Foreign Investment and Domestic Productivity: Identifying Knowledge Spillovers and Competition Effects

Revision: October 2018

Christian Fons-Rosen[1]
Sebnem Kalemli-Ozcan[2]
Bent E. Sorensen[3]
Carolina Villegas-Sanchez[4]
Vadym (V.) Volosovych[5]

1: Universitat Pompeu Fabra, CEPR, and Barcelona Graduate School of Economics, Spain
2: University of Maryland, CEPR, and NBER, the USA
3: University of Houston and CEPR, the USA
4: ESADE - Universitat Ramon Llull, Spain
5: Erasmus University Rotterdam, Erasmus Research Institute of Management, the Netherlands; Tinbergen Institute, The Netherlands

Foreign Investment and Domestic Productivity: Identifying Knowledge Spillovers and Competition Effects[*]

Christian Fons-Rosen, Şebnem Kalemli-Özcan, Bent E. Sørensen,
Carolina Villegas-Sanchez, and Vadym Volosovych

October 2018

Abstract

We study the impact of FDI on the productivity of host-country firms. FDI has positive spillovers only when foreign and domestic firms use similar technologies. Channeling FDI to sectors where firms share similar technology would significantly increase productivity spillovers from FDI. We show that inventor mobility across sectors is a key channel of technology transfer. To deal with endogeneity concerns we control for sectoral productivity growth, construct a Bartik-style instrument based on the productivity growth of neighboring countries, and exploit differences in knowledge flows across sectors captured by an asymmetric patent citation matrix.

JEL: E32, F15, F36, O16.
Keywords: Multinationals, Competition, Technology, Selection, FDI, TFP.

1 Introduction

According to the 2017 World Investment Report of UNCTAD, China is the third largest foreign direct investment (FDI) recipient in the world after the United States and the United Kingdom. China has followed a development strategy since 2008, allowing FDI only in particular sectors with the justification that such policies promote growth.[1] The Chinese Administration for Industry and Commerce divides sectors as "encouraged" and "restrictive" for foreign direct investment, where the encouraged sectors are all technology-intensive sectors. At the other extreme, the United States is currently deliberating the "Foreign Investment Risk Review Modernization Act (FIRRMA)," which aims to restrict FDI in strategic sectors, most importantly, the technology-intensive sectors.[2] The United States is not the only country that restricts FDI in such sectors on the grounds of the national security concerns. A recent OECD study (Kowalski, Rabaioli and Vallejo (2017)) investigates the industrial policy of 24 developed and developing countries and finds that almost all developing countries use measures to encourage FDI in technology-intensive sectors, with China being a very active user of such policies. At the same time, countries such as the United States, the United Kingdom, and Germany do not typically encourage such FDI and the European Commission, following a request by Germany, France and Italy, is currently considering a proposal for European foreign investment screening regulation.

Our paper provides a rationale for policies that encourage FDI into technology-intensive sectors. FDI benefits the host country most when domestic firms use similar production technologies to those of the foreign firms. We investigate the productivity effects of FDI on domestic firms in six European countries, where domestic firms are defined as firms with no foreign ownership at any time during our sample period of 1999–2008.[3] Our sample of domestic firms is nationally representative for the countries we study and our FDI data that is aggregated from firm-level foreign ownership shares match OECD aggregate FDI data during our sample period.

Previous work has investigated the same question by regressing the total factor productivity (TFP) of domestic firms on output produced by foreign firms in the same two-digit sector, in the majority cases, in a single country. The effects found within the same sector are referred to as the "horizontal spillovers." The literature has typically found negligible or even negative horizontal spillovers.[4] Aitken and Harrison (1999) interpret these results as emanating from two counteracting

---

[1]See *Made in China 2025.*

[2]See https://www.congress.gov/bill/115th-congress/senate-bill/2098.

[3]Identifying domestic firms is challenging: in order to insure that the definition is consistent over time, one needs data on the ownership structure of all firms for all countries at an annual basis during the entire decade, together with changes in ownership stakes. We use the Orbis/Amadeus data which provides detailed information on parent-subsidiary linkages over time, recording the percentage of equity capital owned by domestic and foreign owners across the globe. To track the possible changes through M&A transactions, we use many vintages of Orbis/Amadeus over time and double check our final time-series on firm-level FDI with Zephyr database that records the universe of M&A transactions in our sample of countries.

[4]See Barrios, Gorg and Strobl (2011), Haskel, Pereira and Slaughter (2007), Keller and Yeaple (2009), and Aitken

forces: positive learning effects and negative competition effects. The literature also finds negative competition effects manifesting in lower market shares for domestic firms, although it is a priori possible that competition between foreign and domestic firms increases average productivity of the latter if less efficient firms are forced to exit. The lack of evidence for positive horizontal spillovers in most countries has led researchers to explore the role of "vertical spillovers"; that is, spillovers from foreign to domestic firms in different two-digit sectors. Positive vertical spillovers have been found in upstream sectors that supply goods to the sectors with FDI presence. The interpretation is that linkages between domestic upstream suppliers and downstream foreign-owned customers support learning and technology transfers between the firms (see Javorcik (2004)). These vertical spillovers are measured using coefficients from the input-output matrix to capture the trade-linkages between upstream and downstream sectors.

Our contribution differs from the previous literature in three major ways. First, we consider FDI at the four-digit sector-level while previous studies focused on FDI at a broader two-digit sector-level. Going to a more granular four-digit sector-level will allow us to better separate the learning effects from the competition effects. We test if horizontal spillovers to domestic firms in a four-digit sector are larger when they are technologically similar to the foreign firms investing in that same four-digit sector. The intuition underlying our analysis is that when foreign and domestic firms operate in the same four-digit sector, domestic firms may be affected negatively from foreign competition in the product markets, but if foreign firms use similar technology to domestic firms, these negative effects may be counteracted by positive learning effects.

Our second contribution is to search for vertical spillovers across different four-digit and technologically close sectors. We demonstrate empirically that technological similarity is not capturing the same variation as input-output linkages between sectors along the product value chain. Finally, our third contribution is to explicitly account for the endogeneity of FDI. The first order endogeneity is the sector-level productivity growth since FDI is likely to come to growing sectors. We exploit our multicountry setup so that we can simultaneously control for sector-level productivity growth and identify the independent effect of sector-level FDI. For the possibility that FDI is endogenous not only to sector-level productivity but endogenous to sector-level productivity only in certain countries, we undertake a Bartik-style instrumental variable analysis using neigbouring country-sector productivity growth.

Our methodology relies on the construction of measures of "technology closeness" across sectors. Firms that file patents have to classify these into patent classes and the main idea is that if firms patent in similar patent classes (or, for brevity, have similar patents), they are technologically close. Hence, a sector has "technologically close firms" when many of its firms patent in the same technology classes. Similarly, if firms in two different sectors tend to have similar patents, then

and Harrison (1999), and Harrison, Martin and Nataraj (2013).

these sectors are considered technologically close to each other. We develop sector-level measures of technology closeness using firm-level patent data from the United States.[5] The fact that technological distances in our European sample of countries comes from the U.S. patent data is crucial to address another endogeneity concern that domestic firms adjust their patenting to the entry of foreign firms.[6] This strategy resembles the common practice of applying the Rajan-Zingales sector-level indicators measuring the need for external finance, which uses U.S. firm-level data and then uses those values in cross-country settings. Specifically, for each firm we construct a vector of patent shares across patent classes. Our sectoral measure is then the average of the pairwise un-centered correlations across these vectors weighted by firm R&D.

With the help of these sectoral technology closeness measures, we can break up the existing measures of horizontal spillovers from FDI, which are used to test for spillovers in a given two-digit sector, into three different measures: (1) the effect of FDI on competition; (2) learning from foreign firms using similar technologies and operating in the *same* four-digit sector (positive horizontal technology spillovers); and (3) learning from foreign firms operating in *different* four-digit sectors that are technologically close to the sector of the domestic firm (positive vertical technology spillovers).

In order to fix ideas, consider an example from the French data in our sample. The French company Valinox, a manufacturer of highly specialized steel tubing for nuclear steam generation, competes with the German multinational Salzgitter Mannesmann Precision (SMP), which operates in France in the same four-digit sector. The presence of SMP may hurt the productivity of Valinox by undercutting Valinox's market shares but, at the same time, Valinox benefits if it can adopt some of SMP's cutting-edge technology. Valinox may buy a patent owned by SMP, learn about new processes, mimic the organization of production and sales, or mimic management practices of SMP, for example, by hiring workers who acquired skills while working for SMP. To be concrete, in the NACE sector 2420, to which Valinox and SMP belong, on average, 23 percent of patents registered belong to the same technological patent classes.[7]

Valinox may also benefit from technological spillovers from foreign firms in other four-digit sectors like the Dutch producer Constellium Montreuil Juigne, a leading supplier of aluminum products and solutions for aerospace, transportation, and defense worldwide. It operates in France in the same broad two-digit sector 24 as Valinox, but in a different four-digit NACE sector (sector 2442). Hence, Constellium does not directly compete with Valinox. However, steel and aluminum are the two most common metals used and they share similarities in applications and design and

---

[5]The U.S. patent data is from Bloom, Schankerman and Van Reenen (2013).

[6]In the Appendix, we show that domestic firms increase their patenting after foreign entry when FDI takes place in sectors that are technologically close.

[7]In our paper we use the statistical classification of economic activities in the European Community, abbreviated as NACE; the term NACE is derived from the French Nomenclature statistique des Activités économiques dans la Communauté Européenne.

therefore Valinox may benefit from technological advances achieved by Constellium.[8] Firms in sector 2420 (such as Valinox) and firms in sector 2442 (such as Constellium) use similar technologies, reflected in the 31 percent patent overlap for firms across these two sectors. We measure the technological distance between four-digit sectors in a similar fashion as within sectors; specifically, it is the average of the pairwise un-centered correlations across firms taken from each of the sectors of these vectors weighted by firm R&D.

We regress TFP of domestic firms on the three measures that we have constructed.[9] Our first measure is a standard horizontal spillover measure, i.e. the fraction of output produced by foreign firms, but in a given four-digit sector. The second measure is technology-closeness-weighted horizontal spillovers in the four-digit sector; namely, the fraction of output produced by foreign firms in a given four-digit sector weighted by our sector-level measure of technology closeness across firms in that same sector. The third measure is technology-closeness-weighted vertical spillovers, that is the share of output produced by foreign firms in each of the four-digit sectors (other than the domestic firm's), weighted by the technological distance from the four-digit sector of a given domestic firm in the same broad two-digit sector. Our approach is related to Bloom, Schankerman and Van Reenen (2013), who test for productivity spillovers from R&D, whereas we test for productivity spillovers from FDI.

Our results are as follows. Regressing firm TFP on our technology-closeness-weighted FDI measure, we find that domestic firms obtain positive horizontal productivity spillovers: domestic firms benefit from operating in the same sector as foreigners if firms in that sector are technologically close. This contrasts to the results we obtain when we add the unweighted horizontal spillovers measure into the same regression; horizontal spillovers are negative in sectors where firms are technologically distant. Similar to Aitken and Harrison (1999), we interpret these negative effects to be the result of the dominating competition effects. We also find positive vertical spillovers from FDI undertaken in other four-digit sectors, only when these other sectors are technologically close to the four-digit sector of the domestic firms. Importantly, we show that this measure does not capture the standard vertical FDI spillovers based on customer-supplier relationships constructed from the input-output matrices since we explicitly include this type of vertical spillovers in the regressions. All in all, our results are similar in spirit to Keller and Yeaple (2009), who show that

---

[8]Both steel and aluminum can be altered by alloying, cold-working, and heat-treating, and formed by rolling, extruding, drawing, machining, and other mechanical processes. They can also be cast to a high tolerance; see the website of the steel industry expert Satyendra Kumar Sarna http://ispatguru.com/comparison-of-steel-with-aluminum/.

[9]While our baseline is a measure of revenue TFP, following the literature, results look very similar when we use a measure of physical TFP, following a procedure suggested by Marin and Voigtlander (2014). As is well known in the literature, changes in revenue TFP can be due to changes in physical productivity or to changes in prices and the separate effects can not be measured in the absence of firm-level price data. This issue is particularly important because markups of domestic firms may respond endogenously to competition, leaving the effect on physical TFP in doubt. We also estimate markups following De Loecker and Warzynski (2012) and use them in a separate regression analysis.

FDI spillovers are stronger in high-technology industries.

A potential source of bias can be due to FDI being endogenous. Foreign companies may invest in sectors that have high productivity growth or in sectors that are technologically close to high productivity sectors. As long as domestic firms' productivity is correlated with any factors that affect the entry decision of the multinationals, the estimates of spillovers from foreign to domestic firms may be biased. We address this issue in several ways. First, taking advantage of the fact that our data spans six countries, we control for sectoral productivity growth that may attract FDI to certain sectors, regardless of location, using sector×year-fixed effects. These fixed effects absorb sector-specific patterns while still allowing us to identify the effect of sector-level changes in FDI over time across countries. This identification methodology has not been used in the literature on FDI spillovers before given that the majority of previous papers have been single-country studies.[10] We also include country×year fixed effects to account for potential country-specific policy changes implemented to attract FDI. As a result, our estimates are not driven by different dynamic patterns in productivity between sectors and/or between countries or any other omitted variable at these levels.

Second, we also address potentially remaining endogeneity at the sector×country×year level. Multinationals might decide to enter to specific country-sector combinations over time driven by motives to acquire technological knowledge from domestic firms that are successful at innovating in the years of foreign entry. This might be more likely to take place when domestic firms are technologically close to the potential foreign entrants. It is also possible that foreign firms may wish to enter the underperforming sectors with high growth potential (maybe taking advantage of "fire-sales" of distressed firms), which could explain the negative competition effects that we have found. To address potential reverse causality resulting from such selection, we perform an instrumental variables analysis (IV), where we implement a Bartik-type IV procedure. We instrument growth of FDI in a given country-sector-year with the neighboring countries' growth in the same sector and year. These IV results support causal horizontal spillover effects of FDI in technologically close sectors.

Our results on vertical spillovers are probably less likely to be affected by endogeneity concerns, as it is less plausible for FDI to take place in a given four-digit sector with the expectation of getting benefits from booming firms in other four-digit sectors. Nevertheless, we examine such possibility and devise the following empirical test. Assume for simplicity of exposition that a two-digit sector consists of three four-digit sectors, labelled as $s4_1$, $s4_2$, and $s4_3$. Let us focus on domestic firms in sector $s4_1$ and assume that foreign firms enter sectors $s4_2$ and $s4_3$. Assume that foreign firms of $s4_2$ enter as a reaction to productivity advances by domestic firms in sector $s4_1$, because globally sector $s4_2$ happens to learn a lot from $s4_1$ and it is easier to benefit from these spillovers in geographical

---

[10]An exception is Gorodnichenko, Svejnar and Terrell (2014), who study FDI in 17 transition economies and show that the vertical spillovers vary by industry.

proximity. Further assume that sector $s4_3$ has never learned anything from $s4_1$, so that the FDI perform in that sector must be for reasons other than events or shocks occuring in sector $s4_1$. If this is the case, foreigners entering $s4_2$ should cite patents of the sector $s4_1$ more often than foreigners entering to sector $s4_3$. This would create a stronger relation between domestic productivity in sector $s4_1$ and FDI in sector $s4_2$ due to reverse causality based on absorbing technology from sector $s4_1$ more intensely. From the perspective of domestic sector $s4_1$, FDI into sector $s4_2$ is endogenous while FDI into sector $s4_3$ is exogenous. In the estimations we split FDI into its endogenous and exogenous components and find estimated coefficients of similar magnitude, giving support for our causal interpretation. To do this split, we make use of the asymmetric nature of patent citations as shown by Acemoglu, Akcigit and Kerr (2016).

What do our technology closeness measures capture? We argue that these measures capture the benefits from "operating in the same technology space," but in general we are not explicit about how domestic firms may acquire knowledge from foreign firms. One channel could be the mobility of inventors who transmit their knowledge from firm to firm. If inventors move frequently between two four-digit sectors, then these sectors are likely to have more patent overlap as a result of such knowledge transfers, thereby scoring high on our technology closeness metric. The data on inventor mobility can provide an important check on our interpretation. We construct a measure of mobility of inventors between sectors (imputed from underlying data for mobility between technology classes). Using this new measure of FDI spillovers weighted by inventor mobility, we find qualitatively similar results to those with spillover measures based on patent and citations data; inventor mobility is a channel of knowledge transfers between technologically close sectors that can benefit domestic firms, indicating true learning spillovers. Similarly to the asymmetric patent citations exercise discussed previously, we again reject reverse causality after decomposing FDI into its endogenous and exogenous components based on how often inventors move across a given pair of sectors. Specifically, we test whether the measured FDI spillovers from different four-digit sectors to domestic firms are stronger when inventors migrate more intensively from the domestic sector to those FDI sectors (which would suggest reverse causality). We find no such differences as the estimated coefficients are similar to each other for both FDI components.

Finally, we perform a quantification exercise that reveals an important policy implication of our results. We calculate the (counterfactual) effects of FDI flowing to technologically close sectors rather than flowing equally to all sectors. Strikingly, this would more than double the productivity spillovers from FDI on domestic firms. If all FDI were flowing to the least technologically close sectors, spillovers would even be negative! These patterns likely explain the wide range of existing estimates for horizontal spillovers in two-digit sectors where competition effects and knowledge spillovers were confounded.

The rest of the paper is structured as follows. Section 2 presents our data and measures.

Section 3 presents our econometric specification and results. Section 4 presents robustness analysis and quantitative implications based on counterfactual exercises. Section 5 deals with endogeneity and presents results from instrumental regressions framework. Section 6 presents results on the inventor mobility channel. Section 7 concludes.

2 Data and Construction of Variables

We use the Orbis global database, from Bureau van Dijk (BvD)—a Moody's Analytics company. Orbis is the largest cross-country firm-level database, covering over 200 countries and 200 million firms that can be used for research focusing on linking firms' financial accounts, ownership structure and production decisions. The database includes all industries and both private and public firms. BvD collects data from various sources, in particular, publicly available national business registries (BR), and harmonizes the data into an internationally comparable format.

The coverage of firms varies both by country, industry, over time and across variables. The reason for variation in firm coverage by country is that different countries have different laws in terms of which firms require to file their financial accounts to the BR.[11] For countries where the laws require every firm to file to the national BR, the data obtained via Orbis will not differ from what is obtained from country statistical offices in terms of financial accounts.[12] We focus on six such European countries, Belgium, Finland, France, Italy, Norway, and Spain, for the years 1999–2008 that has representative national coverage as shown in Table A1.[13]

The coverage of firms in Orbis database can vary by time and industry and this may be a source of discrepancies between various studies because it is a common practice in the literature of using a single vintage of Orbis database. As, we explain in detail in Kalemli-Ozcan, Sørensen, Villegas-Sanchez, Volosovych and Yesiltas (2015), the only way to get around this problem and have consistent coverage of firms over time and by industry is to use the historical vintages and match the firm data over time using unique firm identifiers. If a single vintage is used, firms will be missing since Orbis drops firms over a certain period of time from the database and also variables, such as value-added and intermediate inputs, will be missing since every vintage does not cover all

---

[11]There is a common misconception that data from countries' national statistical offices always have better coverage than Orbis. Certain countries' national statistical offices might have administrative surveys on certain samples of small firms that can cover some of the differences in coverage of firms' financial accounts, if these countries laws do not require small firms to file their financial accounts. A case in point is the United States, where private firms are not required to file financial accounts but there are select surveys covering certain set of firms in certain years such as the Federal Reserve Board of Governors' survey on "small business finance," which is a repeated cross-section that comes in four waves and covers only 3000-5000 firms.

[12]Country censuses are administrative datasets and will cover the universe of firms in a country, however, census datasets typically do not provide information on firms' financial accounts as business registries do.

[13]Finland is the exception since in Finland only firms with more than 25 employees are required to file financial accounts. We keep Finland in the analysis because our final analysis focuses on firms with more than 10 employees to be consistent with the FDI literature.

the variables. The industry classification will also be misleading since these classifications change over time due to firms' expanding their operations and/or firm and industry ID changes done by the national statistical offices. With the missing information, most of the time Orbis single vintage data will over-represent larger firms and under-represent smaller firms, leading to imputations and re-weighing of the data to increase the representation of small firms. As we show in Kalemli-Ozcan, Sørensen, Villegas-Sanchez, Volosovych and Yesiltas (2015), there will not be a need to re-weigh the data and doing imputations if the historical vintages were used, which will deliver the nationally representative data that mimics the firm size distributions given by the official statistics of each country as shown in Table A2.[14]

In the case of the ownership structure, which we need to track for constructing our FDI statistics over time, using historical vintages is the absolute minimum practice since the current vintage of Orbis database will only provide current ownership information. As a result, we use several vintages of Orbis database together with Zephyr data that provides detailed financial and ownership information on parent-subsidiary linkages with percentage of capital owned by domestic and foreign owners across the globe. Firms are linked to their domestic and foreign parents through unique ID numbers, and this allows us to construct precise firm-level measures of changes in FDI over time based on changes in ownership stakes.[15]

The financial information on the balance sheets requires checks and cleaning. We follow Kalemli-Ozcan, Sørensen, Villegas-Sanchez, Volosovych and Yesiltas (2015) for cleaning the data and we refer the reader there for details. However, we would like to point out a key issue that is critical for the current paper. We need to make sure we assign correctly the foreign and domestic owners of the companies that span the European countries as business groups such as the BMW and Vodafone groups. Financial accounts of companies come both as consolidated accounts, which aggregates all parent and subsidiary accounts, and also at unconsolidated level, that shows individual subsidiary accounts separately. Using both will be double-counting as shown by Kalemli-Ozcan, Sørensen and Volosovych (2014); hence, we use only unconsolidated accounts. Finally, to be consistent with the FDI spillovers literature, we use only data from the manufacturing sector, with information on all the four-digit sectors within the manufacturing sector. We exclude micro enterprises (those with less than ten employees according to the European Commission definition), also to be consistent with the FDI literature. Our final sample accounts for more than 90% of manufacturing output in the countries under study, and it mimics the size distribution of firms provided by official national sources, delivering national representation. Aggregate output of the foreign firms also tracks the official statistics on FDI, verifying that we capture all the foreign firms in a given economy as shown in Figure A.1.

---

[14]Again with Finland being the exception for reasons explained above.

[15]Official, balance-of-payments definition of FDI is equity ownership in access of 10% of a firm's capital stock. 90 percent of FDI is conducted through acquisitions, according to Barba-Navaretti and Venables (2004).

We next describe the main firm-level variables used in the analysis: revenue productivity, foreign ownership, and the spillover variables. More details on the cleaning process and firm-level statistics are provided in Appendix A.

2.1 Firm-Level Productivity

Our main dependent variable is total factor productivity of domestic firms.[16] We assume that firm i's output is given by a Cobb-Douglas production function,

$$Y_{it} = A_{it} L_{it}^{\beta_\ell} K_{it}^{\beta_k} \,, \tag{1}$$

where firm value added, Y_{it}, is a function of physical productivity (A_{it}) and firm inputs (L_{it}, K_{it}). L_{it} is labor input, K_{it} is capital input, β_k is the output elasticity of capital, and β_ℓ is the output elasticity of labor. We measure nominal value added, $P_{it}Y_{it}$, as the difference between gross output (operating revenue) and materials. We do not observe prices at the firm level, and we calculate "real" output, Y_{it}, by dividing nominal value added with the Eurostat two-digit industry price deflators.[17] Labor input, L_{it}, is measured as the firm's wage bill (deflated by the same two-digit industry price deflator).[18] Finally, we measure the capital stock, K_{it}, as the book value of fixed assets, deflated by the price of investment goods.[19]

To obtain firm-level productivity estimates, we follow the approach suggested in Wooldridge (2009)—see Appendix B for a detailed description of the estimation procedure. We estimate the production function by country and two-digit sector (Table B.1 in Appendix B shows the estimated elasticities) and winsorize the resulting distribution at the 1 and 99 percentiles by country.

2.2 Firm-Level Foreign Ownership

As explained below, to have the FDI measures based on a foreign ownership definition that is consistent over time requires data on the ownership structure of all firms for all countries in our sample at an annual basis during the entire decade together with the changes in ownership stakes. We use the Orbis-Zephyr data that provides detailed ownership information on parent-subsidiary linkages

---

[16]See next section for the definition of a domestic firm.

[17]Norway and France do not have good coverage of industry price deflators at the two-digit level, and we use the total manufacturing industry price deflator for these two countries.

[18]We follow Hsieh and Klenow (2009) and use the wage bill, rather than the head count, to help adjust for differences in the quality of workers across firms.

[19]We use country-specific prices of investment from the World Development Indicators to deflate the book value of fixed assets. The capital stock includes both tangible and intangible assets because in 2007 there was a change in the accounting system in Spain and leasing items that until 2007 had been part of intangible fixed assets were from 2008 included under tangible fixed assets. To avoid breaks in the time series, we opt to use the sum of tangible and intangible fixed assets as our measure of capital stock. Our results are robust to estimating TFPR using tangible fixed assets as a measure of capital in the other countries of the sample.

with percentage of capital owned by domestic and foreign owners across the globe. The ownership section of Orbis contains detailed information on owners of both listed and private firms, including name, country of residence, and type (e.g., bank, industrial company, private equity, individual) and tracks changes in ownership over time. Orbis collects ownership data from official registers (including SEC filings and stock exchanges), annual reports, private correspondence, telephone research, company web-sites, and news wires. We define a firm to be "domestic" if it never had *any* foreign owner during the sample period.

The database refers to each record of ownership as an "ownership link." An ownership link indicating that an entity A owns a certain percentage of firm B is referred to as a "direct" ownership link. BvD records direct links between two entities even when the ownership percentages are very small (sometimes less than one percent). For listed companies, very small stockholders are typically unknown.[20] We compute "foreign ownership" as the sum of all percentages of *direct* ownership by foreigners. For example, if a company has three foreign owners with stakes of 10, 15, and 35 percent, the foreign ownership fraction for this company is 60 percent. If the information on "direct" ownership is missing, we substitute this missing information with the global ultimate owner information.

2.3 Sector-Level FDI Spillovers

In the literature, measures of FDI have typically been calculated using two-digit sector classifications. Because the number of firms in our dataset is very large and a four-digit industry classification is available, we are able to refine the measures to the four-digit level as described below.

Horizontal FDI/Spillovers. We follow Javorcik (2004) and compute exposure to foreign-owned firms that produce in the same sector. We define the measure HORIZONTAL as the share of "foreign output" in sector output:

$$\text{HORIZONTAL}_{s4,t} = \frac{\sum_{i \in s4} \text{fo}_{i,t} \times \text{go}_{i,t}}{\sum_{i \in s4} \text{go}_{i,t}} , \tag{2}$$

where $s4$ refers to the four-digit sector of firm i, $\text{go}_{i,t}$ refers to gross output (operating revenue) of firm i at time t, and $\text{fo}_{i,t}$ refers to the percentage of capital of firm i that is foreign owned.

Vertical FDI/Spillovers. Again, we follow Javorcik (2004) and define

$$\text{VERTICAL\_IO}_{s4,t} = \sum_{\substack{\tilde{s}4 \in s2(s4) \\ \tilde{s}4 \neq s4}} \alpha_{s4,\tilde{s}4,t} \times \text{HORIZONTAL}_{\tilde{s}4,t} , \tag{3}$$

where $\alpha_{s4,\tilde{s}4,t}$ is the input-output coefficient that records the fraction of its own output that sector $s4$

---

[20]Countries have different rules for when the identity of a minority owner needs to be disclosed. France requires listed firms to disclose all owners with a stake larger than five percent while Italy requires listed firms to disclose all owners with a stake larger than two percent.

supplies to each given sector $\tilde{s}4$. This measure is extensively used in the literature, often exploiting data from The World Input-Output Database (WIOD). This data set provides time-series of world input-output tables for forty countries at the two-digit industry level. To construct our measure at the four-digit level, we use the U.S. input-output table from the Bureau of Economic Analysis (BEA).[21, 22]

2.4 Sector-Level Technology Closeness

We utilize the technology closeness measures provided by Bloom, Schankerman and Van Reenen (2013) for pairs of firms in the United States. They use firm-level accounting data (sales, employment, capital, R&D expenditure, etc.) and market value data from Compustat over the period 1980–2001 to match firms to U.S. Patent and Trademark Office data from the NBER data archive. After matching, an unbalanced panel of 715 listed firms with at least four observations between 1980 and 2001 is available. Bloom, Schankerman and Van Reenen (2013) assign patents to different technology clusters and compute a technology closeness measure, calculated as the uncentered correlation between firm pairs, as follows: calculate the average share of patents each firm holds in each of 426 different technology classes over the period 1970–1999 and define, for each firm i, the vector of i's technological activity $t_i = (t_{i1}, t_{i2}, ..., t_{i426})$, where t_{ix} is the share of patents of firm i in technology class x. For each firm pair i, j in the sample, measures of technology closeness are constructed, following Jaffe (1986), as the uncentered correlation of patent share vectors t_i and t_j:

$$\text{tech}_{ij} = \frac{\left(t_i t_j'\right)}{\left(t_i t_i'\right)^{1/2}\left(t_j t_j'\right)^{1/2}}. \tag{4}$$

Bloom, Schankerman and Van Reenen (2013) collect firm-level R&D expenditures, labeled rd_{it} for firm i at time t, and test if a given firm i is affected by the R&D of other firms $j \neq i$ that are technologically close, as measured by tech_{ij}. Specifically, they construct a measure $\text{spilltech}_i = \Sigma_j \text{tech}_{ij} \text{rd}_j$ and test if this empirically predicts the R&D expenditures and other outcomes of firm i, finding large effects of R&D on technologically close firms.

We use these firm-level measures of technology closeness, but aggregate them to the four-digit sector-level, matched to our NACE industrial classification. The number of four-digit industries covered is lower than the number of four-digit industries available in our dataset,[23] but Table D.1

---

[21]Using the U.S.-based measures implicitly assumes that the patterns of input and knowledge flows in the six countries of our sample are close to those of the United States. If the U.S. production and input structures are imperfect for European countries, we are introducing random error in the measurement of our regressors and, therefore, reducing the probability of finding statistically significant results.

[22]The variable VERTICAL_IO$_{s4,t}$ represents what has been named in the literature as backward spillovers. A similar variable can be constructed using the input-output coefficient that records the fraction of its own output that sector $s4$ buys from each given sector $\tilde{s}4$, this is the case of forward spillovers. In Appendix C, we show further results on vertical spillovers.

[23]In our final dataset, we have 135 four-digit industries.

in Appendix D shows that there are no major differences in company size due to the use of a lower number of four-digit sectors, meaning the firms that are in the dropped sectors are of similar size to the firms in our final sample. We use time-invariant measures of technology closeness at the bilateral sector-level that pre-date the analysis. The combination time-invariant and the U.S.-firm based measures helps alleviate concerns about endogeneity of the sector-level technology closeness measures to sector-level FDI.

The following steps describe our procedure. (We use the notation of Bloom, Schankerman and Van Reenen (2013), except that we use a small-cap font for firm-level variables to contrast with the country-sector-level variables in capitals.) First, for each four-digit sector pair, we compute the *sectoral* technological closeness as the R&D-weighted sum of the technology closeness of firms i and j operating in sector pairs $s4$ and $\tilde{s}4$ respectively, normalized by the sum of R&D of all k and l firms in the two sectors, as:

$$\text{SPILL\_RD}_{s4,\tilde{s}4} = \sum_{i \in s4} \sum_{j \in \tilde{s}4} \text{tech}_{ij} \times \left(\frac{\text{rd}_i + \text{rd}_j}{\sum_{k \in s4} \sum_{l \in \tilde{s}4} (\text{rd}_k + \text{rd}_l)} \right) , \qquad (5)$$

where $s4$ and $\tilde{s}4$ refer to four-digit sectors.[24] The technology closeness measure and the firm-level R&D expenditure, rd, are time invariant because we use the average R&D expenditure for each firm.

Figure 1 displays heat maps for the values of SPILL_RD in Panel A and input-output coefficients on Panel B for the randomly selected sector 24. The figure illustrates that there is significant variation across sectors in both measures and that patterns of technological closeness are not simply capturing input-output relations.

---

[24]A given $\text{rd}_k + \text{rd}_l$ sum will enter twice in the sum in the denominator. This insures that the weights sum to unity over the double summation.

Figure 1: Heatmaps for technological closeness and input-output links

A. Technological Closeness

B. Input-Output Coefficients

2420: Manuf. of tubes, pipes, hollow profiles and related fittings, of steel
2442: Aluminium production
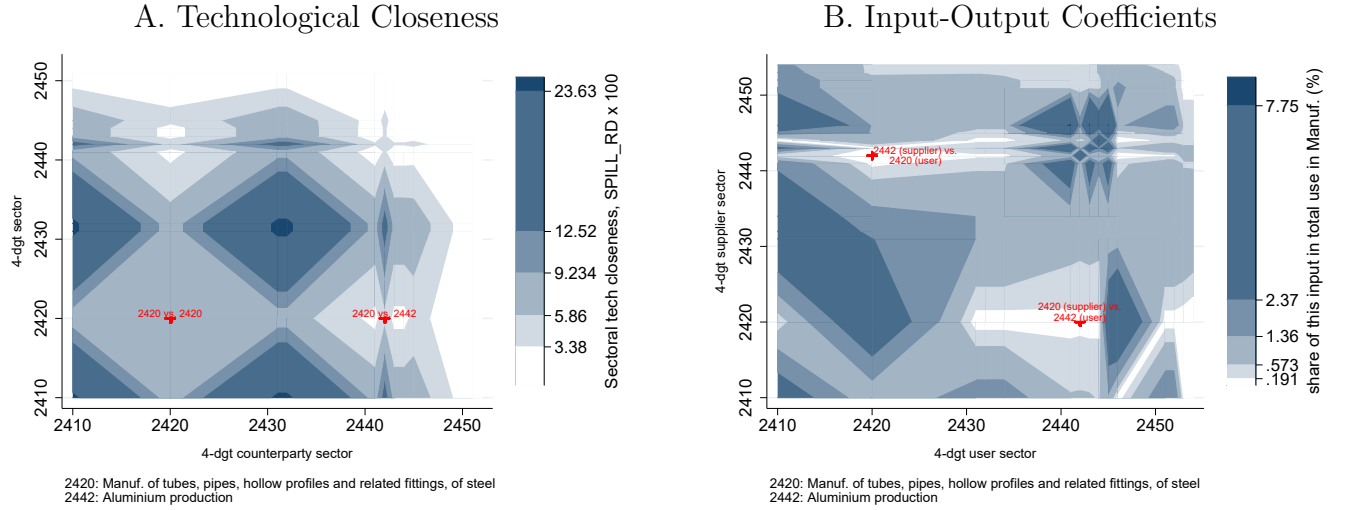
Notes: For the randomly-chosen sector 24, Panel A displays the four-digit technology closeness-measure SPILL_RD, and Panel B displays the four-digit input-output coefficients. The input-output coefficients are α-parameters in equation (3) in the main text, and they record the fraction of its own output that a given sector supplies to each other sector as intermediate input. SPILL_RD is the R&D weighted similarity of patenting between four-digit sectors, calculated using equation (5) in the main text. The data for these measures comes from U.S. sources, because of the lack of four-digit data outside of the U.S., and the measures are time-invariant. The sectors discussed in the example in Introduction are marked.

Figure 2 shows the distribution of the sectoral technology closeness measure in equation (5), where we split sectors by the OECD classification of the technology intensity of the sector.[25] Sectors that are classified by the OECD as technologically intensive (represented by the solid green line) feature larger values of our technology closeness measure, equation (5). From the figure, the SPILL_RD-measure tends to be larger in technologically intensive sectors.

Second, to account for the economic importance of each sector that is technologically linked to a given four-digit sector $s4$, we use output weights as:

$$\text{WTECH}_{s4,\tilde{s}4,t} = \frac{\text{SPILL\_RD}_{s4,\tilde{s}4} \times \text{GO}_{\tilde{s}4,t}}{\sum_{\substack{\hat{s}4 \in s2(s4) \\ \hat{s}4 \neq s4}} \text{SPILL\_RD}_{s4,\hat{s}4} \times \text{GO}_{\hat{s}4,t}}, \tag{6}$$

where $\text{GO}_{s4,t}$ is gross output of sector $s4$ at time t, and $s2(s4)$ is the 2-digit sector that includes the 4-digit sector $s4$. The numerator is the technological similarity between sectors $s4$ and $\tilde{s}4$ weighted by output and the denominator sums over all four-digit sectors in the two-digit sector. Our empirical analysis use dummy variables to control for factors affecting two-digit sectors, and we therefore design the weights to reflect the relative technological closeness *within* two-digit sectors.

---

[25]The OECD classification is based on both direct R&D intensity and R&D embodied in intermediate and investment goods. Four categories were introduced: high-, medium-high, medium-low and low technology. Figure D.1 in the appendix shows a more detailed breakdown by two-digit industry.

Figure 2: Distribution of R&D weights by technology intensity of sectors

A. All values

B. Truncated at 99 percentile
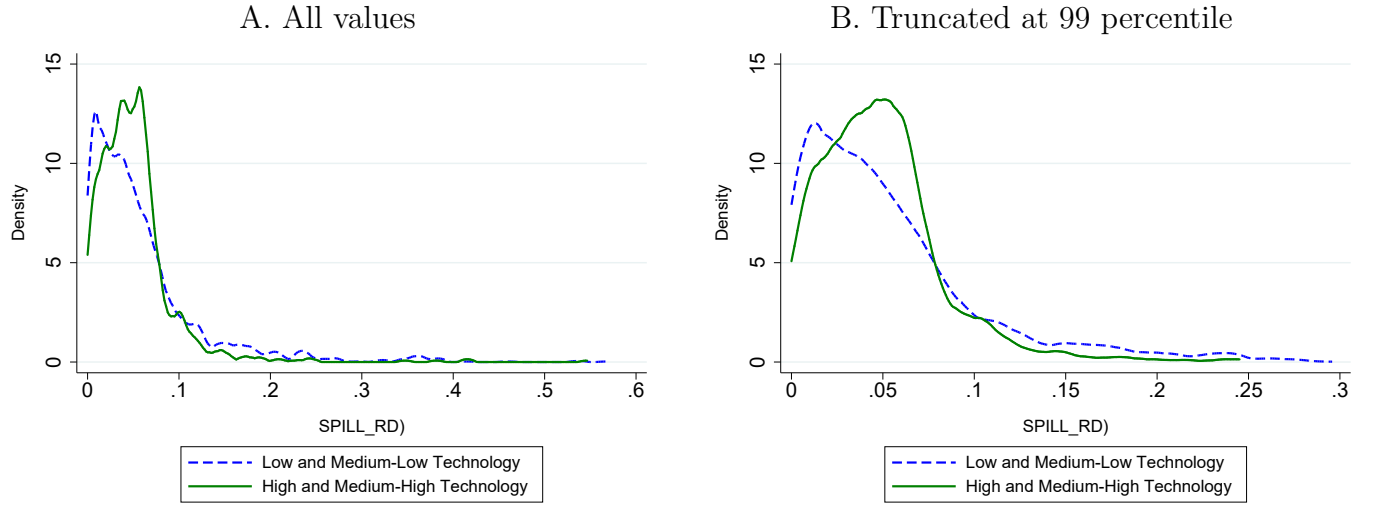
Notes: The figure displays the frequency distribution of the four-digit technology closeness-measure SPILL_RD. The SPILL_RD is the R&D weighted similarity of patenting between four-digit sectors, calculated using equation (5) in the main text. The data for this measure comes from the U.S. sources, because of the lack of four-digit data outside of the U.S. The measure is time-invariant. In Panel B, the distribution is truncated at the 99th percentile in order to more clearly bring out the differences between sectors with lower and higher technology.

2.5 Sector-Level Technology Closeness-Weighted FDI

Using the measures of technology closeness, we define the "technology closeness-weighted FDI/spillover" variables that will be our regressors. For the "weighted horizontal FDI" variable, we multiply the share of output produced by MNCs in each four-digit sector with our measure of technology closeness of firms within the sector:

$$\text{HORIZONTAL\_TEC}_{s4,t} = \text{WTECH}_{s4,s4,t} \times \text{HORIZONTAL}_{s4,t} . \tag{7}$$

The HORIZONTAL_TEC measure interacts the presence of foreign firms in the four-digit sector (HORIZONTAL) with the measure (WTECH) of the importance of technological linkages between firms within the sector. HORIZONTAL_TEC is obviously correlated with the HORIZONTAL measure, but when we include both in a regression, HORIZONTAL_TEC mainly captures the importance of closeness in the technology space, while HORIZONTAL captures the competition effect from producing similar products.[26]

For potential knowledge spillovers between different four-digit sectors, we construct a "weighted vertical FDI" variable as follows: for each four-digit sector in the two-digit sector, we multiply our measure of technology closeness with the output share of MNCs and average across all four-digit

---

[26]In order to avoid outliers, we normalize the "diagonal" indices $\text{WTECH}_{s4,s4,t}$ with the sums over all sectors.

sectors within the two-digit sector, apart from the sector that the measure is constructed for:

$$\text{VERTICAL\_TEC}_{s4,t} = \sum_{\substack{\tilde{s}4 \in s2(s4) \\ \tilde{s}4 \neq s4}} \text{WTECH}_{s4,\tilde{s}4,t} \times \text{HORIZONTAL}_{\tilde{s}4,t} \ . \tag{8}$$

This measure is constructed similarly to the VERTICAL_IO measure, but using technology closeness weights instead of weights from the input-output tables.

Figure 3 shows the relative importance of each measure in high- versus low-technology sectors (as classified by the OECD). The importance of VERTICAL_TEC (the blue bars) is higher in high-technology sectors.

Figure 3: Spillover measures by technology sector in 2007.

.
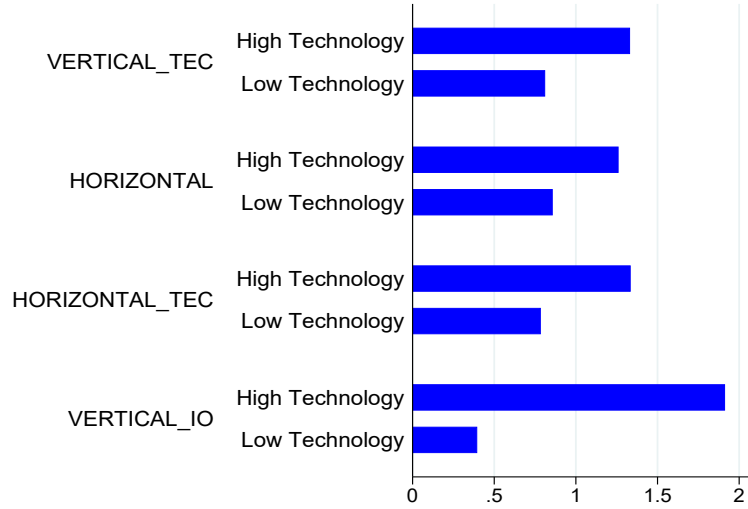
Notes: The figure displays the average values of the main regressors, defined in Section 2.3, by technological intensity of the four-digit sector. The data in these figures spans all countries in our sample but, unlike in the main paper, is for a single year 2007. We take the value for each country-sector-year and normalize with the average value for the corresponding country-year (because the scale differs across countries), and average over countries and years. The technological intensity classification of sectors used here for presentation purposes follows the OECD methodology and is based on two indicators of technology intensity reflecting, to different degrees, "technology-producer" and "technology-user" aspects: i) R&D expenditures divided by value added; ii) R&D expenditures divided by production (see OECD (2011)).

Table D.2 in Appendix D reports summary statistics for the main variables and for the final sample of firms used in the analysis. Table D.3 shows the matrix of correlations between our main variables.

3 Empirical Analysis

We regress domestic firms' TFP on our three measures and a measure of vertical FDI spillovers based on input-output linkages:

$$
\begin{aligned}
\log\left(\text{TFPR}_{i,s4,c,t}\right) \quad = \quad & \beta_1 \text{HORIZONTAL}_{s4,c,t-1} + \beta_2 \text{HORIZONTAL\_TEC}_{s4,c,t-1} \qquad\qquad (9) \\
& + \beta_3 \text{VERTICAL\_TEC}_{s4,c,t-1} + \beta_4 \text{VERTICAL\_IO}_{s4,c,t-1} \\
& + \alpha_i + \phi_{s4,t} + \delta_{c,t} + \epsilon_{i,s4,c,t}\,,
\end{aligned}
$$

where $\text{TFPR}_{i,s4,c,t}$ refers to total revenue factor productivity of firm i, in sector $s4$, country c, at time t; and the terms $\delta_{c,t}$ and $\phi_{s4,t}$ represent country-year and sector-four-digit-year fixed effects, respectively. We expect β_1 to be negative, capturing the effect of competition, while we expect β_2 to be positive, capturing the possible productivity improvement effects derived from the presence of foreign-owned firms operating in the same sector as the domestic firms. It is important that both measures of foreign presence in the four-digit sector are included because the measures are correlated by construction; for example, both will be zero if no foreign firms are present in the sector. When both measures are included, the regression estimates will for each variable, by the Frisch-Waugh theorem, capture the contribution which is orthogonal to that of the other; therefore, the simple measure of FDI is likely to capture pure competition effects, while the weighted measure of FDI is likely to capture the effect of knowledge spillovers. β_3 will be positive if domestic firms benefit from the presence of foreign investors in closely related technologically close sectors, and we expect β_4 to be positive, capturing standard backward spillovers based on input-output relationships. We include sector-year fixed effects to control for the sectoral productivity growth that can attract FDI in the first place.

Firms are heterogeneous and while most of the existing literature estimates spillover regressions by Ordinary Least Squares (OLS), heteroskedasticity is of particular concern in the current study that pools firms across different countries with potentially different firm-level volatility. Therefore, all our results are estimated by two-step feasible Generalized Least Squares (GLS), which is the efficient estimator in this case. We cluster standard errors at the country-four-digit sector-level (allowing for autocorrelation).[27]

---

[27]The first step estimates the equation by OLS, and for each firm the standard error of the innovation term is estimated as the square root of the mean (over the observations for the firm) squared residuals. In the second step, the least squares regression is repeated, weighting each firm by the inverse of its estimated residual standard error. Because the time dimension is relatively short, some firms may get extremely high weights and to avoid this, we winsorize the lower tail of the weights distribution at 5 percent. Graphical inspection of a partial correlation plot of the regression reveals that there are no obvious outliers in our second step regression. Similar results were found if the weights were obtained with a parametric model of the error variance (i.e., estimating standard errors as a function of firm characteristics).

Table 1: Horizontal Technology Spillovers Conditional on Competition

DEPENDENT VARIABLE: log FIRM REVENUE TFP
SAMPLE: DOMESTIC FIRMS

	(1)	(2)	(3)
HORIZONTAL$_{s4,t-1}$	-0.096*	-0.176***	-0.320***
	(0.048)	(0.048)	(0.080)
HORIZONTAL_TEC$_{s4,t-1}$			0.223**
			(0.076)
Observations	322,698	322,698	322,698
Firm FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Sec4-Year FE		✓	✓
Cluster	cs4	cs4	cs4
s.d.(HORIZONTAL)	0.16	0.16	0.16
s.d.(HORIZONTAL_TEC)			0.07

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. The dependent variable is log revenue firm-level productivity at time t (log TFPR$_{i,t}$). See section 2.3 for a full description of the regressors. The sample comprises domestic firms, i.e., firms that have no foreign participation over the years of analysis. All right hand side variables are lagged one period. Standard errors are clustered at the country-four-digit-sector-level. Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

In Table 1, we report the benchmark results for horizontal spillovers. Because our regressors do not have magnitudes which are immediately intuitive, we report standardized coefficients obtained from normalizing each regressor by its standard deviation (reported at the bottom of the tables of results). The benefit of reporting standardized coefficients is that the magnitudes of the reported parameter-estimates reflect the importance of the corresponding regressor in explaining the dependent variable.

The specification reported in column (1) uses the regressor HORIZONTAL and finds negative, (statistically) significant spillovers—broadly consistent with many previous papers. In column (2), where we use fixed effects to control for factors that affect entire sectors in each given year, we find that foreign presence in the four-digit sector significantly decreases the productivity of domestic firms. The estimated coefficients are semi-elasticities with the left-hand side scaled by 100; i.e., the estimated coefficients are interpreted as the predicted percentage change in the dependent variable following a unit change in the regressor. Therefore, the coefficient of –0.176 implies that an increase in foreign presence in the four-digit industry of one standard deviation decreases domestic firms' productivity by about 0.18 percent.

In column (3), we attempt to sort out competition effects and knowledge-spillover effects by including both the competition variable HORIZONTAL and the technology-weighted variable, HORIZONTAL_TEC. We obtain highly significant coefficients for both competition and knowledge spillovers, with coefficients of –0.320 and 0.223, respectively. Our interpretation is that there are large knowledge spillovers to domestic firms from foreign-owned firms in the same four-digit sector, but those are offset on average by negative competition effects. The knowledge spillover effect dominates in sectors where the foreign-owned firms are technologically close, while the negative competition effect dominates in sectors where the foreign-owned firms are not technologically close.

Our main regression, presented in Table 2, utilizes our three measures of potential FDI spillovers and condition on standard vertical spillovers measures based on input-output linkages.[28] We include the variables sequentially, in order to evaluate if the effects are robust to specification.

In the first three columns, we display regressions using only the vertical measures. In column (1), we include only VERTICAL_TEC while in column (2), we include only VERTICAL_IO. Column (3) displays results from regressions with both measures included. The VERTICAL_TEC measure is robustly estimated with a coefficient above 0.25 and highly significant. The backward-spillovers measure, VERTICAL_IO, takes a positive coefficient, but it is not significant when the technology closeness-weighted measure is included. Notice, the measure of backward spillovers is constructed at the four-digit industry level while previous studies on the role of vertical linkages have used vertical measures using I-O coefficients at the two-digit industry level. In Appendix Table E1, we show the corresponding results that are significant when we calculate input-output based vertical spillovers at the two-digit level in our sample of firms.

Column (4) verifies that the horizontal and the vertical measures are not highly correlated, as the estimated coefficients of the horizontal variables are similar to those found in Table 1 and not sensitive to whether the vertical variables are included and vice versa. This is intuitive because the measures are constructed from non-overlapping groups of firms. Column (5) is the preferred specification with our three main measures, all being robustly significant.

Overall, productivity increases with foreign presence in technologically close sectors. On average, the overall results imply that the beneficial effect of foreign presence in the four-digit sector roughly cancels out the negative competition effect; however, FDI is overall beneficial for the productivity of domestic firms because of the further impact from FDI outside the four-digit sector. An important implication of this result is that the amount of knowledge spillovers that a country enjoys from FDI depends not just on *how much* foreign investment there is, but also critically on *which sectors* foreign investment takes place in. In particular, as we quantify below, the impact will be larger when FDI is concentrated in sectors where technological spillovers to other sectors are high, and

---

[28]The variable VERTICAL_IO refers to backward spillovers and we omit from the analysis the forward spillover measure because it was never statistically significant. See Appendix A.4 for further results regarding the vertical I-O measures.

Table 2: Horizontal and Vertical Technology Spillovers Conditional on Vertical I-O Linkages

<div align="center">DEPENDENT VARIABLE: log FIRM REVENUE TFP
SAMPLE: DOMESTIC FIRMS</div>

	(1)	(2)	(3)	(4)	(5)
VERTICAL_TEC$_{s4,t-1}$	0.297***		0.267**	0.282**	0.312***
	(0.074)		(0.089)	(0.089)	(0.074)
VERTICAL_IO$_{s4,t-1}$		0.113**	0.042	0.049	
		(0.035)	(0.042)	(0.042)	
HORIZONTAL$_{s4,t-1}$				-0.336***	-0.336***
				(0.080)	(0.080)
HORIZONTAL_TEC$_{s4,t-1}$				0.271***	0.271***
				(0.076)	(0.076)
Observations	322,698	322,698	322,698	322,698	322,698
Firm FE	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓
Sec4-Year FE	✓	✓	✓	✓	✓
Cluster	cs4	cs4	cs4	cs4	cs4
s.d.(VERTICAL_TEC)	0.15	0.15	0.15	0.15	0.15
s.d.(VERTICAL_IO)	0.07	0.07	0.07	0.07	0.07
s.d.(HORIZONTAL)	0.16	0.16	0.16	0.16	0.16
s.d.(HORIZONTAL_TEC)	0.07	0.07	0.07	0.07	0.07

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. The dependent variable is the logarithm of revenue firm-level productivity at time t, (log TFPR$_{i,t}$). See section 2.3 for the details of the construction of the variables. The sample comprises domestic firms; i.e., firms that have no foreign participation over the years of analysis. All right-hand-side variables are lagged one period. Standard errors are clustered at the country-four-digit-sector-level. Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

lower if FDI is concentrated in "technologically isolated" sectors. This may well explain why the estimated productivity spillovers from FDI differ across countries.

4 Robustness and Quantitative Implications

4.1 Technology or Pricing?

So far, we have studied the effect of spillovers on revenue total factor productivity which combines the influence of technology and demand factors. Without further analysis, it is impossible to know whether the effects found in Table 2 reflect changes in domestic firms' technology or prices. To see

this, let us express revenue total factor productivity as:

$$\text{TFPR}_{it} \equiv P_{it}\text{TFPQ}_{it} = \mu_{it} \times \text{MC}_{it} \times \text{TFPQ}_{it}\,, \tag{10}$$

where P_{it} refers to the firm's output price, and TFPQ_{it} is physical productivity. If firm-specific prices were available to deflate nominal output, physical productivity (TFPQ_{it}) could be isolated. However, in the absence of firm prices, the use of sectoral price indexes to deflate introduces biases if there are within-industry deviations between firm and sectoral prices. One can write the firm price as the product of marginal cost and a markup μ_{it}, as is done in equation (10). Re-arranging and writing the variables in growth rates (denoted by Δ), gives the following relationship between efficiency gains and changes in TFPR, markups, and marginal costs: $\Delta\text{TFPR}_{it} - \Delta\mu_{it} = \Delta\text{TFPQ}_{it} + \Delta\text{MC}_{it}$. Reordering, we obtain a formula for the change in physical productivity:

$$\Delta\text{TFPQ}_{it} \;=\; \Delta\text{TFPR}_{it} - \Delta\mu_{it} - \Delta\text{MC}_{it}\,. \tag{11}$$

If one models firm-level marginal cost and markups, one can derive an estimate of changes in productivity. In order to provide an estimate of markups, we follow the influential approach of De Loecker and Warzynski (2012). Hall (1986) notes that under imperfect competition, input growth is associated with disproportional output growth (as measured by the relevant markup). Based on this insight, De Loecker and Warzynski (2012) re-arrange the first order condition of the firm cost minimization problem with respect to the flexible input $\mathcal{J}$ to derive the firm markup (defined as price over marginal cost) according to this expression:

$$\mu_{it} \equiv \frac{P_{it}}{MC_{it}} = \underbrace{\frac{\partial\mathcal{F}_{it}(\cdot)}{\partial\mathcal{J}_{it}}\frac{\mathcal{J}_{it}}{\mathcal{F}_{it}(\cdot)}}_{OutputElasticity}\bigg/\;\underbrace{\frac{P_{it}^{\mathcal{J}_{it}}\mathcal{J}_{it}}{P_{it}\text{y}_{it}}}_{ExpenditureShare}\quad, \tag{12}$$

where P_{it} is the output price, MC_{it} is marginal cost, $\mathcal{F}_{it}(\cdot)$ is the production function, $\mathcal{J}_{it}$ is inputs, and $P_{it}\text{y}_{it}$ is nominal value added.

Following De Loecker and Warzynski (2012), we consider labor as a flexible input $\mathcal{J}$. Labor's expenditure share is straightforward to compute from the data as the ratio of the labor cost to value added. The output elasticity with respect to labor is given by the elasticity obtained from the estimation of the production function β_ℓ.[29] We find empirical estimates of the median markup close to 1.5 on average, which is similar to the estimates in the literature.[30]

---

[29]We estimated the firm-level production function separately for two-digit industry and country so β_ℓ is constant across firms within two-digit industries in each country.

[30]Notice we follow De Loecker and Warzynski (2012) strategy to recover markups as the ratio of the technical coefficient to the input share but for consistency the total factor productivity measure used throughout the text is the one obtained from Wooldridge (2009) estimation procedure and therefore, the technical coefficients that go into the estimation of the firm markups as well.

Following Marin and Voigtlander (2014), who express marginal cost as a function of physical productivity and input prices, we provide a more indirect estimate of the effect of FDI on TFPQ, which we label the effect on "implied TFPQ" by applying equation (11). We have data for the total material and wage cost, and normalizing this with revenue provides us with a measure of average cost. We regress this measure on our regressors. While the change over time in average cost over revenue is not literally the same as marginal cost, since firm-level constants are averaged out via firm fixed effects, we take the estimated coefficients as estimates of the effect of the regressors on marginal cost. Armed with estimates of the effect of spillovers on revenue TFP, markups, and marginal cost, we can provide estimates of (implied) TFPQ.

Table 3: Revenue TFP and Markups

SAMPLE: DOMESTIC FIRMS

Dependent Variable:	$\log \text{TFPR}$ (1)	$\log \mu$ (2)	$\log MC$ (3)	Implied $\log \text{TFPQ}$ (4)
VERTICAL_TEC$_{s4,t-1}$	0.312*** (0.074)	0.267*** (0.074)	-0.267*** (0.045)	0.312*** (0.114)
HORIZONTAL$_{s4,t-1}$	-0.336*** (0.080)	-0.192** (0.064)	0.144*** (0.032)	-0.288*** (0.107)
HORIZONTAL_TEC$_{s4,t-1}$	0.271*** (0.076)	0.076 (0.076)	-0.104** (0.042)	0.299*** (0.115)
Observations	322,698	322,698	322,698	
Firm FE	✓	✓	✓	
Country-Year FE	✓	✓	✓	
Sec4-Year FE	✓	✓	✓	
Cluster	cs4	cs4	cs4	
s.d.(VERTICAL_TEC)	0.15	0.15	0.15	
s.d.(HORIZONTAL)	0.16	0.16	0.16	
s.d.(HORIZONTAL_TEC)	0.07	0.07	0.07	

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. In column (1), the dependent variable is the logarithm of revenue firm-level productivity at time t, ($\log \text{TFPR}_{i,t}$). In column (2), the dependent variable is the logarithm of firm markup ($\log \mu_{i,t}$). In column (3), the dependent variable is the logarithm of marginal cost MC, computed as the sum of the cost of employment and the expenditure on materials over total output. See section 2.3 for the details of the construction of the variables. The sample comprises domestic firms; i.e., firms that have no foreign participation over the years of analysis. All right-hand-side variables are lagged one period. Standard errors are clustered at the country-four-digit-sector-level in column (1)-(3). Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. The numbers in column (4) are obtained by subtracting the numbers in the second and third columns from the number in the same row in column (1). The standard errors in column (4) are calculated assuming independence of the estimates in columns (1)-(3). *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

We consider the effect of each measure. Column (1) in Table 3 repeats our main specification (showing standardized coefficients) with the dependent variable log TFPR. A one standard deviation increase in VERTICAL_TEC, capturing vertical spillovers from outher four-digit sectors, increases domestic firms' revenue productivity by 0.312 percent. Column (2) uses the log of firm markup as the dependent variable and shows that a unit increase in VERTICAL_TEC leads to a 0.267 percent increase in domestic firms' markup. Column (3) uses the average cost of domestic firms as the dependent variable and shows that a unit increase in VERTICAL_TEC leads to a drop in the marginal average cost of 0.267 percent.[31] And, adding the numbers in columns (1)-(3), we obtain, in column (4), the effect on physical TFP. The effect of a one standard deviation change in technology-weighted knowledge spillovers is 0.312 percent, which for this variable happens to be exactly the value of the impact on revenue TFP.

Consider foreign presence in the same four-digit sector as captured by HORIZONTAL. The effect on TFPR of a one standard deviation increase in HORIZONTAL is –0.336 percent. Columns (2), (3), and (4) show that this effect can be decomposed into a –0.192 percent change in markups, a 0.144 percent increase in marginal cost, and a 0.288 percent decrease in TFPQ. Because the technology-weighted four-digit measure is included, the interpretation is that of an increase in competition with no knowledge spillovers, while the impact of HORIZONTAL_TEC is the effect of knowledge spillovers in the four-digit industry when product competition is controlled for.

Consider knowledge spillovers from within the four-digit sector. The predicted effect of a one-standard-deviation change in HORIZONTAL_TEC is a 0.271 percent increase in revenue productivity, and columns (2), (3), and (4) show that this effect can be decomposed into a 0.076 percent increase in markups, a –0.104 percent change in marginal cost, and a 0.299 percent increase in TFPQ. It is reasonable that increased competition decreases markups and increases marginal costs due to a lower scale of production. Knowledge spillovers can be expected to lower costs, and we see negative relations between (approximate) marginal cost and the technological-spillover measures. The lower cost then allows firms to increase margins.

To sum up, increases in revenue productivity, induced by higher foreign ownership in technologically related sectors outside the same four-digit sector, are mainly driven by changes in physical productivity. The presence of technologically related foreign firms in the same four-digit sector has a similar effect on revenue productivity, but a larger fraction of this is driven by changes in markups and marginal costs. Competition lowers revenue productivity, with about two-thirds of the effect coming from physical productivity.

---

[31]That VERTICAL_TEC has exactly the opposite effect on markup and average cost to the third decimal point is purely coincidental.

4.2 Extra Robustness

Table 4 examines if our results are robust to the use of OLS instead of GLS, to reasonable changes in the way the sample is selected, to the use of dummies at the two-digit sector×year level, and to the inclusion of a direct measure of competition.

Table 4: Robustness

DEPENDENT VARIABLE: log FIRM REVENUE TFP
SAMPLE: DOMESTIC FIRMS

	(1) Benchmark	(2) OLS	(3) Balanced	(4) Empl10	(5) Sec2-year	(6) HHN
VERTICAL_TEC$_{s4,t-1}$	0.312***	0.342**	0.407**	0.442***	0.461***	0.282***
	(0.074)	(0.163)	(0.136)	(0.091)	(0.089)	(0.074)
HORIZONTAL$_{s4,t-1}$	-0.336***	-0.336*	-0.494***	-0.527***	-0.528***	-0.288***
	(0.080)	(0.176)	(0.132)	(0.099)	(0.080)	(0.080)
HORIZONTAL_TEC$_{s4,t-1}$	0.271***	0.09	0.548***	0.362***	0.688***	0.341***
	(0.076)	(0.195)	(0.153)	(0.094)	(0.083)	(0.076)
HERFIN$_{s4,t-1}$						-0.582***
						(0.116)
Observations	322,698	322,698	101,424	221,907	322,705	322,698
Firm FE	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓	✓
Sec4-Year FE	✓	✓	✓	✓		✓
Sec2-Year FE					✓	
Cluster	cs4	cs4	cs4	cs4	cs4	cs4
s.d.(HORIZONTAL)	0.16	0.16	0.16	0.16	0.16	0.16
s.d.(HORIZONTAL_TEC)	0.07	0.06	0.07	0.07	0.07	0.07
s.d.(VERTICAL_TEC)	0.15	0.15	0.15	0.15	0.15	0.15
s.d.(HERFIN)						1.16

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. The dependent variable is the logarithm of revenue firm-level productivity at time t, (log TFPR$_{i,t}$). See section 2.3 for details of the construction of the variables. The sample comprises domestic firms; i.e., firms that have no foreign participation over the years of analysis. All right-hand-side variables are lagged one period. Column (1) reports benchmark results; column (2) reports results in a stable sample of firms that we observe during the full time period of analysis; column (3) reports the OLS results; column (4) reports results for the subsample of firms that always report employment greater than ten employees; column (5) reports results including two-digit-sector (rather than four-digit-sector) and year fixed effects; column (6) reports results controlling for country-four-digit-sector concentration. Standard errors are clustered at the country-four-digit-sector-level. Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

The first column displays our preferred regression that uses GLS. The literature has typically presented OLS estimates, which do not downweigh firms with volatile productivity. We demonstrate in Column (2) our results are broadly similar whether we use OLS or GLS. The horizontal technology closeness-weighted spillover measure is insignificant using GLS, while the effects of competition and vertical technology closeness-weighted spillovers are estimated with coefficients that are very similar

to those of the GLS estimation.

The results of column (3), estimated using a balanced sample of firms, are similar to those obtained using the full sample, although the estimated coefficients are larger for the balanced sample of firms (the standardized coefficients can be compared to those of Table 2 because the standard errors of the variables in the balanced sample are similar to those of the larger sample used in that table). This indicates that the results we have found so far are not driven by entry and exit of firms.

One may worry about the somewhat artificial sample selection of keeping firms in the sample only in the years where they have more than 10 employees. If we keep only firms whose employment never dips below 10 employees, we get the results presented in column (4). The estimated coefficients for this sample are larger than the benchmark coefficients, but the qualitative conclusions remain unchanged.

Using four-digit-year fixed effects absorbs a lot of degrees of freedom, and we examine if the results are sensitive to that. If we use two-digit-year fixed effects, rather than the four-digit-year fixed effects used so far—see column (5)—the impacts of all measures are larger, which is not surprising given that less variation is absorbed by dummies; however, the overall patterns are similar to those found earlier. We prefer the estimates with four-digit-year fixed effects because they are less likely to be impacted by reverse causality, even if the four-digit-year fixed effects may absorb some of the impact of FDI on productivity.

We interpret the effect of HORIZONTAL as a result of increased competition, and we examine if the coefficients are sensitive to including a direct measure of competition. Including a measure of concentration, see column (6), affects the estimated effects little, with no coefficients moving out of the standard 95 percent confidence intervals from the preferred regression, although the impact of the horizontal variables become smaller consistent with the interpretation. Because concentration is likely endogenous to FDI, we prefer the specification without the concentration measure.

Finally, the results may conceivably depend on the exact way that we calculate TFP, but if we run our regressions with labor productivity as the dependent variable and we obtain similar signs and t-values, which indicate that the details of how TFP is calculated is not of importance for our results. Because the coefficients in those regressions are not comparable to those obtained using TFP, we do not tabulate them.

Our results are robust to reasonable modifications of samples and specifications, and the results of the robustness regressions indicate that, if anything, our reported magnitudes are conservative.

4.3 Short Run and Long Run Effects

It is possible that spillover effects take time to materialize, in which case the results will be more economically significant at longer horizons—our previous level-regressions with firm-fixed effects are

equivalent to estimating growth effects over the full sample. Table 5 repeats our main specification estimated in one-, two-, and four-year (non-overlapping) differences.[32]

Table 5: Lagged Spillovers

	(1)	(2)	(3)	(4)
Dependent Variable:	$\log \text{TFPR}$	$\Delta^{j=1} \log \text{TFPR}$	$\Delta^{j=2} \log \text{TFPR}$	$\Delta^{j=4} \log \text{TFPR}$
$\Delta^j \text{VERTICAL\_TEC}_{s4}$	0.021***	0.001	0.032***	0.053***
	(0.005)	(0.005)	(0.004)	(0.004)
$\Delta^j \text{HORIZONTAL}_{s4}$	-0.021***	-0.023***	-0.014**	-0.017***
	(0.005)	(0.006)	(0.006)	(0.005)
$\Delta^j \text{HORIZONTAL\_TEC}_{s4}$	0.039***	0.010	0.039**	0.054***
	(0.011)	(0.015)	(0.015)	(0.012)
Observations	322,698	374,592	172,806	72,752
Firm FE	yes	no	no	no
Country-Year FE	✓	✓	✓	✓
Sec4-Year FE	✓	✓	✓	✓
Cluster	cs4	cs4	cs4	cs4

Notes: The sample comprises domestic firms; i.e., firms that have no foreign participation over the years of analysis. The dependent variable is the non-overlapping year difference in the logarithm of revenue firm-level productivity at time t ($\Delta^j \log, \text{TFPR}_{i,t}$) where j indicates, first, second, and fourth differences. See section 2.3 for the details of the construction of the variables. Standard errors are clustered at the country four-digit-sector-level. Results are obtained by GLS estimation using as weights the square root of the firm mean squared predicted residuals from an initial OLS regression. Column (1) displays (non-standardized) results from our main regression, column (2) is estimated in first differences, column (3) is estimated in second non-overlapping differences, and column (4) is estimated in fourth non-overlapping differences. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

In Table 5, the coefficients are not standardized (nor multiplied by 100), because we want to compare across the different frequencies (columns), rather than across different variables (rows), and this would be hard to do with standardized coefficients. Further, the coefficients are easier to compare with existing results in the literature.[33]

---

[32]The difference-specifications do not include firm fixed effects which "washes out" with differencing. We use non-overlapping data to avoid the mechanically induced autocorrelation that would result from using overlapping data—an autocorrelation that typically cannot be well corrected for. Because we use non-overlapping differences, the number of observations decline with the length of differencing.

[33]We regress over long differences rather than running regressions with many lags. While the later might allow more precise estimates about the timing of effects, for higher lags this benefit is to be weighed against the many lag coefficients themselves being imprecisely estimated. On net, we prefer the long-difference specification, which delivers a compact set of results that are simple to communicate.

In the first column, we repeat the preferred specification from Table 2, but displaying the non-standardized coefficients. The results imply that competition spillovers affect productivity instantly while knowledge spillovers occur only gradually. Competition effects, as captured by the variable HORIZONTAL, are significant already at the first year of a change and the effect remains negative and with no-significant changes over two- and four-year differences. The one-year-difference coefficient of –0.023 is very close to the coefficient –0.021 from the level-regression. The point estimates of the competition effects are numerically slightly smaller at higher differencing intervals but the coefficients are not statistically different from the ones obtained using one-year-differenced data.

The estimated impact of knowledge spillovers within the four-digit sector increases with the length of the differencing period and it is significant only after two years, where the coefficient is 0.039 (identical to the coefficient from the regression in levels with fixed effects). After 4 years the coefficient is 0.054, even larger than the level effect.[34] Finally, knowledge spillovers from outside the four-digit sector also become significant after two years and almost double again from two- to four years, with a four-year coefficient of 0.053, which is larger than the estimated level effect of 0.021. These results provide strong intuitive support for our interpretation because knowledge spillovers are likely to be caused by low-frequency phenomena, such as workers bringing skills from foreign to domestic firms or domestic firms studying the business practices of foreign firms before trying to adapt them.

The estimated coefficients from the level regressions are preferred because the long differences use a much smaller sample, but the conclusions are otherwise very robust and it is clear that knowledge spillovers take time to materialize. The larger coefficients at longer differencing intervals are likely an artifact of changes in the sample, with firms in existence for only a few years not entering the four-year-differenced sample—from Table 4, the estimated coefficients are larger for a sample of firms in existence throughout the sample period.

Haskel, Pereira and Slaughter (2007) estimate spillovers from FDI in the UK—a developed country comparable to the countries in our sample—and they use differences at various frequencies, as we do. They find that spillovers take time to materialize and although their results are not directly comparable with ours, we can compare magnitudes.[35] Our VERTICAL_TEC measure increases more or less proportionally with foreign ownership on average, so the non-standardized coefficient to VERTICAL_TEC of 0.053 in column (4) implies that a change in VERTICAL_TEC of 0.1 predicts a change in TFP of 0.53 percent, which can be compared to the effect of 10 percentage point change in foreign ownership found by Haskel, Pereira and Slaughter (2007) for five-year differences of 0.63 percent. Because the specifications differ in several respects, we will not compare more extensively

---

[34]Estimating long differences is not the same as estimating lags, so referring to the effect, when TFP and the regressor are both differenced over two years, as the effect "after" two years involves an obvious abuse of notation. More precisely, the two-year differenced results capture average impacts over two years.

[35]They consider FDI in either the same sector or the same geographical area as domestic firms, and do not separate between competition and knowledge effects, or between closeness in product or technology space.

to their results, but we find it reassuring that our estimates and those of the literature are similar in magnitude.

4.4   Economic Significance and Quantitative Implications

We illustrate how important FDI is in explaining the observed TFP improvements of domestic firms by performing two different exercises. First, we undertake a simple prediction exercise and calculate the variation explained in the actual revenue TFP data by our estimated coefficients. And second, in order to highlight the importance of heterogeneity of sectors in terms of technological closeness, we run a counterfactual experiment where we hypothetically change FDI *only* in technologically connected sectors.

Figure 4 shows the result of the first exercise. On the y-axis, we plot the average annual TFP growth rate at the country-four-digit-sector-level. We average the firm-level TFP data to the four-digit-sector-level by country and calculate the annual growth rate for the country-four-digit-sector TFP for each year. We average these growth rates over time and plot them on the y-axis. On the x-axis, we plot the predicted TFP growth rates—we use the estimates in column (4) of Table 2 (including all fixed effects) to predict average firm-level TFP over time. We average the firm-level predicted TFP effects to the country-four-digit-sector-level for each year, calculate TFP growth rates, and plot the average of these growth rates on the x-axis. Our predicted TFP growth rates explain 43.5 percent of the variation in the actual TFP growth rates.

Figure 4: Actual versus predicted TFP growth by four-digit sector 2000 and 2008.
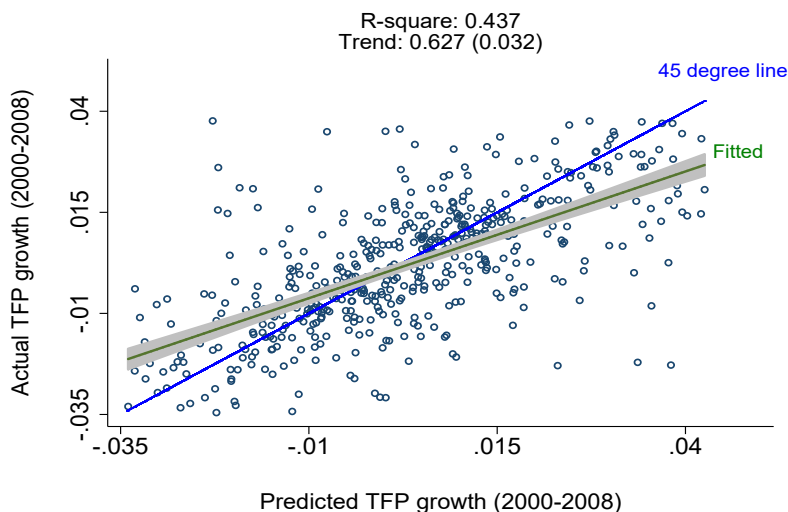
Notes: The figure plots estimated actual TFP growth, averaged by four-digit sector, against firm-level predicted TFP growth, averaged by four-digit sector. In order to focus on the central mass of the distribution, the four-digit-level TFP variables are truncated at the 5 and 95 percentiles in this figure.

Next, we quantify the effects on TFP of hypothetical changes in foreign ownership of a certain

magnitude using only the estimated coefficients on the FDI/spillover variables in column (4) of Table 2. We consider the effect of a similar amount of FDI in all sectors and the effect of certain patterns in the amount of FDI across sectors.

Consider the effect of an across-the-board change in foreign ownership in sector $s4$ of magnitude $\Delta \mathrm{fo}_{s4}$, which should be considered evenly distributed across firms in each sector. The predicted change in TFPR as a result of change in each spillover variable from column (4) of Table 2 can be written as:

$$\Delta \log \widehat{(\mathrm{TFPR}_{s4})} = \hat{\beta}_1 \Delta \mathrm{VERTICAL\_TEC}_{s4} + \hat{\beta}_2 \Delta \mathrm{HORIZONTAL}_{s4} \tag{13}$$
$$+ \hat{\beta}_3 \Delta \mathrm{HORIZONTAL\_TEC}_{s4} \ .$$

Because the right-hand-side spillover variables in equation (13) are sector-level variables, we obtain the predicted average sectoral TFPR increase. We can transfer a change in firm-level foreign ownership, which is what we measure in our data, to a change in the country-four-digit-sector spillover variables as follows:

$$\Delta \mathrm{VERTICAL\_TEC}_{s4} = \sum_{\tilde{s}4, \tilde{s}4 \in s2(s4), \tilde{s}4 \neq s4} \mathrm{WTECH}_{s4,\tilde{s}4} \times \Delta \mathrm{HORIZONTAL}_{\tilde{s}4} \tag{14}$$
$$= \sum \mathrm{WTECH}_{s4,\tilde{s}4} \times \Delta \mathrm{fo}_{\tilde{s}4}$$
$$= \Delta \mathrm{fo}_{\tilde{s}4} \ .$$

Because the weights are constructed to sum to unity, the impact of a change in foreign ownership if $\Delta \mathrm{fo}_{s4}$ for every firm in an $s4$ sector is

$$\Delta \mathrm{HORIZONTAL}_{s4} = \frac{\sum_{i \in s4} \Delta \mathrm{fo}_{s4} \times \mathrm{go}_i}{\sum_{i \in s4} \mathrm{go}_i} = \Delta \mathrm{fo}_{s4} \ . \tag{15}$$

We have

$$\Delta \mathrm{HORIZONTAL\_TEC}_{s4} = \mathrm{WTECH}_{s4,s4} \times \Delta \mathrm{fo}_{s4} \ . \tag{16}$$

For an across-the-board change of $\Delta \mathrm{fo}_i = \Delta \mathrm{fo}_{s4} = \Delta \mathrm{fo}$, the total effect is

$$\Delta \log \widehat{(\mathrm{TFPR}_{s4})} = \left(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 \, \mathrm{WTECH}_{s4,s4} \right) \Delta \mathrm{fo} \ . \tag{17}$$

FDI may not take place as an equal across-the-board change in all sectors. As is clear from equations (7) and (8), the impact of FDI will be higher if it is concentrated in sectors where the WTECH coefficients are higher; i.e., in sectors that (on average) are technologically close to other sectors (including the average technological closeness of firms within each sector). We construct a

dummy for such sectors in order to evaluate the effect of FDI being concentrated in these sectors.

For each sector $s4$, we can calculate the "average technology connectedness" (ATC) to other sectors as

$$\text{ATC}_{s4} = \frac{1}{N} \sum_{\tilde{s}4 \in s2} \text{WTECH}_{\tilde{s}4, s4} \,, \tag{18}$$

where the average is taken over the first subscript of WTECH. We construct a dummy, $D_{connected\ s4}$ which is unity for "high connectedness sectors," with a value of ATC above the median (across four-digit sectors) and 0 otherwise. We calculate the effect of focusing FDI in highly connected sectors. Specifically, we calculate the effect of an increase of $2\,\Delta\text{fo}$ in the sectors which have ATC above the median and 0 in the other sectors ($\Delta\text{fo}_i = \Delta\text{fo}_{s4} = D_{connected\ s4} \times 2\,\Delta\text{fo}$). By construction, attributing twice the amount of FDI to the "high connectedness" sectors, which make up 50 percent of the sectors, the total amount in FDI is the same as an increase of Δfo in all sectors . We then calculate the predicted effect as

$$
\begin{aligned}
\Delta \log \widehat{(\text{TFPR}_{s4})} \;=\; & \hat{\beta}_1 \sum_{\tilde{s}4, D_{connected\ \tilde{s}4}=1} \text{WTECH}_{s4, \tilde{s}4} \times (2\,\Delta\text{fo}) \\
& + \; D_{connected\ s4} \times (\hat{\beta}_2 + \hat{\beta}_3 \text{WTECH}_{s4, s4}) \times (2\,\Delta\text{fo}).
\end{aligned}
\tag{19}
$$

Notice that the last two terms only are non-zero for sectors that themselves have above-median ATC, while the first term is non-zero for all sectors. In Table 6, we tabulate all the predicted effects. We first consider the effect of an increase in foreign ownership of 10 percentage points across the board. The first row in Table 6 shows that the average predicted increase in productivity across all four-digit sectors is 0.14 percent. This effect is the sum of three partial effects as given in equation (17).

In the second row, we show the predicted effect on average TFP of a similar aggregate increase in FDI, but concentrated in the 50 percent of sectors that are closest to other sectors in technology space on average. The predicted effect is twice as large in this case. There is a larger negative competition effect of –0.34 (which is an artifact of the more connected sectors being larger), a knowledge-spillover effect from firms in the same sector of 0.27, and a spillover effect from firms outside the four-digit sector of 0.32. The total effect then is a predicted increase of average TFP of 0.26 percent, almost twice the predicted impact of an across-the-board increase in TFP. More strikingly, the impact of concentrating TFP in the least connected sectors is virtually nil with a spillover effect of 0.04 percent. This results from a negative competition effect of –0.09 percent, small knowledge spillovers of 0.02 percent from within the four-digit sector, and knowledge spillovers of 0.10 percent from outside the four-digit sector. These results shed light on why researchers have had a difficult time identifying horizontal spillovers. Knowledge spillovers will be negative if domestic

Table 6: Predicted Effects of Increases in FDI

Sample: Domestic Firms

Δ fo	Targeted Sectors	(1)=(2)+(3)+(4) $\Delta \log \text{TFPR}$	(2) $\Delta_{\text{HORIZONTAL}}$	(3) $\Delta_{\text{HORIZONTAL\_TEC}}$	(4) $\Delta_{\text{VERTICAL\_TEC}}$
10%	Across the Board Increase	0.14	-0.21	0.14	0.21
20%	Highly Connected Sectors	0.26	-0.34	0.27	0.32
20%	Less Connected Sectors	0.04	-0.09	0.02	0.10

Notes: Row (1) displays the predicted average increase in TFPR in manufacturing as a result of a 10 percentage point increase in FDI, equally distributed across sectors. Row (2) displays the predicted average increase in TFPR in manufacturing as a result of a 10 percentage point increase in FDI, distributed as a 20 percentage point increase in FDI in half of sectors, chosen as those that are closer to other sectors on average and 0 percentage points in the more isolated sectors. "Closeness to other sectors on average" is defined according to equation (18). Row (3) shows the predicted impact of the same amount of FDI, but focused in the sectors that are below the median in terms of being close to others in technology space on average. Column (1) shows the predicted average effect in TFPR in manufacturing. Columns (2), (3), and (4) show the contribution of the associated corresponding change in HORIZONTAL, HORIZONTAL_TEC, and VERTICAL_TEC to the predicted average effect in TFPR in manufacturing.

firms operate in sectors that are not technologically close to the sectors that MNCs enter.

5 Endogeneity of FDI

So far we have controlled for global sector-level productivity shocks with sector-year fixed effects. This rules out the concern that these global shocks drive both the entry decision of foreigners and domestic firms' productivity. Nonetheless, it is still possible that certain sectors are booming *more* in certain countries than in others and we therefore need to address endogeneity concerns at the country-sector-year level.

First, we instrument our horizontal spillover variables following the approach of Bartik (1991). For each firm i in sector $s4$ with foreign ownership in any year, we use the value of foreign output (which may be zero) and total output in the first year the firm is observed and let these variables grow over time in line with output in the same sector in all other countries in the sample:

$$\text{HORIZONTAL}^{IV}{}_{s4,t} = \frac{\sum_{i \in s4} \text{fo}_{i,o} \times (\text{go}_{i,o} \times \widehat{\text{cumm\_gr}}_{i,t})}{\sum_{i \in s4} \text{go}_{i,o} \times \widehat{\text{cumm\_gr}}_{i,t}}, \tag{20}$$

where $\text{go}_{i,o}$ is total output the first year we observe the firm, $\text{fo}_{i,o} \times \text{go}_{i,o}$ is the initial foreign part of output, and $\widehat{\text{cumm\_gr}}_{i,t}$ is the predicted growth of output for firm i from period 0 to period t.[36]

---

[36]The subscript 0 does not refer to any given year, but is short-hand for the first period each firm i is observed. The formula for predicted growth similar involve a slight abuse of notation, by not marking all time indices as firm

Firm-level predicted output growth is calculated as:

$$\widehat{\text{cumm\_gr}}_{i,t} = \Pi_{s=1}^{t}(1 + \widehat{\text{gr}}_{s2(i),t}) - 1 \tag{21}$$

where $\widehat{\text{gr}}_{s2(i),t}$ is the growth-rate of the two-digit sector of firm i in all other countries in the sample except for the country of firm i. In order to retain information about the industry structure in country c, we weigh the growth of the four-digit sectors with the initial industry shares found in country of firm i, as:

$$\widehat{\text{gr}}_{s2(i)t} = \sum_{s4 \in s2(i), c' \neq c} \left(\frac{\text{GO}_{s4,c,0}}{\text{GO}_{s2(i),c,0}} \right) \times \text{gr}_{s4,c',t} \ . \tag{22}$$

As instrument for HORIZONTAL_TEC, we use $\text{HORIZONTAL\_TEC}^{IV} = \text{WTECH}^{IV}_{s4,s4,t} \times \text{HORIZONTAL}^{IV}$, where $\text{WTECH}^{IV}_{s4,s4,t}$ is instrumented in the same fashion using output growth in other countries.[37]

Table 7: Instrumental Variables—First Stage

DEPENDENT VARIABLES: HORIZONTAL SPILLOVER MEASURES
SAMPLE: DOMESTIC FIRMS

	(1)	(2)
Dependent Variable	HORIZONTAL$_{s4,t-1}$	HORIZONTAL_TEC$_{s4,t-1}$
HORIZONTAL$^{IV}_{s4,t-1}$	4.026***	0.177*
	(0.279)	(0.101)
HORIZONTAL_TEC$^{IV}_{s4,t-1}$	4.003***	3.558***
	(0.35)	(0.202)
VERTICAL_TEC$_{s4,t-1}$	0.148	-0.044
	(0.341)	(0.193)
Firm FE	✓	✓
Country-Year FE	✓	✓
Sec4-Year FE	✓	✓
Cluster	cs4	cs4
F-stat	186 ***	126 ***
s.d.(HORIZONTAL^{IV})	0.24	0.24
s.d.(HORIZONTAL_TEC^{IV})	0.13	0.13

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

First stage regressions in Table 7 show that the instruments are highly significant. The second stage regressions are reported in Table 8; where Column (1) presents our main OLS results for comparison. Column (2) shows the second stage results after instrumenting the variables HORIZONTAL

---

specific, in order to simplify the equations.

[37]The exact formula is $\text{WTECH}^{IV}_{s4,s4,t} = \frac{\text{SPILL\_RD}_{s4,s4} \times (\sum_{i \in s4} \text{go}_{i,o} \times \widehat{\text{cumm\_gr}}_{i,t})}{\sum_{\tilde{s}4 \in s2(s4)} \text{SPILL\_RD}_{s4,\tilde{s}4} \times (\sum_{i \in \tilde{s}4} \text{go}_{i,o} \times \widehat{\text{cumm\_gr}}_{i,t})}$.

and HORIZONTAL_TEC. The estimated coefficients tell the same story as the un-instrumented horizontal spillover variables, although the coefficients are much larger for the instrumented variables. This pattern is consistent with the instrument correlating better with the potentially endogenous horizontal variables in some regions or sectors than others (a "local treatment effect"). We conclude that the qualitative impact of the horizontal spillover-variables is robust, while the magnitudes of the effects may be more tentative. The estimated coefficient on the variable VERTICAL_TEC is almost unchanged when comparing the OLS and IV specifications. Because the vertical measures are nearly orthogonal to the horizontal measures, the coefficient stays very close across specifications.

Column (3) undertakes an exercise that deals with the possibility that VERTICAL_TEC$_{s4,t}$ may be endogenous, even if FDI firms and domestic firms are in different sectors. Assume for simplicity of exposition that a two-digit sector consists of three four-digit sectors, labelled as $s4_1$, $s4_2$, and $s4_3$. Let us focus on domestic firms in sector $s4_1$ and assume that foreign firms enter sectors $s4_2$ and $s4_3$. Assume that foreign firms of $s4_2$ enter as a reaction to productivity advances by domestic firms in sector $s4_1$, because globally sector $s4_2$ happens to learn a lot from $s4_1$ and it is easier to benefit from these spillovers in geographical proximity. Further assume that sector $s4_3$ has never learned anything from $s4_1$, so that the FDI in that sector must be for reasons other than events in sector $s4_1$. If this is the case, foreigners entering $s4_2$ should cite patents of the sector $s4_1$ more often than foreigners entering to sector $s4_3$. This would create a stronger relation between domestic productivity in sector $s4_1$ and FDI in sector $s4_2$ due to reverse causality based on absorbing technology from sector $s4_1$ more intensely. From the perspective of domestic sector $s4_1$, FDI into sector $s4_2$ is endogenous while FDI into sector $s4_3$ is exogenous. We, therefore, estimate two separate coefficients and test whether the relation between domestic productivity and FDI is stronger for foreign entry sectors that cite the sector of the domestic firms more heavily or not.[38]

For each sector $s4$, we split the other sectors into two groups, based on how often they cite patents belonging to sector $s4$. Consider the vector cite$_{s4,\tilde{s}4}$ giving us the number of citations made by firms in sector $\tilde{s}4 \in s2(s4)$ to patents originating in sector $s4$, for $\tilde{s}4 \in s2(s4), \tilde{s}4 \neq s4$. In other words, this vector should tell us how much each $\tilde{s}4$ cites our baseline sector $s4$. We define the dummy variable D_HC$_{s4,\tilde{s}4}$ taking value 1 for sectors with above median citations made to sector $s4$ and D_LC $= 1 -$ D_HC taking value 1 for sectors with below median citations made to sector $s4$, i.e. sectors that are less likely to receive knowledge-spillovers from sector $s4$. This latter dummy variable is the one that we will interpret causally as the reverse causality concern is considerably mitigated.

---

[38]See Appendix A.3 for a full description of the citation data as well as the construction of the variables.

Table 8: Instrumental Variables—Second Stage (Asymmetric Patent Citation)

DEPENDENT VARIABLE: log FIRM REVENUE TFP
SAMPLE: DOMESTIC FIRMS

	(1)	(2)	(3)
Specification for Horiz. Spillovers	GLS	IV	IV
HORIZONTAL$_{s4,t-1}$	-0.336***	-1.617***	-1.617***
	(0.080)	(0.240)	(0.240)
HORIZONTAL_TEC$_{s4,t-1}$	0.271***	1.224***	1.217***
	(0.076)	(0.202)	(0.202)
VERTICAL_TEC$_{s4,t-1}$	0.312***	0.342***	
	(0.074)	(0.074)	
VERTICAL_TEC$^{HC}_{s4,t-1}$			0.327***
			(0.098)
VERTICAL_TEC$^{LC}_{s4,t-1}$			0.318**
			(0.098)
Obs.	322698	322698	321474
R^2	0.99631	0.99631	0.99629
Firm FE	✓	✓	✓
Country-Year FE	✓	✓	✓
Sec4-Year FE	✓	✓	✓
Cluster	cs4	cs4	cs4
s.d.(HORIZONTAL)	0.16	0.16	0.16
s.d.(HORIZONTAL_TEC)	0.07	0.07	0.07
s.d.(VERTICAL_TEC)	0.15	0.15	0.15
P-value F-test			0.58

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. The dependent variable is the logarithm of revenue firm-level productivity at time t, (log TFPR$_{i,t}$). The regressions are estimated by 2SLS where the variables HORIZONTAL$_{s4,t-1}$ and HORIZONTAL_TEC$_{s4,t-1}$ are instrumented by, correspondingly, HORIZONTAL$^{\text{Bartik IV}}_{s4,t-1}$ and HORIZONTAL_TEC$^{\text{Bartik IV}}_{s4,t-1}$. The construction of the latter variables is explained in Section 5. The first stage for column (2) is reported in Table 7; the first stage for column (3) is similar. All right-hand-side variables are lagged one period. Standard errors are clustered at the country-four-digit-sector level. Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance. "P-value F-test" reports the p-value of the test of the equality of the coefficients of VERTICAL_TEC$^{HC}_{s4,t-1}$ and VERTICAL_TEC$^{LC}_{s4,t-1}$ under the null.

We then define

$$\text{VERTICAL\_TEC}^{HC}_{s4,t} = \sum_{\substack{\tilde{s}4 \in s2(s4) \\ \tilde{s}4 \neq s4}} \text{WTECH}_{s4,\tilde{s}4,t} \times \text{HORIZONTAL}_{\tilde{s}4,t} \times \text{D\_HC}_{s4,\tilde{s}4} , \quad (23)$$

and

$$\text{VERTICAL\_TEC}^{LC}_{s4,t} = \sum_{\substack{\tilde{s}4 \in s2(s4) \\ \tilde{s}4 \neq s4}} \text{WTECH}_{s4,\tilde{s}4,t} \times \text{HORIZONTAL}_{\tilde{s}4,t} \times \text{D\_LC}_{s4,\tilde{s}4} , \quad (24)$$

which satisfies $\text{VERTICAL\_TEC}_{s4,t} = \text{VERTICAL\_TEC}_{s4,t}^{HC} + \text{VERTICAL\_TEC}_{s4,t}^{LC}$

If the estimated coefficient to $\text{VERTICAL\_TEC}_{s4,t}^{HC}$ was larger than the one to $\text{VERTICAL\_TEC}_{s4,t}^{LC}$, this could reflect reverse causality. No significant difference in the coefficients would make it unlikely that causality goes from productivity by domestic firms in sector $s4$ leading to FDI in other sectors that are technologically close. Column (3) of Table 8 shows that there is no significant difference in these estimated coefficients to $\text{VERTICAL\_TEC}_{s4,t}^{LC}$ and $\text{VERTICAL\_TEC}_{s4,t}^{HC}$, which makes us be less concerned about reverse causality.

6   A Primer on the Channel: Inventor Mobility

In this section, we explore one potential channel for spillovers highlighted in the theoretical literature; namely, that of labor mobility (see Fosfuri, Motta and Ronde (2001)). In particular, if one believes that knowledge is embedded in people, then a channel by which firms learn from each other is through employee mobility.[39]

In Table 9, we use a different measure of vertical spillovers, where the technological distance (again using U.S. data) is based on inventor mobility across sectors. Instead of the patent-based measure used in previous tables, now we rather use the fraction of inventors $\text{tech}_{s4,\tilde{s}4}^{INV}$ in sector $s4$ that previously worked for firms in sector $\tilde{s}4$. As the $\text{tech}_{s4,\tilde{s}4}^{INV}$ measure is not symmetric, we will use this feature for identification as we did with the patent citation matrix. The imputation of inventor-flows between sectors from inventor-flows between technology classes is described in detail Appendix A.4.

We proceed using tech^{INV} as the new weight and calculate

$$\text{SPILL\_RD}_{s4,\tilde{s}4}^{INV} = \sum_{i \in s4} \sum_{j \in \tilde{s}4} \text{tech}_{s4,\tilde{s}4}^{INV} \times \left(\frac{\text{rd}_i + \text{rd}_j}{\sum_{k \in s4} \sum_{l \in \tilde{s}4} (\text{rd}_k + \text{rd}_l)} \right), \tag{25}$$

which is equivalent to equation (5), except for using an inventor-mobility weight instead of a patent similarity weight. We further calculate

$$\text{VERTICAL\_TEC}_{s4,t}^{INV} = \sum_{\substack{\tilde{s}4 \in s2(s4) \\ \tilde{s}4 \neq s4}} \text{SPILL\_RD}_{s4,\tilde{s}4,t}^{INV} \times \text{HORIZONTAL}_{\tilde{s}4,t}, \tag{26}$$

The GLS-results obtained using this measure are presented in column (1) of Table 9. Results are very similar to those obtained in column (5) of Table 2 indicating that our findings are not an artifact of the exact way the technology weights are constructed. Domestic firms in sector $s4$ are

---

[39]Other studies like, for example Poole (2013), who use data from Brazil, show that higher-skilled former multinational workers are better able to transfer information.

more likely to benefit from FDI that is concentrated in sectors $\tilde{s}4$ when the inventor matrix shows that there is higher inventor mobility from sector $\tilde{s}4$ to $s4$. In addition, the inventor mobility-based measure does not rely on output weights therefore implying that our main results are not caused by some particular patterns in those weights. Further, results indicate an important role for inventor mobility for knowledge flows across sectors.

These results might be driven by endogeneity if foreign firms enter particular sectors to hire inventors from other sectors. To deal with this concern and following a similar intuition to the one used in the previous section, we divide sectors into high inventor mobility (HM) and low inventor mobility (LM) sectors. If inventors in a given domestic sector do not tend to move to the foreign entry sector (i.e. the LM case), then it is less likely that foreigners do FDI with the hope of stealing inventors from the domestic sector. Consequently, from the perspective of the domestic sector, the presence of FDI in an LM sector is less likely to be driven by reverse causality. We use the transition matrix of inventor mobility across technology classes previously described. From column (2) of Table 9, we see that there is no significant difference between the estimated coefficients of HM and LM sectors, supporting our causal interpretation. Furthermore, it also highlights inventor mobility as one channel of technology transfer captured with our patent based measure.

7 Conclusion

We study the impact of FDI on host-country firms' productivity. We show that the positive FDI spillovers only occur when foreign and domestic firms use similar technologies. When they use dis-similar technologies, the effect of FDI on host country firms may be negative. Our estimates show that productivity gains from FDI doubles, when FDI is concentrated in sectors where firms use similar technologies compared to the case where the same amount of FDI is spread evenly across all sectors.

We account for potential endogeneity of FDI to the productivity of domestic firms in the same sector using two strategies. First, we control for sector×year effects and second, we use Bartik-style instrumentation, where we instrument sector-country growth with productivity growth in the same sector of neighboring countries. Regarding potential endogeneity of FDI to the productivity of domestic firms in other sectors, we use data on asymmetric citations and inventor mobility, both of which supports our causal interpretation. We also show that inventor mobility between sectors is a channel of technology transfer between foreign and domestic firms.

Our work is related to work on competition, innovation, and R&D (Aghion, Bloom, Blundell, Griffith and Howitt (2005), Bloom, Schankerman and Van Reenen (2013), Jaffe (1988), Aghion and Howitt (1992)). Acemoglu, Akcigit and Kerr (2016) argue that technological progress is not only a cumulative process of "standing on the shoulders of giants," but also a process where innovation

Table 9: GLS and IV (Asymmetric Inventor Mobility)

DEPENDENT VARIABLE: log FIRM REVENUE TFP
SAMPLE: DOMESTIC FIRMS

	(1)	(2)
HORIZONTAL$_{s4,t-1}$	-0.320***	-0.320***
	(0.080)	(0.080)
HORIZONTAL_TEC$_{s4,t-1}$	0.229**	0.236**
	(0.076)	(0.076)
VERTICAL_TEC$^{INV}_{s4,t-1}$	0.199***	
	(0.046)	
VERTICAL_TEC$^{INV-HM}_{s4,t-1}$		0.268**
		(0.084)
VERTICAL_TEC$^{INV-LM}_{s4,t-1}$		0.147**
		(0.059)
Obs.	322698	322698
Firm FE	✓	✓
Country-Year FE	✓	✓
Sec4-Year FE	✓	✓
Cluster	cs4	cs4
s.d.(HORIZONTAL)	0.16	0.16
s.d.(HORIZONTAL_TEC)	0.07	0.07
P-value F-test		0.40

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. The dependent variable is the logarithm of revenue firm-level productivity at time t, (log TFPR$_{i,t}$). All right-hand-side variables are lagged one period. Standard errors are clustered at the country-four-digit-sector level. Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance; $^{+}$ denotes 15% significance. "P-value F-test" reports the p-value of the test of the equality of the coefficients of VERTICAL_TEC$^{INV-HM}_{s4,t-1}$ and VERTICAL_TEC$^{INV-LM}_{s4,t-1}$ under the null.

in one firm affects firms in technologically close fields. Our findings are in line with the existence of such innovation networks. The unique contribution of our paper is to show that these network effects extends across country-borders with the help of FDI, which constitutes a strong argument in favor of globalization.

References

Acemoglu, Daron, Ufuk Akcigit, and William Kerr, "Innovation Network," *Proceedings of the National Academy of Sciences*, 2016, *113* (4).

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, *60* (2), 323–51.

_ , Nicholas Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics*, 2005, *120* (2), 701–728.

Aitken, Brian J. and Ann E. Harrison, "Do Domestic Firms Benefit from Direct Foreign Investment?" *American Economic Review*, 1999, *89* (3), 605–618.

Barba-Navaretti, G. and A. Venables, *Multinational Firms in the World Economy*, Princeton University Press, 2004.

Barrios, Salvador, Holger Gorg, and Eric Strobl, "Spillovers through backward linkages from multinationals: Measurement matters!" *European Economic Review*, August 2011, *55* (6), 862–875.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?* number wbsle. In 'Books from Upjohn Press.', W.E. Upjohn Institute for Employment Research, December 1991.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 07 2013, *81* (4), 1347–1393.

De Loecker, Jan and Frederic Warzynski, "Markups and Firm-Level Export Status," *American Economic Review*, 2012, *102* (6), 2437–2471.

Fosfuri, Andrea, Massimo Motta, and Thomas Ronde, "Foreign direct investment and spillovers through workers' mobility," *Journal of International Economics*, February 2001, *53* (1), 205–222.

Gorodnichenko, Yuriy, Jan Svejnar, and Katherine Terrell, "When does FDI have positive spillovers? Evidence from 17 transition market economies," *Journal of Comparative Economics*, 2014, *42* (4), 954–969.

Hall, Robert E., "Market Structure and Macroeconomic Fluctuations," *Brookings Papers on Economic Activity*, 1986, *17* (2), 285–338.

Harrison, Ann E., Leslie A. Martin, and Shanthi Nataraj, "Learning versus Stealing: How Important Are Market-Share Reallocations to India's Productivity Growth?" *World Bank Economic Review*, 2013, *27* (2), 202–228.

Haskel, Jonathan E., Sonia C. Pereira, and Matthew J. Slaughter, "Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?" *The Review of Economics and Statistics*, 2007, *89* (3), 482–496.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Jaffe, Adam B., "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, December 1986, *76* (5), 984–1001.

Jaffe, Adam B, "Demand and Supply Influences in R&D Intensity and Productivity Growth," *The Review of Economics and Statistics*, August 1988, *70* (3), 431–37.

Javorcik, Beata Smarzynska, "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages," *American Economic Review*, 2004, *94* (3), 605–627.

Kalemli-Ozcan, Sebnem, Bent E. Sørensen, and Vadym Volosovych, "Deep Financial Integration And Volatility," *Journal of the European Economic Association*, December 2014, *12* (6), 1558–1585.

_ , _ , Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yesiltas, "How to Construct Nationally Representative Firm Level data from the ORBIS Global Database and an Application to Industry Concentration," September 2015, (21558).

Keller, Wolfgang and Stephen R. Yeaple, "Multinational Enterprises, International Trade, and Productivity Growth: Firm-Level Evidence from the United States," *The Review of Economics and Statistics*, 2009, *91* (4), 821–831.

Kowalski, Przemyslaw, Daniel Rabaioli, and Sebastian Vallejo, "International Technology Transfer measures in an interconnected world: Lessons and policy implications," OECD Trade Policy Papers 206, OECD Publishing November 2017.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70* (2), 317–342.

Marin, Alvaro F. Garcia and Nico Voigtlander, "Exporting and Plant-Level Efficiency Gains: It's in the Measure," *Journal of Politcal Economy*, 2014, *Forthcoming*.

OECD, "Technology Intensity Definition, ISIC REV. 3," 2011.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64* (6), 1263–97.

Petrin, Amil, Jerome Reiter, and Kirk White, "The Impact of Plant-level Resource Reallocations and Technical Progress on U.S. Macroeconomic Growth," *Review of Economic Dynamics*, 2011, *14* (1), 3–26.

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

Wooldridge, Jeffrey M., "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 2009, *104* (3), 112–114.

A   Data

[Kalemli-Ozcan, Sørensen, Villegas-Sanchez, Volosovych and Yesiltas](2015) provide a comprehensive comparison of our data, relying on the Orbis database, with the best available national sources. Here we briefly touch upon some statistics relevant for the countries in our study and refer the reader to [Kalemli-Ozcan, Sørensen, Villegas-Sanchez, Volosovych and Yesiltas](2015) for more details.

A.1   Fraction of Total Output Covered

Table A.1 shows how much of the official gross output data, from Eurostat's *Structural Business Statistics* (SBS) database, is covered by the firms in our dataset in the manufacturing sector.[40] Each cell is the ratio of gross output produced by the firms in our dataset, relative to the value of gross output in the Eurostat data. Missing ratios appear in some country-year cells due to missing Eurostat data. For year 2008, our dataset accounts for close to 50 percent of output in Finland and for 69–90 percent in the other countries. [Kalemli-Ozcan, Sørensen, Villegas-Sanchez, Volosovych and Yesiltas](2015) demonstrate that the BvD has good coverage for the data used in the estimation of TFP, such as employment, wage bill, tangible fixed assets, and cost of materials. In addition, our dataset is representative when we count the number of firms and compare to numbers from Eurostat or to numbers from the sector-level CompNet database, which is constructed by the European Central Bank using sources similar to those of the BvD.[41]

Important for our study is the representativeness in terms of firm sizes within the manufacturing sector. Table A.2 reports the size distribution in the manufacturing sector by gross output accounted for by firms belonging to three size categories in year 2006 (other years are similar). Row entries denote the fraction of total economic activity accounted for by firms belonging to each size class. We match the official statistics well in terms of size distribution of economic activity in the manufacturing sector.

---

[40] "Gross output" in the BvD data is firms' operating revenue and "gross output" in the Eurostat database is called "turnover." The definitions are very similar and quite standard, and if there are any "fine print" deviations, they are unlikely to be important for our comparison.

[41] The comparison in terms of the number of businesses is somewhat noisy because many firms in Eurostat have zero employment (self-employed entrepreneurs), while neither our database nor CompNet includes the self-employed. Nonetheless, our data covers most of the total economy in terms of number of firms.

Table A.1: Coverage Based on Gross Output (Turnover) in the Manufacturing Sector, BvD vs. Eurostat Data

	Belgium	Finland	France	Italy	Spain	Norway
1999	0.75	0.30	0.64	0.61	0.75	0.60
2000	0.80	0.34	0.76	0.66	0.77	0.60
2001	0.78	0.36	0.79	0.65	0.78	na
2002	na	0.37	0.82	0.71	0.80	0.75
2003	0.81	0.39	0.79	0.70	0.79	0.68
2004	0.80	0.41	0.83	0.73	0.79	0.72
2005	0.80	0.41	0.82	0.77	0.78	0.69
2006	0.78	0.4	0.84	0.79	0.83	0.75
2007	0.79	0.45	0.87	0.79	0.81	0.76
2008	0.78	0.49	0.90	0.90	0.85	0.69

Notes: The table presents the ratio of output covered by the firms in the BvD sample to output from Eurostat's *Structural Business Statistics* (SBS) database. For a given country-year, the reported fraction is the ratio of aggregated gross output value reported by BvD firms in the manufacturing sector over total manufacturing output reported in the SBS database. "na" indicates that the latter number is 0.

Table A.2: Size Distribution by Gross Output (Turnover) in the Manufacturing Sector, 2006, BvD vs. Eurostat Data

	Belgium	Finland	France	Italy	Spain	Norway
A: Orbis-Amadeus						
1 to 19 employees	0.05	0.08	0.05	0.12	0.13	0.11
20 to 249 employees	0.30	0.38	0.23	0.49	0.40	0.40
250 + employees	0.66	0.54	0.72	0.40	0.47	0.49
B: EUROSTAT (SBS)						
0 to 19 employees	0.08	0.06	0.09	0.20	0.14	0.13
20 to 249 employees	0.27	0.21	0.27	0.41	0.38	0.36
250 + employees	0.65	0.74	0.64	0.39	0.49	0.51

Notes: The table presents the share of gross output accounted for by firms belonging to each of three size categories in the year 2006. The sample consists of firms that report data with positive values of gross output (turnover). Panel A reports the measures from our data based on Orbis-Amadeus and Panel B reports the same numbers from Eurostat's SBS database. Row entries denote the fraction of total economic activity accounted for by firms belonging to each size class. Each column is a different country.

A.2 Foreign Ownership

We validate our foreign ownership data by comparing it to the country-level *Activities of Foreign Affiliates* (AFA) database from the Organization for Economic Cooperation and Development (OECD), covering "affiliates under foreign control" (available at http://stats.oecd.org/Index.aspx?DataSetCode=AFA_IN3). The data is provided, largely, at the one-digit level of the ISIC rev. 3 or 4 classification, but it allows us to estimate the share of foreign output in manufacturing. The data is not directly comparable to our firm-level data. In our data, foreign investors are the direct foreign owners and we observe all stakes, including very small stakes held by multiple foreign entities. In the OECD data, foreign investors are, in principle, the ultimate foreign owners, although it is the immediate controlling entity for some countries. In addition, the notion of foreign affiliate in the OECD AFA database is based on the concept of controlling interest, which varies across countries. In most countries, the controlling interest is based on majority ownership (50%), while other countries also consider minority control (between 10% and 50%). Moreover, some countries include indirectly owned foreign affiliates in addition to directly controlled affiliates.

We compute foreign turnover from our data as the ratio of the foreign part of turnover, aggregated over firm-sector-countries, to the total turnover. That is, we compute $\frac{\sum_{s4,c}(\sum_{i\in s4,c} \text{fo}_{i,c,t} \times \text{go}_{i,c,t})}{\sum_{s4,c}(\sum_{i\in s4} \text{go}_{i,c,t})}$, in the full sample of firms and, separately, for the firms which remain in the sample in all years (the balanced sample of firms).

For the OECD data, we aggregate the multinational turnover from the AFA database, expressed in a single currency, add them up across countries, and then divide by the total monetary value of overall manufacturing turnover taken from the OECD's *STAN Database for Structural Analysis* (available at http://stats.oecd.org/Index.aspx?DataSetCode=STAN08BIS).

Figure A.1 presents a comparison of foreign shares of turnover based on data from the OECD and from the BvD. The dashed line, based on the all-firms BvD sample, and the solid line with circles, based on OECD data, almost coincide. The shares in the smaller BvD balanced sample of firms follow the same trend, but are a bit larger (by about a percentage point), because this sample consists, on average, of larger firms which are more likely to have foreign ownership. Overall, our data matches the foreign presence reported by government agencies to the OECD very well.

Figure A.1: Foreign Shares in Turnover: BvD vs. OECD Data
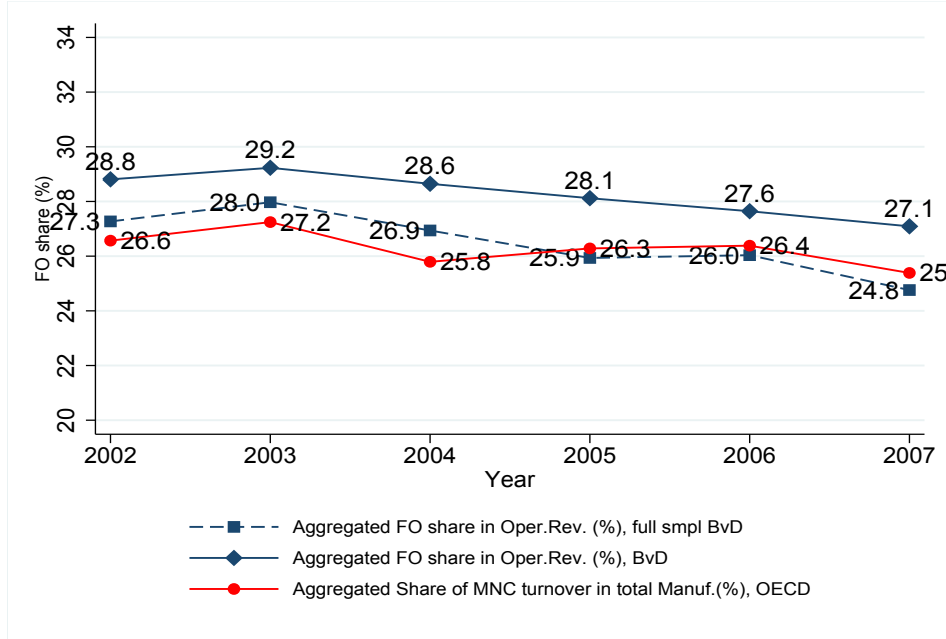
Notes: All ratios are in percent. The shares from the BvD data are computed as the ratios of the aggregated foreign turnover to total turnover as $\frac{\sum_{i \in s4,c} \mathrm{fo}_{i,c,t} \times \mathrm{go}_{i,c,t}}{\sum_{i \in s4} \mathrm{go}_{i,c,t}}$ over firms i, sectors $s4$, and countries c in the balanced firm sample (solid line with diamonds) and in the full sample (dashed line with squares). Foreign presence from the OECD data (solid line with circles) is the sum of the multinational turnover in manufacturing over countries divided by the total manufacturing turnover in these countries. Countries included are Finland, France, Italy, Norway, and Spain. Belgium is omitted because of missing OECD data due to confidentiality issues.

A.3 Citation Based Technology Closeness

We construct a measure of technology closeness between sectors based on patent citations by merging the information in two databases from the United States. The first is the *USPTO database*, which lists the number of patents by cited-citing technology class. Technology classes are defined according to 37 different subcategories[42]. This classification was created by Hall, Jaffe, and Trajtenberg and is based on the Patent Classification System as of 12/31/1999. The unit of observation is the citation between a pair of firms, and we use data on granted patents between 1976 and 2006. The second dataset is *Compustat*, which we merge with the information from the USPTO database in order to construct a sectoral measure of the overlap in technology class citation based on cited and citing sector. In particular, using a firm identifier that is present in both datasets, we can incorporate Compustat's sectoral information to the citation data, which allows us to generate an asymmetric citation matrix at the sector-pair level.

Specifically, we begin by merging the two files from the USPTO database: i) *pat76_06_assg.dta*, which contains information on patent id, granted year, and technology class; and ii) *cite76_06*, which has two columns: one for citing patent and another for cited patent. We merge the two files, so that we know for each citing-cited pair from the latter file in which year the cited patent was granted for the first time and in which year the patent is cited. We explored other cuts of the data in order to examine if the results are sensitive to using, for instance, more recent data or predetermined observations. We obtained similar results for the alternative cuts, specifically:

- 1976-1998: from the starting year of patent dataset to the year before our dataset starts so that everything is pre-sample. Further, in the citations dataset we see truncation from roughly 1999 onwards, which is another reason for ending in 1998 (truncation of citations is due to patent applications still not granted or incorporated in the official dataset).

- 1989-1998: past 10 years prior to the starting data of our paper's dataset in 1999.

- 1987-2006: past 20 years with available information.

The resulting file includes all the citations in the USPTO, which are uniquely identified at the citing patent and cited patent pair level. We aggregate this data from the patent- to the technology class-pair level and make it a fully balanced panel. Since we have 37 technological classes in our dataset, we end up with a dataset with 1,369 observations and a measure of the number of citations made between each pair of technological classes. We refer to this dataset as the *"Technological Input-Output Matrix."*

---

[42]For information on what they are, see page 41 of *http://www.nber.org/papers/w8498.pdf*

The final step is to match the USPTO database with Compustat and construct a sectoral measure which indicates the overlap in technology class citation based on cited and citing sector. The resulting database relies on the information on 6,325 unique currently listed or listed in the past Compustat firms that are associated with a total of 1,124,198 patents with the time period spanning 1976–2006. Once we match Compustat and USPTO, we know for each U.S. firm operating in sector $s4$ the number of patents created in each technology class.

A.4 Inventor Based Technology Closeness

We obtain the patents information from the NBER patent data project. It covers US patents data for the period 1976-2006 and several files, including assignee matches to Compustat are available on the downloads page.[43] The Community Members (Bronwyn Hall, Jim Bessen, and Grid Thoma) are responsible for providing this data for the scientific community. For each patent, the file `pat76_06_ipc.dta` provides us with the technology class, named `subcat`. This variable will allow us to assign a main technology class to each patent assignee. Crucially for us, they also provide a file that has standardized the patent applicant, named `pdpass` to be found in the file `assignee.dta`. The official documentation of the website provides an excellent explanation of their procedure:

> Finally, it is also worth noting what we mean by a patent "assignee". Assignee names are listed in the patent data from the USPTO, but they are not standardized. For example, there are over 100 different spellings, misspellings, abbreviations, etc. for the assignees of patents assigned to IBM. Using extensive name standardization and matching routines, we grouped these into a single "assignee" that is assigned a unique number stored in the variable PDPASS. This number is based on the first patent the firm was granted after January 1, 1976. Note that some firms do not have utility patents, but do have other patents, e.g., design patents. These firms have negative PDPASS numbers.

The next step is to combine the previous two files, i.e. `assignee.dta` and `pat76_06_ipc.dta`, in order to obtain the main technology class by patent assignee.

With this dataset at hand, what is left is to match this to the inventor side. To this end, we benefit from the inventor name disambiguation undertaken by Harvard Business School and its Patent Network Dataverse.[44] Large-scale studies of inventors are challenging, as the United States Patent and Trademark Office does not provide unique inventor identifiers. Many scholars of invention have implemented ad-hoc disambiguation methods based on string similarity thresholds and string comparison matching, but such methods have been shown to be vulnerable to a number

---

[43]See https://sites.google.com/site/patentdataproject/Home/downloads.

[44]See https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/15705 and also https://funginstitute.berkeley.edu/research/innovation-in-tech/tools-and-data/.

of problems that can adversely affect research results. The disambiguation method used in this paper is represented below and based on this paper: "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)" by Ronald Lai, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming.[45] The following quotation summarizes what the paper does:

> Using a Bayesian supervised learning approach, we identify individual inventors from the U.S. utility patent database, from 1975 to the present. An interface to calculate and illustrate patent co-authorship networks and social network measures is also provided. The network representation does not require bounding the social network beforehand. We provide descriptive statistics of individual and collaborative variables and illustrate examples of networks for an individual, an organization, a technology, and a region. The paper provides an overview of the technical algorithms and pointers to the data, code, and documentation, with the hope of further open development by the research community.

Once we generate a panel of inventors and track them over time, the last step is to match those patents to their patent assignees we identified later. This final match allows us to track the changes in patent assignees during the lifetime for a given inventor. For each change in employer that an inventor undertakes, we can identify the main technology class of both the previous and the new patent assignee. This information allows us to generate an asymmetric transition matrix of inventor mobility across technology classes which we then use in the regressions.

In our study, there is a potential endogeneity concern if foreign firm A enters sector S1 expecting to hire a lot of inventors from sector S2. To deal with this concern we divide sectors from the perspective of S1 into those that historically have transferred a lot of inventors to S1 (High Inflow (HI)) and those that have hardly transferred any inventors to S1 (Low Inflow (LI)). Suppose S2 falls under the LI category, in this case, the motive for firm A to enter S1 cannot be because it expects a lot of inventor mobility from S2 to S1, therefore, mitigating the selection concern. However, at the same time, given the asymmetric nature of the inventor mobility matrix even if S1 does not supply inventors to S2 it can still be that S2 supplies inventors to S1 leading to knowledge spillovers. To identify the HI and LI sectors, we use the transition matrix of inventor mobility across technology classes previously described.

---

[45]See https://hdl.handle.net/1902.1/15705.

B Production Function Estimation

B.1 Methodology

To obtain firm-level productivity estimates, we estimate the log-value added production function

$$y_{it} = \beta_0 + \beta_\ell \ell_{it} + \beta_k k_{it} + \omega_{it} + \epsilon_{it} \,, \tag{B.1}$$

where y_{it} is the logarithm of real output, ℓ_{it} is the logarithm of labor input, k_{it} is the logarithm of capital input, ω_{it} is the logarithm of physical productivity, and ϵ_{it} is a production shock that is not observable by the firm before making their input decisions at time t. The main concern, when estimating output elasticities with respect to the inputs in equation B.1, is whether the firm observes its own productivity ω_{it} at the time of making input choices. This would render input quantities endogenous to productivity and ordinary least squares (OLS) estimates of β_ℓ and β_k would be inconsistent. We follow the approach suggested in Wooldridge (2009), which builds on previous work by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP), which addresses the concerns raised by Ackerberg, Caves and Frazer (2015), who argue that if the flexible labor input is chosen as a function of unobserved productivity, the coefficient on labor input is not identified in the previous approaches.

The estimation is based on a two-step procedure to achieve consistency of the coefficient estimates for the inputs of the production function. Wooldridge (2009) suggests a generalized method of moments estimation of TFPR to overcome some limitations of OP and LP, including correction for simultaneous determination of inputs and productivity, no need to maintain constant returns to scale, and robustness to the Ackerberg, Caves and Frazer (2015) critique.[46] The following discussion is based on Wooldridge (2009), accommodated to the case of a production functions with two production inputs (see Wooldridge (2009) for a general discussion).

For firm i in time period t define:

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + e_{it} \,, \tag{B.2}$$

where y_{it}, l_{it}, and k_{it} denote the natural logarithm of firm value added, labor (a variable input), and capital, respectively. The firm-specific error can be decomposed into a term capturing firm-specific productivity ω_{it} and an additional term that reflects measurement error or unexpected productivity shocks e_{it}. We are interested in estimating ω_{it}.

---

[46]Ackerberg, Caves and Frazer (2015) highlight that if the variable input (labor) is chosen prior to the time when production takes place, the coefficient on variable input is not identified.

A key assumption of the OP and LP estimation methods is that for some function $g(.,.)$:

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

B.3

where m_{it} is a proxy variable (for investment in OP, for intermediate inputs in LP). Under the assumption,

$$E(e_{it}|l_{it}, k_{it}, m_{it}) = 0 \qquad t = 1, 2, ..., T,$$

B.4

substituting equation B.3 into equation B.2, we obtain the regression

$$
\begin{aligned}
E(y_{it}|l_{it}, k_{it}, m_{it}) &= \alpha + \beta_l l_{it} + \beta_k k_{it} + g(k_{it}, m_{it}) \\
&\equiv \beta_l l_{it} + h(k_{it}, m_{it}),
\end{aligned}
$$

B.5

where $h(k_{it}, m_{it}) \equiv \alpha + \beta_k k_{it} + g(k_{it}, m_{it})$.

In order to identify β_l and β_k, we need some additional assumptions. First, rewrite equation B.4 in a form allowing for more lags:

$$E(e_{it}|l_{it}, k_{it}, m_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, ..., l_{i1}, k_{i1}, m_{i1}) = 0 \qquad t = 1, 2, ..., T.$$

B.6

Second, assume productivity follows a first-order Markov process:

$$E(\omega_{it}|\omega_{i,t-1}, ..., \omega_{i1}) = E(\omega_{it}|\omega_{i,t-1}) \qquad t = 2, 3, ..., T,$$

B.7

and assume that the productivity innovation $a_{it} \equiv \omega_{it} - E(\omega_{it}|\omega_{i,t-1})$ is uncorrelated with current values of the state variable k_{it} as well as past values of the variable input l, the state k, and the proxy variables m:

$$
\begin{aligned}
E(\omega_{it}|k_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, ..., l_{i1}, k_{i1}, m_{i1}) \\
= E(\omega_{it}|\omega_{i,t-1}) \equiv f[g(k_{i,t-1}, m_{i,t-1})].
\end{aligned}
$$

B.8

Recall from equation(B.3) that $\omega_{i,t-1} = g(k_{i,t-1}, m_{i,t-1})$.

Plugging $\omega_{i,t} = f[g(k_{i,t-1}, m_{i,t-1})] + a_{it}$ into equation B.2 gives:

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + f[g(k_{i,t-1}, m_{i,t-1})] + a_{it} + e_{it}.$$

B.9

Now it is possible to specify two equations which identify (β_l, β_k):

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + g(k_{i,t}, m_{i,t}) + e_{it}$$

B.10

and

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + f[g(k_{i,t-1}, m_{i,t-1})] + u_{it}\,, \qquad \text{B.11}$$

where $u_{it} \equiv a_{it} + e_{it}$.

Important for the GMM estimation strategy, the available orthogonality conditions differ across these two equations. The orthogonality conditions for equation B.10 are those outlined in equation B.6, while the orthogonality conditions for equation B.11 are

$$E(u_{it}|k_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, ..., l_{i1}, k_{i1}, m_{i1}) = 0 \qquad t = 2, ..., T\,. \qquad \text{B.12}$$

To proceed with the estimation, we estimate these equations parametrically. We follow Petrin, Reiter and White (2011) and use a third-degree polynomial approximation using first order lags of variable input as instruments.[47]

B.2   Estimation Results

Table B.1 reports summary statistics for the output elasticities estimated using the Wooldridge (2009) approach. The results are consistent across countries with no major differences except for Belgium, where the number of observations is slightly lower and the coefficient on labor is on average marginally lower (0.601) and the average coefficient on capital marginally higher (0.113). Summary statistics are computed excluding sectors in which the WLP procedure delivers either missing, negative, or zero coefficients. These cases are few and mainly correspond to sectors 12 "Manufacture of Tobacco products" and 19 "Manufacture of coke and refined petroleum products," which have very few observations and contribute little to overall manufacturing output.

Table B.1: Summary Statistics of the Production Function Output Elasticities

	Labor Elasticity (β_ℓ)	Capital Elasticity (β_k)
Mean	0.719	0.083
Median	0.725	0.079
Standard Deviation	0.099	0.057
Max	0.943	0.573
Min	0.133	0.004

---

[47]We use the Stata routine suggested in Petrin, Reiter and White (2011).

C   Vertical Spillovers

Our analysis focuses on distinguishing between competition and knowledge spillover effects *within* two-digit sectors. There is an extensive literature, following the pioneer work of Javorcik (2004), which explores the role of inter-industry linkages and we verify that our data delivers similar results. The variable HORIZONTAL$_{s2,t-1}$ proxies for horizontal spillovers and is the two-digit sector counterpart of equation (2). The variable BACKWARD$_{s2,t-1}$ captures backward linkages between two-digit sectors and is computed at the two-digit industry level according to equation (3). To be consistent with the previous literature, we add the variable FORWARD$_{s2,t-1}$ that represents forward linkages between two-digit industries. The rationale for this variable is that domestic firms might experience productivity improvements as a result of buying cheaper or higher quality inputs from foreign firms. The variable is constructed following equation similarly to (3), but summing over the first index in the input-output coefficients; i.e., the amount that sector $s4$ buys on average from sector $\tilde{s}4$:

$$\text{VERTICAL\_IO}_{s2,t} = \sum_{\tilde{s}2 \neq s2} \alpha_{\tilde{s}2,s2,t} \times \text{HORIZONTAL}_{\tilde{s}2,t} \; . \tag{C.1}$$

Column (1) in Table C.1 shows the results without sector-year fixed effects, which corresponds to pooled country studies, and column (2) shows the results with sector-year fixed effects. The results in column (1) are consistent with productivity spillovers through backward linkages between two-digit sectors but no evidence of forward spillovers is found. Javorcik (2004) finds a similar result and argues that a potential explanation for the lack of evidence supporting forward linkages is that domestic firms lack the technological capability to benefit from the higher quality product produced by foreign firms. Column (2) confirms that the importance of backward linkages between two-digit sectors is not driven by omitted sector characteristics that vary over time and corroborates the lack of evidence supporting forward spillovers in this study.[48]

---

[48]Similarly, we do not find evidence of forward linkages within the same two-digit sector and therefore, we exclude this variable from our analysis.

Table C.1: Two-Digit Sector Vertical Linkages

DEPENDENT VARIABLE: log FIRM REVENUE TFP
SAMPLE: DOMESTIC FIRMS

	(1)	(2)
HORIZONTAL$_{s2,t-1}$	-0.183*	0.444***
	(0.105)	(0.105)
BACKWARD$_{s2,t-1}$	0.281***	0.140**
	(0.054)	(0.054)
FORWARD$_{s2,t-1}$	0.177	0.089
	(0.101)	(0.101)
Observations	322,705	322,705
Firm FE	✓	✓
Country-Year FE	✓	✓
Sec2-Year FE		✓
Cluster	cs2	cs2
s.d.(HORIZONTAL$_{s2}$)	0.13	0.13
s.d.(BACKWARD$_{s2}$)	0.11	0.11
s.d.(FORWARD$_{s2}$)	0.13	0.13

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. The dependent variable is the logarithm of revenue firm-level productivity at time t, (log TFPR$_{i,t}$). See section 2.3 for the detailed construction of the variables. The sample comprises domestic firms; i.e., firms that have no foreign participation over the years of analysis. All right-hand-side variables are lagged one period. Standard errors are clustered at the country-two-digit-sector-level. Results are obtained by GLS estimation using as weights the square root of each firm's mean squared predicted residuals from an initial OLS estimation. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

D Appendix Tables and Figures

Table D.1: Employment Characteristics Across Samples and Countries

Country	Sample	Mean	Median	SD
Total	original	61.4	26	157.1
	merged	65.2	27	173.4
Belgium	original	151.6	55	410.3
	merged	171.4	58	506.8
Spain	original	57.4	25	124.9
	merged	58.7	25	136.7
Finland	original	56.4	24	113.7
	merged	57.5	24	115.5
France	original	75.5	27	202.9
	merged	85.3	30	227.1
Italy	original	49.2	25	100.2
	merged	50.5	25	101.1
Norway	original	54.8	24	121.9
	merged	58.3	24	135.9

Notes: The table displays descriptive statistics for firms by country. "Original" sample refers to the sample of firms from the BvD with more than 10 employees, while the "merged" sample refers to the sample we end up with after the BvD dataset is merged with the dataset for which the technology closeness-measures can be calculated. See the text for further details.

Table D.2: Summary Statistics

	Observations	Mean	SD
log TFPR	322,698	3.55	0.93
μ	322,698	1.94	0.90
MC	322,698	0.68	0.13
$\#patents$	322,698	0.00	0.20
L	322,698	54.48	135.21
HORIZONTAL	322,698	17.45	16.01
HORIZONTAL_TEC	322,698	5.00	6.95
VERTICAL_IO	322,698	1.99	7.06
VERTICAL_TEC	322,698	21.23	14.86
VERTICAL_TEC^{CIT}	322,698	23.22	15.84

Notes: The table displays sample size, means, and standard deviations for our main variables for the sample of domestic firms used in the regressions. log TFPR is the logarithm of revenue total factor productivity; μ is the estimated markup; MC is the estimated marginal cost; #PATENTS is the number of granted patents; L is the number of employees; VERTICAL_TEC is the constructed measure of "technology-weighted knowledge spillovers" from outside the four-digit sector; HORIZONTAL_TEC is the constructed measure of "technology-weighted competition spillovers" from within the four-digit sector; HORIZONTAL is the constructed measure of "competition spillovers;" VERTICAL_TEC^{CIT} is the constructed measure of "technology-weighted knowledge spillovers" from outside the four-digit sector based on citations. Monetary values are deflated and expressed in constant 2010 dollars.

Table D.3: Correlation

	logTFPR	log μ	log MC	HORIZONTAL	HORIZONTAL_TEC	VERTICAL_IO	VERTICAL_TEC	VERTICAL_TEC^{CIT}
logTFPR	1							
log μ	0.9312	1						
log MC	-0.702	-0.7351	1					
HORIZONTAL	-0.0046	-0.0047	0.0021	1				
HORIZONTAL_TEC	0.0045	0.0012	-0.0011	0.7139	1			
VERTICAL_IO	0.003	0.0021	-0.0015	-0.0637	-0.1552	1		
VERTICAL_TEC	0.0049	0.0049	-0.0042	-0.1105	-0.2111	0.5231	1	
VERTICAL_TEC^{CIT}	0.0071	0.0083	-0.0024	-0.0901	-0.1519	0.428	0.7216	1

Notes: The table displays the correlation matrix of the "demeaned" variables (i.e., firm, country-year, and four-digit-sector-year fixed effects have been removed). The construction of these variables is explained in the main text. log TFPR is the logarithm of revenue total factor productivity; log μ is the logarithm of the estimated markup; log MC is the logarithm of (approximate) marginal cost; VERTICAL_TEC is "technology-weighted knowledge spillovers;" HORIZONTAL_TEC is "technology-weighted competition spillovers;" VERTICAL_TEC^{CIT} is "technology-weighted knowledge spillovers;" HORIZONTAL is "competition spillovers;" HORIZONTAL is "competition spillovers;" VERTICAL_TEC^{CIT} is "technology-weighted knowledge spillovers" based on citations.

Figure D.1: Spillover Measures by Two-Digit sector in 2007
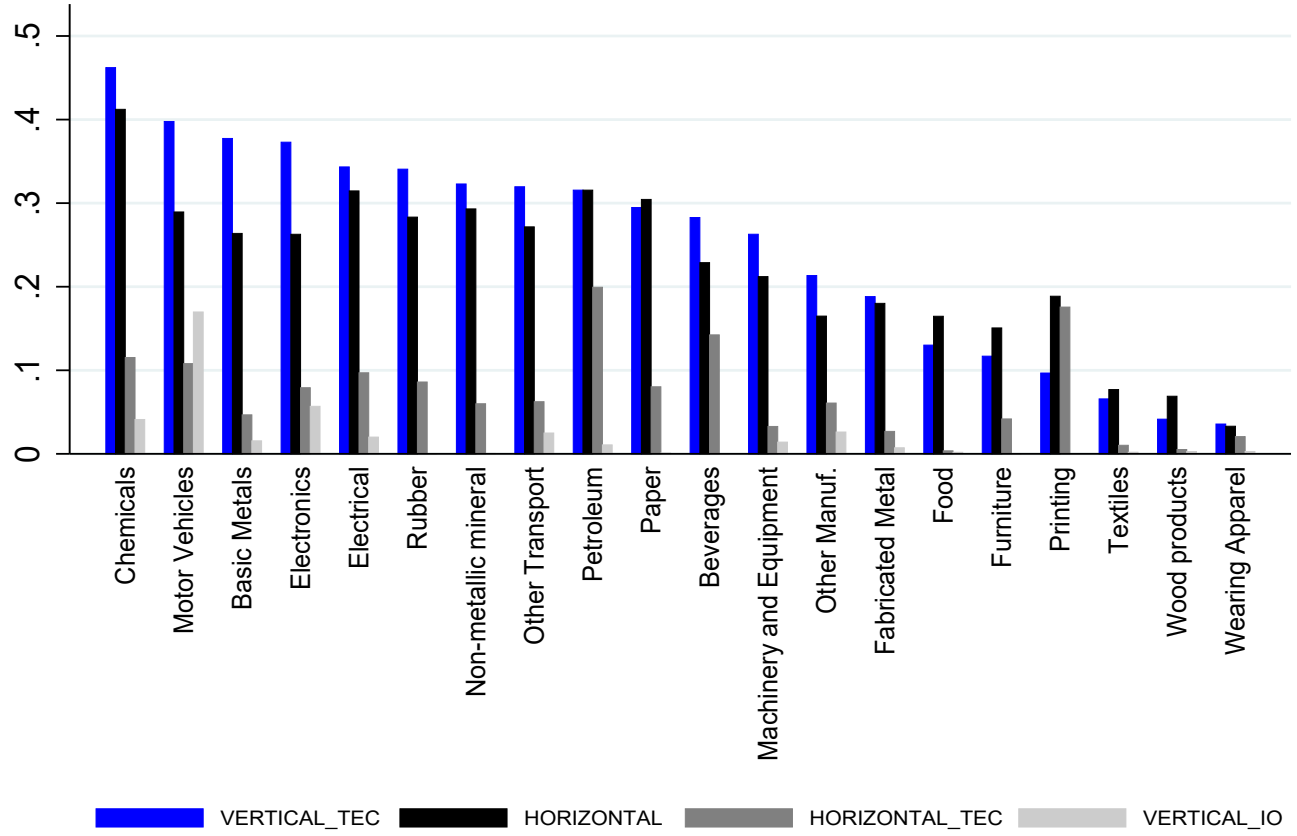
Notes: The figure displays the average values, by two-digit sector, of the main regressors, defined in Section 2.3. The data in these figures spans all countries in our sample but, unlike in the main paper, is for year 2007 only. We average over countries for that year.