References

Aitken, B.J., G.H. Hanson, and A.E. Harrison (1997). 'Spillovers, Foreign Investment, and Export Behavior'. *Journal of International Economics*, 43(1–2): 103–32.

Albeaik, S., M. Kaltenberg, M. Alsaleh, and C.A. Hidalgo (2017). 'Improving the Economic Complexity Index'. *arXiv*, preprint: 1707.05826[q-fin.EC].

Aterido, R., A. Hlatshwayo, D. Pieterse, and A. Steenkamp (2019). 'Firm Dynamics, Job Outcomes, and Productivity: South African Formal Businesses, 2010–14'. Policy Research Working Paper 8788. Washington, DC: World Bank.

Atlas of Economic Complexity (2019). 'What did South Africa Export in 1995?'. Cambridge, MA: Growth Lab, Center for International Development, John F. Kennedy School of Government, Harvard University. Available at: http://atlas.cid.harvard.edu/explore?country=246&product=undefined&year=1995&productClass=HS&target=Product&partner=undefined&startYear=undefined (accessed September 2019)

Bajgar, M., and B. Javorcik (2019). 'Climbing the Rungs of the Quality Ladder: FDI and Domestic Exporters in Romania'. EBRD Working Paper 235. London: European Bank for Reconstruction and Development (EBRD).

Balsvik, R. (2011). 'Is Labor Mobility a Channel for Spillovers from Multinationals? Evidence from Norwegian Manufacturing'. *The Review of Economics and Statistics*, 93(1): 285–97.

Barrios, S., H. Görg, and E. Strobl (2003). 'Explaining Firms' Export Behaviour: R&D, Spillovers and the Destination Market'. *Oxford Bulletin of Economics and Statistics*, 65(4): 475–96.

Benli, M. (2016). 'FDI and Export Spillovers using Heckman's Two Step Approach: Evidence from Turkish Manufacturing Data'. *Theoretical and Applied Economics*, 28(4): 315–42.

Brambilla, I. (2009). 'Multinationals, Technology, and the Introduction of Varieties of Goods'. *Journal of International Economics*, 79(1): 89–101.

Caves, R.E. (1974). 'Multinational Firms, Competition, and Productivity in Host-Country Markets'. *Economica*, 41(162): 176–93.

Clerides, S.K., S. Lach, and J.R. Tybout (1998). 'Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco'. *The Quarterly Journal of Economics*, 113(3): 903–47.

Cohen, W.M., and D.A. Levinthal (1990). 'Absorptive Capacity: A New Perspective on Learning and Innovation'. *Administrative Science Quarterly*, 35(1): 128–52.

Cristelli, M., A. Gabrielli, A. Tacchella, G. Caldarelli, and L. Pietronero (2013). 'Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products'. *PLoS ONE*, 8(8): e70726.

Eapen, A. (2013). 'FDI Spillover Effects in Incomplete Datasets'. *Journal of International Business Studies*, 44(7): 719–44.

Eck, K., and S. Huber (2016). 'Product Sophistication and Spillovers from Foreign Direct Investment'. *Canadian Journal of Economics/Revue canadienne d'économique*, 49(4): 1658–84.

Gereffi, G., J. Humphrey, and T. Sturgeon (2005). 'The Governance of Global Value Chains'. *Review of International Political Economy*, 12(1): 78–104.

Goldberg, P., A. Khandelwal, and N. Pavcnik (2013). 'Variety In, Variety Out: Imported Input and Product Scope Expansion in India'. In J. Bhagwati and A. Panagariya (eds), *Reforms and Economic Transformation in India*. Oxford: Oxford University Press.

Görg, H., and E. Strobl (2001). 'Multinational Companies and Productivity Spillovers: A Meta-Analysis'. *The Economic Journal*, 111(475): 723–39.

Görg, H., and E. Strobl (2005). 'Spillovers from Foreign Firms through Worker Mobility: An Empirical Investigation'. *The Scandinavian Journal of Economics*, 107(4): 693–709.

Greenaway, D., N. Sousa, and K. Wakelin (2004). 'Do Domestic Firms Learn to Export from Multinationals?' *Europäische Zeitschrift für politische Ökonomie*, 20(4): 1027–43.

Gualier, G., and S. Zignago (2010). 'BACI: International Trade Database at the Product Level'. Working Paper 2010-23. Paris: Centre d'Études Prospectives et d'Informations Internationales (CEPII).

Harding, T., and B.S. Javorcik (2012). 'Foreign Direct Investment and Export Upgrading'. *Review of Economics and Statistics*, 94(4): 964–80.

Hausmann, R., and B. Klinger (2006a). 'South Africa's Export Predicament'. CID Working Paper 129. Cambridge, MA: Center for International Development (CID) at Harvard University.

Hausmann, R., and B. Klinger (2006b). 'Structural Transformation and Patterns of Comparative Advantage in the Product Space'. CID Working Paper 128. Cambridge, MA: CID.

Hausmann, R., and D. Rodrik (2003). 'Economic Development as Self-Discovery'. *Journal of Development Economics,* 72(2): 603–33.

Hausmann, R., C.A. Hidalgo, S. Bustos, M. Coscia, A. Simoes, and M.A. Yildrim (eds) (2013). *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, MA: The MIT Press.

Hausmann, R., J. Hwang, and D. Rodrik (2007). 'What You Export Matters'. *Journal of Economic Growth*, 12(1): 1–25.

Havranek, T., and Z. Irsova (2011). 'Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is'. *Journal of International Economics*, 85(2): 234–44.

Hidalgo, C.A. (2011). 'Discovering East Africa's Industrial Opportunities'. Economic Policy Paper Series. Washington, DC: The German Marshall Fund of the United States.

Hidalgo, C.A., and R. Hausmann (2009). 'The Building Blocks of Economic Complexity'. *Proceedings of the National Academy of Sciences*, 106(26): 10570–75.

Javorcik, B.S. (2004). 'Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages'. *The American Economic Review*, 94(3): 605–27.

Javorcik, B.S. (2008). 'Can Survey Evidence Shed Light on Spillovers from Foreign Direct Investment?' *The World Bank Research Observer*, 23(2): 139–59.

Javorcik, B.S., A. Lo Turco, and D. Maggioni (2018). 'New and Improved: Does FDI Boost Production Complexity in Host Countries?' *The Economic Journal*, 128(614): 2507–37.

Kneller, R., and M. Pisu (2007). 'Industrial Linkages and Export Spillovers from FDI'. *The World Economy*, 30(1): 105–34.

Kokko, A., M. Zejan, and R. Tansini (2001). 'Trade Regimes and Spillover Effects of FDI: Evidence from Uruguay'. *Weltwirtschaftliches Archiv*, 137(1): 124–49.

Kreuser, F., and C. Newman (2018). 'Total Factor Productivity in South African Manufacturing Firms'. *South African Journal of Economics*, 86: 40–78.

Lo Turco, A., and D. Maggioni (2019). 'Local Discoveries and Technological Relatedness: The Role of MNEs, Imports and Domestic Capabilities'. *Journal of Economic Geography*, 19(5): 1077–98.

Mariani, M.S., A. Vidmer, M. Medo, and Y.C. Zhang (2015). 'Measuring Economic Complexity of Countries and Products: Which Metric to Use?' *The European Physical Journal B*, 88(11): 293.

Markusen, J.R., and A.J. Venables (1999). 'Foreign Direct Investment as a Catalyst for Industrial Development'. *European Economic Review*, 43: 335–56.

Mayer, T., and S. Zignago (2011). 'Notes on CEPII's distances measures: The GeoDist database'. CEPII Working Paper 2011/25. Paris: CEPII.

Mayneris, F., and S. Poncet (2015). 'Chinese Firms' Entry to Export Markets: The Role of Foreign Export Spillovers'. *The World Bank Economic Review*, 29(1): 150–79.

Melitz, M.J. (2003). 'The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity'. *Econometrica*, 71(6): 1695–725.

Moran, T. (2001). *Parental Supervision: The New Paradigm for Foreign Direct Investment and Development*. Washington, DC: Peterson Institute for International Economics.

Moran, T. (2007). 'How to Investigate the Impact of Foreign Direct Investment on Development, and Use the Results to Guide Policy '. *Brookings Trade Forum*, 2007: 1–60.

Moran, T. (2010). *Foreign Direct Investment and Development: Launching a Second Generation of Policy Research*. Washington, DC: Peterson Institute for International Economics.

Moran, T. (2018). 'FDI and Supply Chains in Horticulture (Vegetables, Fruits, and Flowers, Raw, Packaged, Cut, and Processed): Diversifying Exports and Reducing Poverty in Africa, Latin America, and Other Developing Economies'. CGD Working Paper 475. Washington, DC: Center for Global Development (CGD).

Newman, C., J. Rand, T. Talbot, and F. Tarp (2015). 'Technology Transfers, Foreign Investment and Productivity Spillovers'. *European Economic Review*, 76: 168–87.

Newman, C., J. Rand, and F. Tarp (2016). 'Imports, Supply Chains, and Firm Productivity'. WIDER Working Paper 2016/90. Helsinki: UNU-WIDER.

OECD (2008). *OECD Benchmark Definition of Foreign Direct Investment*. Paris: Organisation for Economic Co-operation and Development (OECD).

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

Poole, J.P. (2013). 'Knowledge Transfers from Multinational to Domestic Firms. Evidence from Worker Mobility'. *The Review of Economics and Statistics*, 95(2): 393–406.

Quantec EasyData (2019). 'SA Standardised Industry Input and Output Structure at Basic Prices'. Pretoria: Quantec. Available at https://www.easydata.co.za (accessed February 2019).

Rodríguez-Clare, A. (1996). 'Multinationals, Linkages, and Economic Development'. *American Economic Review*, 86(4): 852–73.

Rodrik, D. (2008). 'Understanding South Africa's Economic Puzzles'. *Economics of Transition*, 16(4): 769–97.

Romer, P.M. (1992). 'Two Strategiesfor Economic Development: Using Ideas and Producing Ideas'. *The World Bank Economic Review*, 6(suppl_1): 63–91.

SARS (2013). *Comprehensive Guide to the ITR14 Return for Companies*. Pretoria: South African Revenue Service (SARS).

SARS (2016). 'HS 12'. Pretoria: SARS. Available at: https://www.sars.gov.za/Legal/Primary-Legislation/Pages/HS-2012.aspx (accessed May 2019).

SARS (2019). 'Dividends Tax'. Pretoria: SARS. Available at: https://www.sars.gov.za/TaxTypes/DT/Pages/default.aspx?fbclid=IwAR0jyeH9DyiiKMNzyE_JtJC2A_cIvjY-m0JSJtah3mwQKFhi3nQbbk8cF5o (accessed September 2019).

Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero (2012). 'A New Metrics for Countries' Fitness and Products' Complexity'. *Scientific Reports*, 2: article 723.

Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero (2013). 'Economic Complexity: Conceptual Grounding of a New Metrics for Global Competitiveness'. *Journal of Economic Dynamics and Control*, 37: 1683–91.

UN (2019). 'HS2007–HS2012 Correspondence Tables'. UN Trade Statistics. Available at: https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp (accessed May 2019).

UNCTAD. (2005). *World Investment Report 2005: Transnational Corporations and the Internationalization of R&D*. New York and Geneva: UN Conference on Trade and Development (UNCTAD).

Wier, L., and H. Reynolds (2018). 'Big and "Unprofitable": How 10% of Multinational Firms Do 98% of Profit Shifting'. WIDER Working Paper 2018/111. Helsinki: UNU-WIDER.

World Bank (2018). 'World Development Indicators'. Washington, DC: World Bank. Available at: https://datacatalog.worldbank.org/dataset/world-development-indicators (accessed January 2019).

Appendix (the appendix has been formatted but not checked)

Appendix A1

Table A1: Literature on Export Upgrading Spillovers from FDI

Authors	**Type**	**Country & period**	**Sample**	**Exclusive export focus**	**Dependent variable measure**	**Level of dependent variable**	**FDI spillover proxy**	**Methodology**	**Type of spillover: result** (+, - , none)
Mayneris and Poncet (2015)	Diversification	China (1997-2007)	Customs data: all provincial export flows	Yes	Province-product-destination export introduction (dummy)	Province-product-destination	Foreign firm export presence in province-product-destination triad	Fixed effects IV	Province-product-destination: +
La Turco & Maggioni (2018)	Diversification	Turkey (2006-2009)	SBS: all firms with 20+ employees	No	Regional product introduction (dummy)	Firm	Foreign firm proximity to not-yet-produced products within region	Fixed effects	Regional product-space: +
Bajgar and Javorcik (2019)	Quality	Romania (2005-2011)	SBS: all firms with 20+ employees plus subsample of smaller firms	Yes	Quality: (unit values and Khandelwal (2013))	Firm-product-destination	Following Javorcik (2004)	Fixed effects Differencing	Horizontal: none Backward: + Forward: +
Eck and Huber (2016)	Complexity	India (2001-2010)	Prowess database: 80% of manufactured output	Yes	Avg. complexity of export basket (PRODY)	Firm	Following Javorcik (2004)	Fixed effects	Horizontal: none Backward: + Forward: -
Javorcik et al. (2018)	Complexity	Turkey (2006-2009)	SBS: all firms with 20+ employees	No	Complexity of new product introductions (complexity-algorithm)	Firm	Following Javorcik (2004)	Fixed effects Heckman IV	Horizontal: none Backwards: + Forward: none
This paper	Complexity	South Africa (2013-2016)	Tax administrative data: all tax paying entities	Yes	Top-line complexity (complexity-algorithm and fitness-algorithm)	Firm	Following Javorcik (2004)	Fixed effects Heckman	Horizontal: none Backwards: none Forward: +

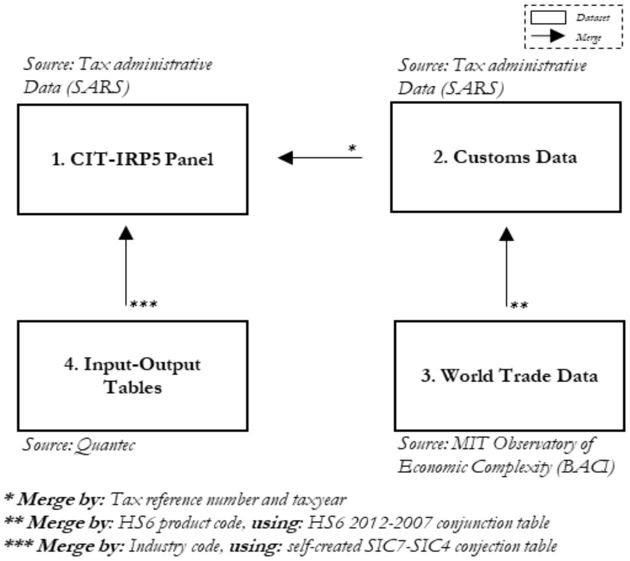
Notes: SBS refers to Structural Business Survey.

Source: author's illustration.

Appendix A2

The final dataset used is built from four different data sources as illustrated in Figure A1. Here, we outline in detail how the data sources are cleaned and merged.

Figure A1: Illustration of merging procedure of datasets



Source: author's illustration.

The CIT-IRP5 Panel is cleaned as follows: In order to obtain the final sample of manufacturing firms for the analysis, we drop all firms with inadequate information on key variables, as shown in Table A2 above. We are left with 77,505 manufacturing firms of which 11,519 are domestic exporters (appr. 15 per cent), whereas 19,271 are foreign firms (appr. 25 per cent). In the regression analysis, we lag all explanatory variables by one period. Doing so leaves us with a final sample of 5,485 South African exporting manufacturers. It is important to notice that a substantial amount of firms drop out of the sample because they lack employment figures. It is currently unclear what causes this attrition (see Table A2.)

Table A2: Sample size through cleaning procedure

Tax Year	All firms	Manu-facturing firms	With geo-data	Non-dormant firms	With employees	Domestic Exporters	With lagged variables	Foreign manu-facturing firms
2013	505,392	32,551	32,246	27,327	19,248	2,812	-	4,972
2014	547,125	33,963	33,822	28,763	19,971	2,795	1,778	5,058
2015	542,709	32,503	32,366	27,687	19,513	2,915	1,840	4,761
2016	536,442	31,165	31,078	26,685	18,773	2,997	1,867	4,480
Total	**2,131,668**	**130,182**	**129,512**	**110,462**	**77,505**	**11,519**	**5,485**	**19,271**

Source: author's calculations based on SARS data.

The SARS customs data is cleaned as follows: The data contains transaction-level information on all imports and exports in South Africa. SARS has provided a code to filter out all transactions that should not be regarded as actual trade (such as movement of goods between warehouses, etc.). Yet, some firms in the data still appear to export more than 700 different products per year at the HS6-level (appr. 5,000 product categories in total). It seems highly unlikely that these are actual manufacturing firms. As a consequence, we winsorize the data, dropping tax reference numbers belonging to the 5 per cent of firms exporting the highest amount of different goods per year. Now, the most diversified firms export in the order of 300 different goods per year, a more believable figure. Additionally, we drop all transactions that cannot be linked to a tax reference number, unclassified product codes, and transactions where the country of destination (for exports) or origin (for imports) is unknown. After discussions with other researchers, we also drop export transactions flowing to members of the Southern African Customs Union (SACU) (Eswatini (Swaziland), Botswana, Namibia, Lesotho, and Namibia). We are only concerned with exports that are actually being produced in South Africa and anecdotal evidence suggests that many goods imported from SACU member states flow through customs offices in South Africa even though they are produced elsewhere. To further reduce noise, we drop all single transactions worth less than R 10 and all transactions per firm per product per tax year worth less than R 7,500. Then, we aggregate the data from the transaction-level to the product-level allowing us to merge in product complexity scores derived from world trade data.

The BACI world trade data (2010) is cleaned as follows: To reduce noise in the data before these calculations, we adopt and adapt the cleaning approach in Albeaik et al. (2017). First, we apply three time-independent filters to the data. We drop all countries whose total export value is less than USD 1 billion in 2012. Then, we exclude countries with unreliable export data (Chad, Iraq, Afghanistan, and Macau). We also drop countries whose population is smaller than 1,250,000 in 2012. The population data is merged from the World Development Indicators (WDI 2018). Second, we apply a set of time-varying filters to the data. We exclude products whose dollar value is zero for 99 per cent of countries in a given year. We also exclude countries in years where their export value equals zero for 95 per cent of all products. Additionally, we exclude products whose yearly world-total export value is less than USD 10 million. Finally, we round all country-product–year export values to zero if they are below USD 5,000. This leaves us with 134 countries and just shy of 5,000 product categories (we conduct a similar procedure at the HS4-level resulting in appr. 1,200 product categories) from which we calculate product complexity scores for each product using the fitness and complexity algorithms. The product codes from the BACI database follow the 2007-revision of the HS6 classification, but the South African customs data follows the 2012-revision[36]. Therefore, we merge the datasets using a HS6 2007-2012 correspondence table obtained from UN Trade Statistics (UN 2019). Finally, we aggregate the data from the product-level to the firm-level, calculate export/import complexity scores for exporting manufacturing firms, and merge the variables with the CIT-IRP5 Panel based on tax reference number and tax year.

The I-O tables are cleaned and merged with our final dataset as described in the Section 3.1

[36] South Africa adopted the 2012-revision in January 2012 (SARS 2011).

Appendix A3

Identifying foreign firms

The benchmark definition by the OECD (2008) defines a foreign direct investment enterprise as a corporation (or quasi-corporation) where at least 10 per cent of its operating assets are owned by a foreign investor. An *associate* is defined as a corporation where foreign investors own 10-50 per cent of its operating assets. In a *subsidiary*, the share of foreign owned assets constitutes more than 50 per cent. *Branches* are also included under the OECD definition of FDI. These are unincorporated direct investment enterprises, not legally separate from their parent companies whom own them fully (OECD 2008).

In the FDI-spillover literature, any firm falling into one of the categories outlined above is classified as foreign. We attempt to follow this approach, but unfortunately there is no information of asset-ownership in the CIT data and it is therefore not straight-forward to identify foreign firms. Different variables in the panel do, however, indicate whether a firm is foreign or not, but it is likely that they do not collectively cover the entire population of foreign firms. Consequently, we take a "shotgun approach" utilizing all variables from the CIT-IRP5 Panel indicating any form of foreign ownership. Some variables do, however, provide some information on the foreign ownership structure. This is important because it can provide an idea about which types of foreign firms we may be missing.

Two previous studies using the tax data (Aterido et al. 2019; Wier and Reynolds 2018) have identified foreign firms solely based on whether the ultimate holding company is foreign. The question, however, only applies to subsidiaries whose share of foreign owned assets exceeds 50 per cent. Hence, the approach underestimates the true population of foreign firms because it fails to identify foreign branches as well as associates where 10-50 per cent of the operating assets are owned by foreign investors. Our "shotgun approach" broadens the scope of the foreign firm-indicator used in these studies.

Table A3: Yearly count of foreign firms by question in ITR14 tax form

	1. Non-resident			2. Subsidiary/Associate		3. Branch
Tax year	Non-resident for tax purposes	Non-resident due to foreign incorporation	Non-resident by virtue of DTA	Foreign holding company (>50%)	Foreign dividends paid exemption (>10%)	Branch/ permanent establishment/ agency
2013	4,211	15	19	564	101	273
2014	4,251	19	10	597	96	294
2015	3,961	12	-	636	101	217
2016	3,677	19	13	640	102	197
Total	**16,100**	**65**	**42**	**2,437**	**400**	**981**

Notes: some counts are removed for sensitivity reasons due to too few observations. Share of foreign owned assets reported in parentheses.

Source: author's calculations based on SARS data.

We categorise firms as foreign if they fall into one of three categories (1.non-resident, 2.subsidiary/associate, or 3.branch) based on their answer to six questions in the ITR14 tax form. Table A3 above displays the yearly counts of firms classified as foreign by each of these six questions. First, we classify firms as foreign if they are not residents in South Africa for tax purposes. Formally, SARS defines a non-resident as a company that 'is not incorporated, established or formed in South Africa and does not have its place of effective management in South Africa. The place of effective management in the case of a company is the place where it is managed on a regular or day-to-day basis' (SARS 2013:53). Furthermore, discussions with a South African tax expert indicate that the category might also include foreign companies conducting once-off contractual work in South Africa. Thus, while it is clear that non-resident companies are foreign and thereby included in our FDI-indicator, the data does not allow us to be specific about their ownership structure. It is currently unclear why the two follow-up questions in the adjacent columns (non-residency due to foreign incorporation or by virtue of double taxation agreement (DTA)) are poorly populated. It is concerning if this is because firms misunderstand and therefore misreport the initial question on tax-residency. If this is the case, the FDI-indicator developed here will exhibit considerable measurement error, and the results reported should therefore be interpreted with this caveat in mind.

Some questions in the ITR14 tax form do allow us to identify foreign subsidiaries/associates and branches of foreign companies. Counts for these categories are shown in the last three columns of Table A3. The column labelled "Foreign holding company" counts the manufacturing subsidiaries classified as foreign in the studies by Wier and Reynolds (2018) and Aterido et al. (2019). In order to identify associates, we also incorporate information on whether firms declare dividends subject to double-taxation relief. In most of South Africa's DTAs, a company's declared dividends to a non-resident are subject to double-taxation relief if the non-resident owns a minimum percentage of shares (typically between 10 and 25 per cent) in the resident company (SARS 2019). While this variable thereby covers both subsidiaries and associates, it has two drawbacks. On one hand, a few of South Africa's DTAs do not enforce a minimum threshold of share-ownership for dividends relief. The variable may therefore capture some portfolio investments by foreign companies that do have a 10 per cent ownership stake in the local company. On the other hand, the variable is unlikely to capture all associates, as not every associate can be expected to declare dividends subject to double-taxation relief. To the best of our knowledge, however, the variable constitutes the only way to partly identify associates, provided that they are not "caught" by the non-resident category.

Finally, companies also indicate whether their ITR14 tax return is in respect of a branch/permanent establishment/agency of a foreign company. This variable constitutes the final bullet in our "shotgun approach".

Having described how we classify foreign firms, it is necessary to comment on the classification of domestic firms applied in the regression analysis. Here, we follow Javorcik et al. (2018) and classify firms as domestic only if they are not foreign in the *current* period, and if they were not classified as foreign in the *previous* period. This trick enables us to clearly distinguish any effect on export complexity of being owned, in full or partially, by a foreign firm from the actual spillover effect occurring from operating independently in the same, upstream, or downstream sectors as multinationals.

Potential consequences of measurement error

A failure to identify a foreign firm in the data has a double whammy-effect: it is not just that the firm will not be classified as foreign, it will automatically be assumed to be domestic. This is problematic because it implies that the identification of spillover effects may suffer from both measurement error and selection problems (Eapen 2013).

First, failure to catch all foreign firms will cause measurement error in the key regressors (*Horizontal*$_{jpt-1}$, *Backward*$_{jpt-1}$, and *Forward*$_{jpt-1}$). If the measurement error is random across province and industries, the spillover coefficients will exhibit classic attenuation bias. But in our case, it is not unlikely that the measurement error is systematically more severe in some industries, for instance those where firms with lower levels of foreign ownership are more prevalent (recall that it is especially firms with a foreign investor share below 50 per cent that we may not be able to identify). Through a series of Monte Carlo simulations, Eapen (2013) shows that an underestimation of foreign presence typically leads to an upwards bias in the spillover coefficients. Intuitively, one can think of the upward bias as a consequence of the fact that changes in domestic firm characteristics (here: export complexity) will be attributed to smaller increases in the presence of foreign firms than they should be, even though the reality is more complicated than this. Thus, the results reported in this paper may be upward biased if our foreign firm-indicator underestimates the true presence of foreign firms. Since we do not know the extent of the measurement error, it is not possible to determine the magnitude of the bias.

Second, the issue of foreign firm-identification can cause a selection-like problem (Eapen 2013), where the effect of FDI is potentially measured on a group of firms that is different from the target population. To be precise, we are interested in the impact of FDI on export complexity among all local, exporting manufacturing firms, but due to the double-whammy effect described above, what we are likely to observe in the regression is the impact of FDI on:

all local, exporting manufacturing firms

plus *any foreign, exporting manufacturing firms falsely identified to be local*

This is problematic if the effect of FDI is different between our target population and the population observed. Again, it is not straightforward to anticipate the potential bias from the selection problem. "Falsely-domestic firms" may either be better at absorbing knowledge from other MNEs, but they might also possess a lot of knowledge already, which would exclude potential learning to take place. Thus, the direction and magnitude of the resulting bias is unclear, and the results reported in this paper should be interpreted keeping this uncertainty in mind.

Appendix A4

Table A4: Distribution of domestic exporters and foreign firms across manufacturing sectors

	Domestic Exporters		Foreign Firms			Export Split	
Manufacturing sectors	Count	%	Count	%	Exporting %	Domestic %	Foreign %
Meat, fish, fruit etc.	137	1.90	338	1.76	16.0	71.7	28.3
Dairy products	40	0.56	76	0.40	17.1	75.5	24.5
Grain mill products	76	1.06	222	1.15	22.5	60.3	39.7
Other food products	153	2.13	416	2.16	27.2	57.5	42.5
Beverages	211	2.93	212	1.10	33.0	75.1	24.9
Textiles	108	1.50	273	1.42	26.0	60.3	39.7
Knitted, crocheted articles	15	0.21	92	0.48	16.3	50.0	50.0
Wearing apparel	328	4.56	889	4.62	21.5	63.2	36.8
Leather and leather and fur products	85	1.18	107	0.56	39.3	66.9	33.1
Footwear	41	0.57	75	0.39	14.7	78.8	21.2
Sawmilling and planing of wood	10	0.14	72	0.37	*Censored*	*Censored*	*Censored*
Products of wood	149	2.07	972	5.06	10.7	58.9	41.1
Paper and paper products	177	2.46	468	2.43	16.9	69.1	30.9
Printing, recorded media	177	2.46	1,232	6.41	8.8	61.9	38.1
Coke, petroleum products and nuclear fuel	27	0.38	106	0.55	23.6	51.9	48.1
Basic chemicals	119	1.65	338	1.76	47.9	42.3	57.7
Other chemical products	503	6.99	849	4.42	36.9	61.6	38.4
Rubber products	87	1.21	377	1.96	24.7	48.3	51.7
Plastic products	436	6.06	1,034	5.38	24.3	63.5	36.5
Glass and glass products	14	0.19	60	0.31	*Censored*	*Censored*	*Censored*
Non-metallic mineral products	118	1.64	295	1.53	24.1	62.4	37.6
Basic iron and steel products	185	2.57	609	3.17	16.4	64.9	35.1
Non-ferrous metal products	114	1.58	344	1.79	23.0	59.1	40.9
Structural metal products	219	3.04	817	4.25	12.6	68.0	32.0
Other fabricated metal products	561	7.80	1,500	7.80	20.1	65.0	35.0
General purpose machinery	490	6.81	1,050	5.46	28.6	62.0	38.0
Special purpose machinery	276	3.84	506	2.63	40.3	57.5	42.5
Office, accounting, computing machinery	18	0.25	51	0.27	*Censored*	*Censored*	*Censored*
Electric motors, generators, transformers	100	1.39	194	1.01	36.1	58.8	41.2
Insulated wire and cables	43	0.60	75	0.39	34.7	62.3	37.7
Other electrical equipment	240	3.34	550	2.86	25.6	63.0	37.0
Radio, television and communication app.	120	1.67	227	1.18	28.6	64.9	35.1
Professional equipment	24	0.33	62	0.32	45.2	46.2	53.8
Motor vehicles	60	0.83	169	0.88	39.1	47.6	52.4
Parts and accessories	60	0.83	283	1.47	59.0	26.4	73.6
Other transport equipment	128	1.78	305	1.59	46.9	47.2	52.8
Furniture	178	2.47	664	3.45	13.0	67.4	32.6
Other manufacturing groups	1,369	19.02	3,318	17.26	23.7	63.5	36.5
Total	**7,196**	**100**	**19,227**	**100**	**23.5**	**61.4**	**38.6**

Notes: all splits are based on firm counts. Some cells are censored due to sensitivity issues. The count of firms differs slightly from the one displayed in Table A2 since we apply the regression analysis-classification of domestic firms (domestic firms should be classified as domestic both in the current and the previous period). Additionally, the table does not display counts from the Tobacco industry because of too low cell-counts. Finally, a set of outliers in terms of product complexity scores has been removed.

Source: author's calculations based on SARS data.

Appendix A5

Table A5: GDP per capita and country-level $EC_{it}^{topline}$ (2014) - OLS

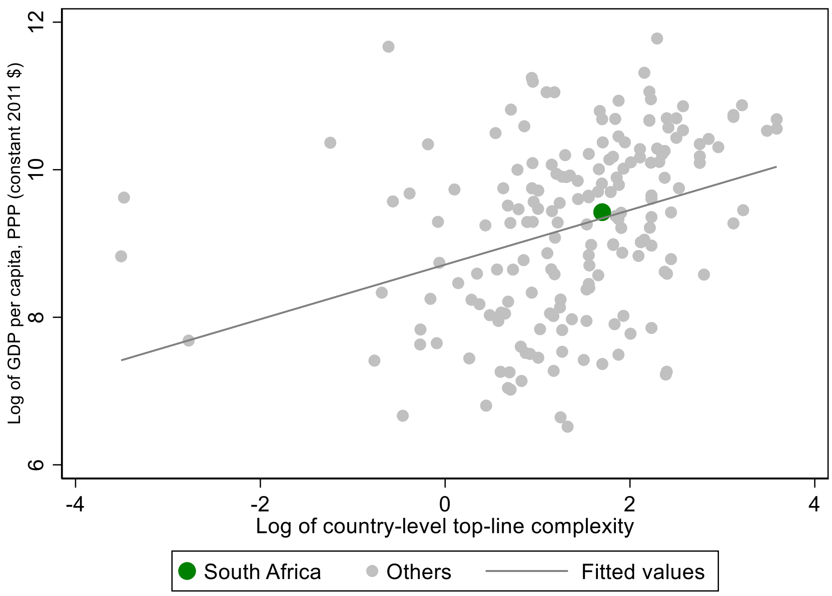
	GDP per capita (log)
	(1)
$EC_{it}^{topline}$	0.370***
	(0.0756)
Constant	8.714***
	(0.133)
Observations	185
R-squared	0.116

Notes: standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Source: author's calculations based on World Bank's WDI (2018) and BACI world trade data (2010).

Appendix A6

Figure A2: Relationship between GDP per capita and country-level $EC_{it}^{topline}$ (2014)



Notes: the country-level $EC_{it}^{topline}$ index is calculated by applying Tacchella et al.'s (2012) fitness-algorithm to world trade data at the HS6 level. Average product complexity over the sample period is used.

Source: author's illustration based on World Bank's WDI (2018) and BACI world trade data (2010).

Appendix A6

Table A6: Leave one out robustness check, $EC_{it}^{topline}$ - OLS

	$EC_{it}^{topline}$										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Horizontal$_{jpt-1}$					-0.067			-0.022	-0.012	-0.008	-0.065
					(0.076)			(0.095)	(0.091)	(0.088)	(0.073)
Backward$_{jpt-1}$						1.561		5.078**	4.743**	-0.115	0.366
						(2.100)		(2.549)	(2.373)	(2.240)	(2.183)
Forward$_{jpt-1}$							5.439***	3.363**	3.759**	4.709***	5.434***
							(1.446)	(1.661)	(1.705)	(1.539)	(1.420)
Size$_{it-1}$	0.126***	0.036**	0.137***	0.051***	0.049***	0.051***	0.050***	0.125***	0.034**	0.137***	0.049***
	(0.013)	(0.014)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.013)	(0.014)	(0.010)	(0.011)
LabourProductivity$_{it-1}$		0.013	0.051***	0.022*	0.021*	0.022*	0.022*		0.012	0.049***	0.020*
		(0.012)	(0.018)	(0.011)	(0.012)	(0.011)	(0.011)		(0.012)	(0.018)	(0.011)
R&DIntensity$_{it-1}$		0.003	0.025***	0.009*	0.008	0.009*	0.009*		0.002	0.024***	0.008
		(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)		(0.006)	(0.005)	(0.005)
Wage$_{it-1}$		0.022**	0.023*	0.022**	0.022**	0.022**	0.022**		0.022**	0.023*	0.022**
		(0.010)	(0.013)	(0.010)	(0.010)	(0.010)	(0.010)		(0.010)	(0.013)	(0.010)
CountryDiversification$_{it-1}$		0.023***		0.019***	0.019***	0.019***	0.019***		0.024***		0.019***
		(0.003)		(0.003)	(0.003)	(0.003)	(0.003)		(0.003)		(0.003)
ProductDiversification$_{it-1}$		0.013***		0.017***	0.017***	0.017***	0.017***		0.013***		0.017***
		(0.002)		(0.002)	(0.002)	(0.002)	(0.002)		(0.002)		(0.002)
EC_{it-1}^{all}			0.403***	0.378***	0.378***	0.378***	0.379***			0.403***	0.379***
			(0.021)	(0.020)	(0.020)	(0.020)	(0.020)			(0.021)	(0.020)
IC_{it-1}^{all}			0.072***	0.076***	0.073***	0.075***	0.076***			0.070***	0.073***
			(0.027)	(0.026)	(0.026)	(0.026)	(0.026)			(0.027)	(0.026)
Fixed effects											
Industry	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Province	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Industry x Year	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Province x Year	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	7,144	7,144	5,487	5,487	5,442	5,487	5,487	7,091	7,091	5,442	5,442
R-squared	0.191	0.282	0.328	0.415	0.417	0.415	0.416	0.195	0.285	0.331	0.418

Notes: robust standard errors, clustered at the province-industry level, in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Source: author's calculations based on SARS data.

Appendix A7

Table A7: Robustness check, EC_{it}^{new} - OLS

	EC_{it}^{new}			
	(1)	(2)	(3)	(4)
$Horizontal_{jpt-1}$	-0.091			-0.090
	(0.122)			(0.123)
$Backward_{jpt-1}$		0.737		0.884
		(2.839)		(2.735)
$Forward_{jpt-1}$			0.789	0.786
			(2.655)	(2.636)
$Size_{it-1}$	0.003	0.003	0.003	0.002
	(0.014)	(0.014)	(0.014)	(0.014)
$LabourProductivity_{it-1}$	-0.002	-0.002	-0.002	-0.002
	(0.012)	(0.012)	(0.012)	(0.012)
$R\&DIntensity_{it-1}$	0.007	0.007	0.007	0.007
	(0.008)	(0.008)	(0.008)	(0.008)
$Wage_{it-1}$	0.015	0.015	0.015	0.015
	(0.011)	(0.011)	(0.011)	(0.011)
$CountryDiversification_{it-1}$	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
$ProductDiversification_{it-1}$	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
EC_{it-1}^{all}	0.266***	0.262***	0.262***	0.265***
	(0.028)	(0.027)	(0.027)	(0.028)
IC_{it-1}^{all}	0.097***	0.100***	0.100***	0.097***
	(0.028)	(0.028)	(0.028)	(0.028)
Fixed effects				
Industry	YES	YES	YES	YES
Province	YES	YES	YES	YES
Year	YES	YES	YES	YES
Industry x Year	YES	YES	YES	YES
Province x Year	YES	YES	YES	YES
Observations	4,168	4,201	4,201	4,168
R-squared	0.195	0.192	0.193	0.195

Notes: robust standard errors, clustered at the province-industry level, in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Source: suthor's calculations based on SARS data.

Appendix A8

In Table A8, we perform three sets of robustness checks by changing *(i)* the dependent variable, *(ii)* the FDI measure, and *(iii)* the time dimension of the specification from Table 3, column 8 (and replicated below in Table A8, column 1).

Changing the Dependent Variable

First, we check whether the result is sensitive to changes in the dependent variable. Column 2 replicates the results from column 1, but here the dependent variable, $EC_{it}^{topline}$, is measured at the HS4-level. In order to derive this dependent variable, we applied the fitness-algorithm to the country-product matrix, M_{cp}, with HS4-level product categories, merged these complexity scores with the customs data, and calculated the most complex HS4-level product exported by each firm. The magnitude of the forward spillover-coefficient is markedly lower than in our preferred specification (3.3 vs. 5.4), but it remains positive and statistically significant at the 1 per cent level. This result is comforting; firstly because the application of a more aggregated product category reduces the volatility of the product-complexity scores derived using the fitness-algorithm. Thus, it does not seem that a high level of volatility in these estimates substantially impacts our initial results. Secondly, the result indicates that the positive FDI-spillover effect is not simply a phenomenon occurring at a very granularly defined margin, but can also be picked up when broader product categories are employed.

In column 3, the dependent variable, $EC_{it}^{topline}$, is calculated using Hidalgo and Hasumann's complexity algorithm to derive product complexity scores. Despite the important differences between the fitness-algorithm and the complexity-algorithm discussed above, the main conclusion of our analysis does not change when applying the latter. Because $K_{p,13}$ is normalized in the PCI-index, it is not necessary to take the log of the dependent variable (see Appendix 9). This means, however, that the effect size should be interpreted differently. The result in column 3 indicates that a 10 percentage point increase in the ratio of foreign firms operating in upstream sectors is associated with an increase of the sophistication of the most complex export goods of 0.39. This corresponds to around 45 per cent of a standard deviation in the sample.

In column 4, we trim the top and bottom 2 per cent of the distribution of $EC_{it}^{topline}$ to check whether the result is robust to the removal of outliers. This check confirms our initial conclusion.

Changing the FDI Measure

In columns 5 and 6, we calculate the FDI-ratio in each sector using foreign firms' employment share instead of their share of output. In column 5, we use employment numbers from the CIT-IRP5 Panel calculated based on all sources of income reported by individuals in the IRP5 data. In column 6, we employ a stricter definition of employment solely based on source code 3601 from the IRP5 data (services rendered, overtime, and pension).[37] In both specifications, $Forward_{jpt-1}$ remains significant at the 1 per cent level. In a meta-analysis of the FDI-TFP literature, Havranek and Irsova (2018) find that employment-based measures of FDI tend to reduce the estimates of spillover coefficients. The coefficients from columns 5 and 6 corroborate this finding when compared to the estimate in column 1 of 5.4.

[37] We refer to Pieterse et al. (2018) for an in-depth discussion of the generation of employment variables.

Changing Time Dimensional Aspects

Next, we check whether the result is robust to two time-dimensional changes. In column 7, we extend the timespan of the panel to incorporate the tax year 2012. One could imagine that a long and persistent exposure to FDI is necessary before learning effects materialise. A longer panel may allow us to pick up such effect. Also, with an additional year come additional observations and a higher precision of our estimates. However, given the continuous updates applied to the CIT tax forms, all variables needed to identify foreign firms are not available in the tax year 2012. This is why we excluded this year from the analysis in the first place. We employ a two-step approach to tackle the issue. First, we use the foreign firm-identifiers that are available in 2012. Second, we use the identification of foreign firms for the years where we do have good data in order to extrapolate whether a firm is foreign or not in 2012. If a firm is classified as foreign in all years it is observed after 2012, we assume that the firm is also foreign in 2012. From Column 7 it is clear that the coefficient on $Forward_{jpt-1}$ does not significantly change when 2012 is included. Neither does the longer panel change the significance level of the backward and horizontal spillover coefficients.

In column 8, we change the time identifier in the regression analysis from tax year to a time indicator from the CIT-IRP5 Panel that is better aligned with the financial year of each firm. The SARS tax year runs from the beginning of March to the end of February. The tax year 2013 ended in February 2013. The tax year 2014 ended in February 2014, and so on. The employment information extracted from the IRP5 data follows the tax year one-to-one. The CIT data, however, follows each firm's individual financial year end (FYE). In the CIT-IRP5 Panel, the *tax year variable* allocates all firms with a FYE in 2013 to the tax year 2013. Similarly, all firms with a FYE in 2014 fall into the tax year 2014, and so on. This means, for instance, that a firm with a FYE at the end of December 2014 would be assigned to the 2014 tax year even though the overlap between its FYE and the tax year only covers January and February of 2014. This is problematic because it means that the employment data from the IRP5 forms (following the tax year) does not always align with the financial data reported in the CIT tax forms (following each firms' FYE). The issue applies to 11-14 per cent of firms, whose FYE do not follow the SARS tax year (Pieterse, Gavin, and Kreuser 2018). The *financial year variable* corrects this issue by aligning the CIT data and the IRP5 data as closely as possible. For instance, if a firm's FYE falls before August 2014, its financial data is kept in the 2014 tax year. If the FYE falls after August, it is pushed to the next tax year. This procedure ensures that the IRP5 data corresponds to at least six months of a firm's financial year. Therefore, the financial year variable is 'the most appropriate time variable to use in regression' (Pieterse et al. 2018:14). However, we have not used the financial year variable in the main analysis because it pushes firms from 2012, where it is difficult to identify foreign firms, into 2013, where our analysis begins. Column 8 shows that our estimates are not sensitive to this judgement call.

Table A8: Robustness checks, $EC_{it}^{topline}$ – OLS

	Preferred specification	**Dependent variable**			**Employment based FDI measure**		**Temporal dimension**	
	From Table 3, column 8	*$EC_{it}^{topline}$, fitness-algo., HS4*	*$EC_{it}^{topline}$, complexity-algo., HS6*	*Outliers removed*	*All source income codes*	*3601 source codes*	*Including 2012*	*Firm financial year*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Horizontal_{jpt-1}$	-0.065	-0.048	0.012	-0.033	-0.165	-0.166	-0.008	-0.034
	(0.073)	(0.053)	(0.052)	(0.070)	(0.104)	(0.104)	(0.060)	(0.075)
$Backward_{jpt-1}$	0.366	-0.064	-0.479	0.590	-0.641	-0.647	0.810	2.579
	(2.183)	(2.008)	(1.651)	(1.962)	(2.329)	(2.294)	(2.014)	(2.291)
$Forward_{jpt-1}$	5.434***	3.294***	3.904***	3.751***	4.371***	4.312***	5.295***	4.032***
	(1.420)	(1.262)	(1.135)	(1.356)	(1.505)	(1.483)	(1.329)	(1.338)
$Size_{it-1}$	0.049***	0.048***	0.042***	0.035***	0.048***	0.048***	0.046***	0.044***
	(0.011)	(0.009)	(0.008)	(0.011)	(0.011)	(0.011)	(0.010)	(0.012)
$LabourProductivity_{it-1}$	0.020*	0.023***	0.025***	0.023*	0.021*	0.021*	0.022*	0.016
	(0.011)	(0.008)	(0.006)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
$R\&DIntensity_{it-1}$	0.008	0.006	0.005	0.013**	0.008	0.008	0.009*	0.006
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$Wage_{it-1}$	0.022**	0.015*	0.022***	0.025**	0.022**	0.022**	0.019**	0.022**
	(0.010)	(0.008)	(0.007)	(0.010)	(0.010)	(0.010)	(0.008)	(0.010)
$CountryDiversification_{it-1}$	0.019***	0.012***	0.012***	0.018***	0.019***	0.019***	0.021***	0.022***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$ProductDiversification_{it-1}$	0.017***	0.013***	0.012***	0.016***	0.017***	0.017***	0.017***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$EC_{\ it-1}^{all}$	0.379***	0.506***	0.510***	0.332***	0.380***	0.380***	0.381***	0.369***
	(0.020)	(0.027)	(0.023)	(0.019)	(0.020)	(0.020)	(0.019)	(0.019)
$IC_{\ it-1}^{all}$	0.073***	0.081***	0.109***	0.067**	0.073***	0.073***	0.090***	0.081***
	(0.026)	(0.020)	(0.022)	(0.026)	(0.026)	(0.026)	(0.025)	(0.025)
Fixed effects								
Industry	YES	YES	YES	YES	YES	YES	YES	YES
Province	YES	YES	YES	YES	YES	YES	YES	YES
Year	YES	YES	YES	YES	YES	YES	YES	YES
Industry x Year	YES	YES	YES	YES	YES	YES	YES	YES
Province x Year	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,442	5,485	5,485	5,261	5,442	5,442	7,149	5,369
R-squared	0.418	0.478	0.545	0.409	0.418	0.418	0.424	0.407

For table A8 above:

Notes: robust standard errors, clustered at the province-industry level, in parenthesis. p<0.01, ** p<0.05, * p<0.1.
Source: author's calculations based on SARS data.

Appendix A9

Hidalgo and Hausmann's complexity-algorithm takes point of departure in the bipartite country-product network encoded in the M_{cp}-matrix as described in Section 3.3. To reiterate, the rows of M_{cp} represent different countries, and the columns different products. The cells in M_{cp} take the value 1 if a country exports a product with a revealed comparative advantage and 0 otherwise.

The first step of the complexity-algorithm involves calculating the *diversity* of countries and the *ubiquity* of products. Assuming that the more a country knows the more products it is able to produce, the country's export basket diversity constitutes a first measure of the complexity of the country. Likewise, assuming that fewer countries are able to produce more complex products means that the ubiquity of a product is negatively correlated with its complexity. Formally, diversity and ubiquity can be defined as the row sum and column sum of M_{cp}, respectively:

$$Diversity = K_{c,0} = \sum_p M_{cp} \qquad (A1)$$

$$Ubiquity = K_{p,0} = \sum_c M_{cp} \qquad (A2)$$

While ubiquity and diversity are good initial measures of country- and product complexity, they do not take into account that certain products (such as diamonds) might be very rare, not because they are difficult to produce, but because only a few countries possess the natural resources necessary for their production. Therefore, it is necessary to correct the ubiquity measure for diamonds by the complexity of countries producing that product. Similarly, the measure of country complexity (diversity) needs to be corrected by the complexity of the products it produces. It follows, that it is possible to use one of the above equations to correct the other and visa versa. In this way, the complexity-algorithm continuously refines the rough complexity measures above by jointly and iteratively calculating the average value of the measures obtained in the previous iteration of the algorithm. Hidalgo and Hausmann (2009) refer to this as the Method of Reflections. For N iterations, the algorithm can be written as:

$$K_{c,N} = \frac{1}{K_{c,0}} \sum_p M_{cp} * K_{p,N-1} \qquad (A3)$$

$$K_{p,N} = \frac{1}{K_{p,0}} \sum_c M_{cp} * K_{c,N-1} \qquad (A4)$$

The interpretation of $K_{c,N}$ and $K_{p,N}$ changes between even and odd iterations. Even numbered iterations of $K_{c,N}$ are generalized measures of diversification, whereas odd numbered iterations of $K_{p,N}$ relate a product to the diversity of countries exporting that product, and it can be interpreted as product complexity. In other words, the complexity-algorithm allows us to classify complex countries as those exporting a large set of complex products, and complex products as those only being exported by a few complex countries.

The algorithm should be stopped when no more information can be derived from M_{cp} and there is perfect rank correlation between the complexity scores in iteration *N* and *N+1*. We stop the algorithm at N13 following the approach in Javorcik et al. (2018)[38]. Finally, the PCI index is obtained based on $K_{p,13}$: the value for each product, *p,* is normalized by subtracting from each value the mean of $K_{p,13}$ and dividing by its standard deviation. In the regression analysis we use complexity scores derived from 2012 world trade data (corresponding to the SARS tax year 2013).

Table A9: Exploring alternative explanations

	Current ratio (liquidity)	Balanced panel (entry/exit)
	(1)	(2)
$Horizontal_{jpt-1}$	-0.069	-0.039
	(0.077)	(0.085)
$Backward_{jpt-1}$	0.244	1.833
	(2.188)	(2.196)
$Forward_{jpt-1}$	5.734***	5.066***
	(1.565)	(1.384)
$CurrentRatio_{it}$	0.000	
	(0.000)	
All controls included	YES	YES
All fixed effects included	YES	YES
Observations	5,354	3,940
R-squared	0.420	0.444

Notes: robust standard errors in parentheses. Controls: $Size_{it-1}$, $LabourProductivity_{it-1}$, $R\&DIntensity_{it-1}$, $Wage_{it-1}$, $CountryDiversification_{it-1}$, $ProductDiversification_{it-1}$, EC_{it-1}^{all}, and IC_{it-1}^{all}. Fixed effects: year, industry, province, province-year and industry-year. *** p<0.01, ** p<0.05, * p<0.1.

Source: author's calculations based on SARS data.

[38] We check the rank correlation using Spearman's correlation. We never achieve a coefficient exactly equal to -1, but we suspect that inadequate decimal numbers in MATA (Stata's matrix program) explain why. In iteration 13, the rank correlation is almost perfect (-0.9987).

Appendix A10

Table A10: Heterogeneous impacts across domestic firm characteristics

	Pre-existing complexity	Size
	(1)	(2)
$Horizontal_{jpt-1}$	-0.066	-0.068
	(0.073)	(0.073)
$Backward_{jpt-1}$	0.203	0.364
	(2.145)	(2.183)
$Forward_{jpt-1}$	7.472***	6.235***
	(2.356)	(1.543)
$Forward_{jpt-1} * D_{it-1}^{complexity}$	-3.390	
	(2.506)	
$D_{it-1}^{complexity}$	0.237***	
	(0.083)	
$Forward_{jpt-1} * D_{it-1}^{size}$		-1.626
		(1.534)
D_{it-1}^{size}		0.009
		(0.070)
All controls included	*YES*	*YES*
All fixed effects included	*YES*	*YES*
Observations	5,442	5,442
R-squared	0.420	0.419

Notes: robust standard errors in parentheses. Controls: $Size_{it-1}$, $LabourProductivity_{it-1}$, $R\&DIntensity_{it-1}$, $Wage_{it-1}$, $CountryDiversification_{it-1}$, $ProductDiversification_{it-1}$, EC_{it-1}^{all}, and IC_{it-1}^{all}. Fixed effects: year, industry, province, province-year and industry-year. *** p<0.01, ** p<0.05, * p<0.1.

Source: author's calculations based on SARS data.