Universitat Pompeu Fabra Barcelona
Department of Economics and Business

Economics Working Paper Series
Working Paper No. 1580

Foreign investment and domestic productivity: Identifying knowledge spillovers and competition

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

July 2017

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

July 2017

Abstract

We study the impact of foreign direct investment (FDI) on total factor productivity (TFP) of domestic firms using a new, representative firm-level data set spanning six countries. A novel finding is that firm-level spillovers from foreign firms to domestic companies can be significantly positive, non-existent, or even negative, depending on which sectors receive FDI. When foreign firms produce in the same narrow sector as domestic firms, the latter are negatively affected by increasing competition and positively affected by knowledge spillovers. We find that the positive spillovers dominate if foreign firms enter sectors where firms are "technologically close," controlling for the endogeneity of their entry decision into such sectors. Positive technology spillovers also affect firms in other sectors, if those sectors are technologically close to the sectors receiving FDI. Increasing FDI in sectors that are technologically close to other sectors boosts TFP of domestic firms by twice as much as increasing FDI by the same amount across all sectors.

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

*Affiliations: Universitat Pompeu Fabra, CEPR, and Barcelona Graduate School of Economics; University of Maryland, CEPR, and NBER; University of Houston and CEPR; ESADE - Universitat Ramon Llull; Erasmus University Rotterdam, Tinbergen Institute, and Erasmus Research Institute of Management. We thank Adam Szeidl for insightful comments on an earlier draft of this paper and discussants and participants of numerous presentations. Carolina Villegas-Sanchez acknowledges financial support from Banco Sabadell. Christian Fons-Rosen acknowledges financial support from the Spanish Ministry of Economy and Competitiveness through a grant (ECO-2014-55555-P) and through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563).. Veronika Penciakova provided superb research assistance.

1 Introduction

How does foreign direct investment affect the productivity of domestic firms? The most productive firms in a country tend to also operate globally and when such firms enter a foreign country, they are likely to affect the productivity of firms in that country. Consider, for example, the French company Valinox, a manufacturer of highly specialized steel tubing for nuclear steam generation. Since late 1990s, the company has competed with the German-owned multinational (MNC) Salzgitter Mannesmann Precision (SMP), which also operates in France in the same four-digit sector. The presence of SMP may hurt the productivity by undercutting Valinox's market shares, but, at the same time, it may benefit from SMP's cutting-edge technology. Valinox may learn about new technology, mimic production, sale and/or management practices of SMP, or hire workers who acquired new skills while working for SMP. Between 2000 and 2008, the productivity of Valinox rose by 25 percent, and we want to understand how much of the productivity increase experienced by Valinox, and other firms like it, is due to knowledge spillovers and competition effects from foreign firms like SMP.

In order to differentiate knowledge spillovers from such competition effects, we develop a sector-level measure of "technology closeness." We base our measure on whether firms within a sector hold similar patents, using data from Bloom, Schankerman and Van Reenen (2013). For NACE sector 2420, to which Valinox and SMP belong, 23 percent of patents registered by firms belong to the same technological patent classes as the patents of other firms in the sector. We construct a measure of horizontal technology-weighted FDI in a narrow four-digit sector, as the fraction of output produced by foreign owned firms weighted by our technological closeness measure for the sector. Using this technology-weighted FDI measure, we test whether FDI has heterogenous productivity effects on domestic firms in the same narrow sector as a function of the technological closeness of the firms in the sector. We find that domestic firms that are technologically close to multinationals become more productive following FDI, while firms that produce similar goods to MNCs, but are not technologically close, become less productive.

Knowledge spillovers may originate from other sectors. Consider Valinox again. The Dutch company Constellium Montreuil Juigne—the leading supplier of aluminum products and solutions for aerospace, transportation, and defense worldwide—operates in France in the same broader two-digit sector 24 as Valinox, but in a different narrow 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; hence, Valinox may also benefit from technological advances achieved by Constellium.[1] Firms in sector 2420 (Valinox) and firms in sector 2442 use similar technology, reflected in that 31 percent of patents registered by firms in sector 2420 overlap with patents of firms in sector 2442. The similarity of technology suggests a high potential for firms to learn from firms in the other sector. To account for such potential learning effects, we develop a measure of vertical technology-weighted FDI, by weighting the share of output produced by foreign firms with the technological distance to other narrow four-digit sectors in the same broad two-digit sector. We find that knowledge spillovers from other sectors are significantly positive when the sectors are technologically close to the sector of a given domestic firm. We show that this result is not due to vertical spillovers from customer-supplier linkages between foreign and domestic firms (inferred from input-output tables).

Identifying causal effects is a challenge. The entry decision of MNCs is endogenous to the performance of a given sector, so the number of MNCs in a sector is a function of sector-specific technological developments or sector-specific demand shocks. If domestic firms' productivity is correlated with such time-varying sector shocks, estimates of spillovers from foreign to domestic firms will be biased. For example, sectors where productivity worldwide is increasing due to technological breakthroughs, often have entry of MNCs and the correlation of MNC entry and productivity will be falsely interpreted as resulting from positive knowledge spillovers to domestic firms. Our data spans six advanced European countries, and because our measures of MNC entry are at the sector $\times$ country $\times$ year level, we can control for sector $\times$ year fixed effects. These fixed effects absorb worldwide sector-specific patterns and still allow us to identify the effect of sector-level changes in FDI over time. This identification methodology has not been used in the spillovers literature before because the literature has focused on one country at a time. We also include country $\times$ year fixed effects to account for potential effects of country-specific policy changes. After including these fixed effects, our results are not driven by different dynamic

[1] 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/.

patterns in productivity between sectors and/or between countries.

If domestic firms benefit from knowledge spillovers (or are hurt by competition effects), we expect that a fraction of them will actively upgrade their technology (or pull back from R&D), which will be reflected in the number of patents they acquire. To further support the interpretation of our findings in terms of knowledge/technology spillovers, we investigate the effect of FDI on the number of patents obtained.[2] We combine our data with a matched patent-firm dataset from Bloom, Draca and Van Reenen (2016) (who use a set of European firms similar to ours), to show that the entry of MNCs in a sector that is technologically close to that of the domestic firm is (on average) associated with an increase in patenting of the domestic firm.

We measure the productivity of domestic firms with revenue TFP. As is well known in the literature, changes in revenue TFP cannot separate changes in physical productivity from changes in prices in the absence of firm-level price data. This issue is particularly important because markups of domestic firms may respond endogenously to competition from MNCs, leaving the effect on physical TFP in doubt. We provide estimates of spillovers to physical TFP, using a method suggested by De Loecker and Warzynski (2012) to compute firm-level markups in order to isolate physical TFP. We find that the spillover effects on physical TFP constitute the larger part of the effects we have estimated for revenue TFP.

There is an extensive literature on FDI spillovers. The macro literature mostly finds a positive correlation between FDI and country-level growth. The empirical evidence in the micro literature is mixed: several papers have found robust evidence that the presence of MNCs generates positive vertical spillovers to domestic firms through customer-supplier relationships; see for example, Javorcik (2004) for evidence of vertical spillovers. There is conflicting evidence for horizontal spillovers. Javorcik (2004) and Barrios, Gorg and Strobl (2011) find little evidence of horizontal spillovers, Haskel, Pereira and Slaughter (2007) and Keller and Yeaple (2009) find positive horizontal spillovers, while Aitken and Harrison (1999) find negative horizontal spillovers. Harrison, Martin and Nataraj (2013) argue that "business-stealing" effects have dominated learning from MNCs in India and Keller and Yeaple (2009) show that FDI spillovers are shown to be strongest

[2]This induced effect is likely to be heterogeneous across firms; for example, Aghion, Bloom, Blundell, Griffith and Howitt (2005) explain how competition induces relatively more innovation in firms that a priori are competitive, while Bloom, Draca and Van Reenen (2016) document how increased competition from Chinese imports leads to innovations for the firms that are affected from Chinese competition, as measured by patents.

in high-technology industries and have a bigger impact within industries when domestic firms are most distant from the productivity frontier.[3] But while this suggests a role for technology that is similar to the one found in this paper, this literature did not attempt to construct sector-level measures to separate learning spillovers from competition effects.

Our contribution to the literature is twofold. First, we explain the mixed results in the literature and isolate the effect of each channel, i.e., negative competition effects and positive technology spillovers, on the TFP of domestic firms. The existing range of estimates for horizontal spillovers is very large due to the confounding of knowledge spillovers and competition effects. Second, we show that positive knowledge spillovers can happen without input-output linkages as long as the firms produce in technologically close sectors. Hence, we identify a central role of technological closeness between the foreign and domestic firms in facilitating knowledge spillovers. In this sense, our work is closely related to work on competition, innovation, and R&D; see, for example, 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 in one firm affects firms in technologically close fields. Our findings of positive knowledge spillovers to firms in sectors that are technologically close are in line with the existence of such innovation networks.

The rest of the paper is structured as follows. Section 2 presents our data and gives a detailed description of how we construct our measures. Section 3 discusses our empirical methodology. Section 4 presents our results and Section 5 concludes.

2 Data, Measurement and Construction of Variables

We use the Orbis global database, combined with Amadeus European database, from Bureau van Dijk (BvD). The Orbis database covers more than 200 countries and over 200 million firms (private

[3] There is an older literature that debates which way TFP effects on the host economy will go as a function of the technology gap between the country of the MNC and the host country. Some papers have argued that the expected positive effect of FDI will depend on the technology gap between the domestic firm and the foreign firm and/or on how far domestic firms are from the technology frontier. This effect can go either way; the potential for spillovers can be higher when the technology gap between foreign and domestic firms is large as in Ronald (1978), Blomstrom and Wolff (1994). Domestic firms will not have the capability to adapt new technologies if the gap is very large as explained in Cantwell (1989) and Kokko (1994).

and publicly listed), where longitudinal dimension and representativeness of the firms vary from country to country depending on whether small firms are required to file information. We focus on six advanced European countries (Belgium, Finland, France, Italy, Norway, and Spain) for the years 1999–2008, which gives us close to 600,000 firm-year observations.[4]

BvD collects data from various sources, in particular, publicly available national business and tax registries, and harmonizes the data into an internationally comparable format. The Orbis database provides consistent representative time series for both private and public firms for the countries we analyze in this paper.[5] The unit of observation is the firm, and, for each firm, we have full balance sheet information over time and unique sector codes at the four-digit NACE level. 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 MNC presence over time based on changes in ownership stakes.[6]

The extensive coverage of smaller, non-listed firms is important when one is interested in aggregate economic effects because large firms (those with +250 employees) account for a small part of total manufacturing output and employment in our sample of European countries. Other vendors compile datasets on foreign ownership—prominent examples are Thompson & Reuters and Dun & Bradstreet, but we prefer the Orbis database because of its extensive representative coverage of domestic firms, which are the objects of study here. The Orbis data set covers up to 70 percent of the real economy in many countries—see Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych and Yesiltas (2015). It has previously been used in the literature (see for example, Klapper, Laeven and Rajan (2006), Bloom, Draca and Van Reenen (2016), and Gopinath, Kalemli-Ozcan, Karabarbounis and Villegas-Sanchez (forthcoming); however, this paper is the first to use foreign ownership information together with estimates of firm-level productivity.

We exclude micro enterprises (those with less than ten employees according to the European

[4]The data for Norway are for the period 2000–2008.

[5]Significant effort is needed to put the longitudinal firm-level data set together, for both the financial series and for the ownership structure. The online dataset, or the current vintage, will only provide ownership current information on firms and if one uses only these data, the results will suffer from survivorship bias. It is also necessary to use older vintages of the data to avoid missing observations in balance sheet items for earlier years. Therefore, the dataset constructed for this study is downloaded from historical vintages of the database. See Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych and Yesiltas (2015) for a detailed explanation on how to construct nationally representative firm-level financial and ownership data from the BvD products.

[6]90 percent of FDI is conducted through acquisitions, according to the comprehensive literature survey of Barba-Navaretti and Venables (2004).

Commission definition) to be consistent with the literature on FDI spillovers. Our final sample accounts for more than ninety percent of manufacturing output in the countries under study, and it mimics the size distribution of firms provided by official national sources.

We next describe the main firm-level variables used in the analysis: revenue productivity, foreign ownership, number of patents, 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, where "domestic firms" are defined as firms that do not have foreign owners throughout the sample. 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.[7] Labor input, L_{it}, is measured as the firm's wage bill (deflated by the same two-digit industry price deflator).[8] Finally, we measure the capital stock, K_{it}, as the book value of fixed assets, deflated by the price of investment goods.[9]

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

[7]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.

[8]Using the wage bill, rather than the head count, helps adjust for differences in the quality of workers across firms because more skilled workers normally are paid more.

[9]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.

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. Our main results are qualitatively robust to estimating productivity from a translog production function following the methodology proposed in De Loecker and Warzynski (2012).

2.2 Firm-Level Patents

We match our sample of firms to the patent data provided by Bloom, Draca and Van Reenen (2016) who identify the firms in Amadeus that filed at least one European Patent Office (EPO) patent since 1978 and provide the number of granted patents by firm. Patents are dated by application year in order to measure the actual year of the invention (see Bloom, Draca and Van Reenen (2016) for more details on the matching procedure between Amadeus and the EPO dataset). Using the unique BvD firm identifier, we match our sample of firms to the patent data. We assume that unmatched firms were not granted any patents.

2.3 Firm-Level Foreign Ownership

To construct our main independent regressors, we rely on details of ownership changes from Orbis. 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. 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.[10] We compute "foreign ownership" as the sum of all percentages of *direct* ownership by foreigners.[11] We define a firm to be "domestic" if it never had *any* foreign owner during the sample period.

[10]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. BvD collects ownership data from official registers (including SEC filings and stock exchanges), annual reports, private correspondence, telephone research, company web-sites, and news wires.

[11]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.

2.4 Constructing Measures of Sector-Level FDI

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 $\alpha_{s4,\tilde{s4},t}$ is the input-output coefficient that records the fraction of its own output that sector $s4$ supplies to each given sector $\tilde{s4}$. 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).[12]

[12] Using the U.S.-based measures implicitly assumes that the patterns of input and knowledge flows in the six advanced European 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.

2.5 Constructing Measures of Sector-Level Technology Closeness

To construct measures of closeness between sectors in technology space, 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 $\mathrm{t}_i = (\mathrm{t}_{i1}, \mathrm{t}_{i2}, ..., \mathrm{t}_{i426})$, where t_{ix} is the share of patents of firm i in technology class x. Then, for each firm pair i, j in the sample, they construct measures of technology closeness, following Jaffe (1986), as the uncentered correlation of patent share vectors t_i and t_j:

$$\mathrm{tech}_{ij} = \frac{(\mathrm{t}_i \mathrm{t}_j')}{(\mathrm{t}_i \mathrm{t}_i')^{1/2} (\mathrm{t}_j \mathrm{t}_j')^{1/2}} \,. \tag{4}$$

They 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 $\mathrm{spilltech}_i = \Sigma_j \mathrm{tech}_{ij} \mathrm{rd}_j$ 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 Bloom, Schankerman and Van Reenen (2013)'s 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,[13] but Table D.1 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 that pre-date the analysis period to avoid potential endogeneity.

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

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:

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

where $s4$ and $\tilde{s}4$ refer to four-digit sectors.[14] 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. Figure 2 shows the distribution of the sectoral technological-closeness measure (5), where we split sectors by the OECD classification of the technology intensity of the sector.[15] Sectors that are classified by the OECD as technologically intensive (represented by the solid green line) feature larger values of our technological closeness measure (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:

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

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 $\tilde{s4}$

[14] 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.

[15] 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.

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.

2.6 Constructing Measures of Technology-Weighted FDI

Using the measures of technology closeness, we can define the "technology-weighted FDI/spillover" variables that will be our regressors. For the "technology-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.[16]

For potential knowledge spillovers between different four-digit sectors, we construct a "technology-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 sectors within the two-digit sector, apart from the sector that the measure is constructed for:

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

This measure is constructed similarly to the VERTICAL_IO measure, but using technology-adjusted size 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-

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

technology sectors.

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 Methodology

We regress domestic firms' TFP on the technology-weighted FDI variables, starting with horizontal measures:

$$\log\left(\mathrm{TFPR_{i,s4,c,t}}\right) = \beta_1 \text{HORIZONTAL}_{s4,c,t-1} + \beta_2 \text{HORIZONTAL\_TEC}_{s4,c,t-1} + \alpha_i + \phi_{s4,t} + \delta_{c,t} + \epsilon_{i,s4,c,t}\ , \quad (9)$$

where $\mathrm{TFPR_{i,s4,c,t}}$ 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 technologically-weighted measure of FDI is likely to capture the effect of knowledge spillovers.

The dependent variable is revenue productivity, so the estimated coefficient could be driven by technology transfers or by demand-price effects. We abstract from this discussion for now and return to it in section 4.1. Foreign investment is likely to gravitate towards rapidly developing sectors with increasing productivity such as Sector 27, Manufacture of Electrical Equipment. Productivity of this sector grew by over 10 percent per year over our sample, which may have attracted FDI to the sector (or both FDI and productivity may have been caused by omitted sector-level variables). We hedge against such effects, to the extent that they are common across countries, by including sector-year fixed effects. Similarly, we can hedge against our results being caused by common reactions to country-level changes in, say, business climate, by including country-year fixed effects. The inclusion of a large number of fixed effects stacks the deck against finding significant results, but lends higher credence to those results that are significant. The inclusion of sector-year dummies implies that our results are identified from the relative differences between sectors, which is why our regressors are designed to measure relative differences *within* two-digit sectors.

Our main empirical regression combines four different channels of potential impact of FDI/MNC presence within each two-digit sector. The specification estimated is

$$\begin{aligned}\log\left(\mathrm{TFPR_{i,s4,c,t}}\right) &= \beta_1 \text{HORIZONTAL}_{s4,c,t-1} + \beta_2 \text{HORIZONTAL\_TEC}_{s4,c,t-1} + \\ &\quad \beta_3 \text{VERTICAL\_TEC}_{s4,c,t-1} + \beta_4 \text{VERTICAL\_IO}_{s4,c,t-1} + \\ &\quad \alpha_i + \phi_{s4,t} + \delta_{c,t} + \epsilon_{i,s4,c,t}\ ,\end{aligned} \tag{10}$$

where $\mathrm{TFPR_{i,s4,c,t}}$ refers to total factor revenue 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 competition, once the potential positive effect from knowledge transfers by direct competitors is accounted for by HORIZONTAL_TEC (with a positive β_2). β_3 will be positive if domestic firms benefit from the presence of foreign investors in closely related technological sectors, and we expect β_4 to be positive, capturing backward spillovers within the two-digit sector.

Firms are heterogeneous and while most of the existing literature estimates equations similar to equation (9) 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). We prefer GLS because it is a more efficient estimator than OLS. We cluster standard errors at the country-four-digit sector level (allowing for autocorrelation).[17]

[17] 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.

4 Results

In Table 1, we report the results from the regression in equation (9).

Table 1 about here

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

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).

competition effect dominates in sectors where the foreign-owned firms are not technologically close.

Our main regression, presented in Table 2, utilizes four different measures of potential FDI spillovers, separating competition and knowledge effects from within the four-digit sector and capturing knowledge spillovers from outside the four-digit sector. As in Table 1, we use the measures of horizontal knowledge FDI (technology-weighted or not), but now we also use the measures of technology-weighted vertical FDI from outside the four-digit sector (captured by VERTICAL_TEC). We also include the standard non-technology weighted backward-spillover measure (captured by VERTICAL_IO) which captures knowledge spillovers from outside the four-digit sector via input-output links. We include the variables sequentially, in order to evaluate if the effects are robust to specification.

Table 2 about here

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 takes a positive coefficient, but it is not significant when the technology-weighted measure is included. 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, where the insignificant backward-spillover measure is not included.

Overall, productivity increases with foreign presence in technologically close sectors. The coefficients are standardized so they can be compared to each other, and 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. The result may not be surprising, given that Bloom, Schankerman and Van Reenen (2013) show that firms learn from the technological innovation of firms that are close in technology space, but the relevance of technological distance for FDI spillovers seems not to have been realized in the literature previously.

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

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:

$$\mathrm{TFPR}_{it} \equiv P_{it}\mathrm{TFPQ}_{\mathrm{it}} = \mu_{\mathrm{it}} \times \mathrm{MC}_{\mathrm{it}} \times \mathrm{TFPQ}_{\mathrm{it}}\,, \tag{11}$$

where P_{it} refers to the firm's output price, and $\mathrm{TFPQ}_{\mathrm{it}}$ is physical productivity. If firm-specific prices were available to deflate nominal output, physical productivity ($\mathrm{TFPQ}_{\mathrm{it}}$) 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 (11). Rearranging 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: $\Delta\mathrm{TFPR}_{it} - \Delta\mu_{it} = \Delta\mathrm{TFPQ}_{\mathrm{it}} + \Delta\mathrm{MC}_{\mathrm{it}}$. Reordering, we obtain a formula for the change in physical productivity:

$$\Delta\mathrm{TFPQ}_{it} \;=\; \Delta\mathrm{TFPR}_{\mathrm{it}} - \Delta\mu_{\mathrm{it}} - \Delta\mathrm{MC}_{\mathrm{it}}\,. \tag{12}$$

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:

$$\mu_{it} \equiv \frac{P_{it}}{MC_{it}} = \underbrace{\frac{\partial \mathcal{F}_{it}(\cdot)}{\partial \mathcal{J}_{it}} \frac{\mathcal{J}_{it}}{\mathcal{F}_{it}(\cdot)}}_{OutputElasticity} / \underbrace{\frac{P_{it}^{\mathcal{J}_{it}} \mathcal{J}_{it}}{P_{it} \mathrm{y}_{it}}}_{ExpenditureShare}, \tag{13}$$

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 $P_{it}\mathrm{y}_{it}$ 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.[18] We find empirical estimates of the median markup close to 1.5 on average, which is similar to the estimates in the literature.

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 (12). 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 foreign ownership using, in particular, firm fixed effects. Regressions with firm fixed effects are identified from changes in the variables over the sample period and, while the change over time in average cost over revenue is not literally the same as marginal cost, any firm-level constants average out, and 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 about here

[18] 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.

We consider the effect of each measure of FDI in turn, starting with knowledge spillovers. Column (1) in Table 3 repeats our main specification (showing standardized coefficients) with the dependent variable log TFPR. Consider knowledge spillovers from outside the four-digit sector: a one standard deviation increase in VERTICAL_TEC 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.[19] 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 then 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

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

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.

4.2 Direct Evidence on Technology Spillovers: FDI and Domestic Firm Patenting

Our results so far document that an increase in the presence of foreign firms significantly impacts the productivity of domestic firms, but the results do not show whether this is the result of passive learning or active innovation by the domestic firms exposed to foreign presence. The firm-level patent data for Europe allows us to dig deeper by examining if patenting activity changes with foreign exposure. We regress the logarithm of one plus the number of patents on our spillover measures.[20]

Table 4 about here

Table 4 presents the results for patenting. Column (1) shows results for the full sample and column (2) shows results for firms in the sample in all years.[21] In column (1), the estimated impact of VERTICAL_TEC is large and significant, while the horizontal measures are insignificant. From column (2), the effect on permanent firms is similar for the vertical measure, while the estimated impact of the horizontal measures are insignificant, but similar in magnitudes to the estimates found for tfp. In the permanent sample, a one-standard-deviation increase in foreign presence in the four-digit sector will lead to a 0.213 percent decrease in patenting if the foreign firms are not technologically close, while the effect of HORIZONTAL_TEC and HORIZONTAL will roughly offset each other if they both increase by a standard deviation.

An implication of these results is that the presence of foreign firms using technology similar to

[20]The distribution of patents by firm is highly right-skewed, and we take the logarithm to avoid our estimates being unduly affected by a few outliers. Most firms have no patents so we add unity to avoid taking the logarithm of zero.

[21]We use OLS estimations in this table, because the large number of zeros for the dependent variable makes this regression unsuitable for weighting because a value of 0 will not change with any heteroskedasticity weighting.

that used by a given domestic firm provides incentives for the domestic firm to actively innovate, as witnessed by the higher patent activity. This is consistent with the findings of Aghion, Bloom, Blundell, Griffith and Howitt (2005) and Bloom, Draca and Van Reenen (2016), who document increased patenting when competition changes, although these papers do not consider the role of foreign firms. The change in patent activity further lends credence to our results, because the intensity of patenting is directly measured, while our measures of productivity are estimated which can lead to bias.

4.2.1 Robustness

Table 5 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, to the use of an alternative (citation-based) measure of technological closeness, and to the inclusion of a direct measure of competition.

Table 5 about here

The first column displays our preferred regression. The literature has typically presented OLS estimates, which do not downweigh firms with volatile productivity. We demonstrate in Column (2) that the choice of GLS is not crucial for our results: the horizontal technology spillover measure is insignificant using GLS, while the effects of competition and vertical knowledge spillovers are estimated with coefficients that are very similar to those of the GLS estimation. On net, the estimated broad effects of FDI on productivity are similar.

The results of column (3), estimated using a permanent sample of firms, are similar to those obtained using the full sample, although the estimated coefficients are larger for the permanent firms (the standardized coefficients can be compared to those of Table 2 because the standard errors of the variables in the permanent 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 with the concentration measure. In column (7), we use a different measure of vertical spillovers, where the technological distance (based on U.S. data) is based on citations. The construction of this measure is described in Appendix C. The results are very similar to those obtained in column (1) indicating that the our findings are not an artifact of the exact way the technology weights are constructed.[22] 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.

[22] In unreported regressions, we attempted to include both measures based on our distance measure and the measure based on citations, but the measures are too highly correlated to find separate effects of technological closeness based on similarity of patenting and on citations.

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 6 repeats our main specification estimated in one-, two-, and four-year (non-overlapping) differences.[23]

Table 6 about here

In Table 6, 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.[24]

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

[23]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.

[24]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.

coefficient is 0.054, even larger than the level effect.[25] 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 5, 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.[26] 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 to their results, but we find it reassuring that our estimates and those of the literature are similar in magnitude.

[25]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.

[26]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.

4.4 Economic Significance and Aggregate 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.

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:

$$\begin{aligned} \Delta \log \widehat{(\mathrm{TFPR}_{s4})} &= \hat{\beta}_1 \Delta \textsc{vertical\_tec}_{s4} + \hat{\beta}_2 \Delta \textsc{horizontal}_{s4} \\ &\quad + \hat{\beta}_3 \Delta \textsc{horizontal\_tec}_{s4} \; . \end{aligned} \tag{14}$$

Because the right-hand-side spillover variables in equation (14) 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:

$$
\begin{aligned}
\Delta\text{VERTICAL\_TEC}_{s4} &= \sum_{\tilde{s}4,\tilde{s}4\in s2(s4),\tilde{s}4\neq s4} \text{WTECH}_{s4,\tilde{s}4} \times \Delta\text{HORIZONTAL}_{\tilde{s}4} \\
&= \sum \text{WTECH}_{s4,\tilde{s}4} \times \Delta\text{fo}_{\tilde{s}4} \\
&= \Delta\text{fo}_{\tilde{s}4}\ .
\end{aligned}
\tag{15}
$$

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

$$
\Delta\text{HORIZONTAL}_{s4} = \frac{\sum_{i\in s4}\Delta\text{fo}_{s4}\times \text{go}_i}{\sum_{i\in s4}\text{go}_i} = \Delta\text{fo}_{s4}\ . \tag{16}
$$

We have

$$
\Delta\text{HORIZONTAL\_TEC}_{s4} = \text{WTECH}_{s4,s4} \times \Delta\text{fo}_{s4}\ . \tag{17}
$$

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

$$
\Delta \log\widehat{(\text{TFPR}}_{s4}) = \left(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3\,\text{WTECH}_{s4,s4}\right)\Delta\text{fo}\ . \tag{18}
$$

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

$$
\text{ATC}_{s4} = \frac{1}{N}\sum_{\tilde{s}4\in s2} \text{WTECH}_{\tilde{s}4,s4}\ , \tag{19}
$$

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

$$\begin{aligned} \Delta \log \widehat{(\text{TFPR}_{s4})} &= \hat{\beta}_1 \sum_{\tilde{s}4, D_{connected\ \tilde{s}4}=1} \text{WTECH}_{s4,\tilde{s}4} \times (2\,\Delta\text{fo}) \\ &+ D_{connected\ s4} \times (\hat{\beta}_2 + \hat{\beta}_3 \text{WTECH}_{s4,s4}) \times (2\,\Delta\text{fo}). \end{aligned} \tag{20}$$

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 7, 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 7 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 (18).

Table 7 about here

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 explain why researchers have had a difficult time identifying horizontal spillovers. As we showed, knowledge spillovers will be negative if domestic firms operate in sectors that are not technologically close to the sectors that MNCs enter. We do not extend our analysis to emerging markets here, but if MNCs typically invest in less connected sectors in emerging markets, our results would explain why positive horizontal knowledge spillovers are not typically found there.

5 Conclusion

We identify the effect of FDI on domestic firms' productivity, separating competition effects which hurt productivity from knowledge spillovers which increase productivity. We take a different route from the existing literature in that we allow the productivity effects to depend on the location of MNC entry in technology space. The standard approach in the spillover literature is to search for spillover effects in the host country when MNCs enter, but an important policy conclusion from our paper is that it matters *in which sectors* MNCs enter. Foreign investment in sectors that are technology close to many sectors in the host economy is associated with an increase in productivity, while investment in technologically disconnected sectors is associated with limited knowledge spillovers. In the latter case, a loss in domestic productivity due to competition effects is dominant.

We use firm-level, granular data for manufacturing with detailed sector classifications at the four-digit level both for domestic and foreign firms from six European economies during 2000–2008—a period where FDI has increased tremendously within Europe. We identify the closeness of domestic and foreign firms in product markets and in technology space. We find that when MNCs enter the same narrow four-digit sector as domestic firms, there are positive knowledge spillovers to domestic firms. These positive spillovers are on average almost fully offset by negative competition effects, although the knowledge spillover effect will dominate in sectors where firms are close in technology space, while the competition effect will dominate if firms are distant in technology space. If MNCs enter into a different four-digit sector than that which a domestic firm is operating in, there will be significant technology spillovers to the firm if the MNC entry takes place in technologically close sectors. Our results are not driven by unobserved sector-level productivity that impacts the entry decision of the MNC. By showing evidence on a new "technology" channel for spillovers from foreign firms, our paper demonstrates that it is possible to have significant positive horizontal knowledge spillovers from FDI.

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.

Blomstrom, Magnus and Edward Wolff, *Multinational corporations and productivity convergence in Mexico* Convergence of productivity: Cross-national studies and historical evidence, Oxford University Press, 1994.

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

_ , Mirko Draca, and John Van Reenen, "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity," *Review of Economic Studies*, 2016, *83* (1), 87–117.

Cantwell, J.A., *Technological Innovation and Multinational Corporations*, Oxford: Basil Blackwell, 1989.

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

Gopinath, Gita, Sebnem Kalemli-Ozcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez, “Capital Allocation and Productivity in South Europe,” *Quarterly Journal of Economics*, forthcoming.

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.

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 Sorensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yesiltas, “How to Construct Nationally Representative Firm Level data from the ORBIS Global Database,” NBER Working Papers 21558, National Bureau of Economic Research, Inc September 2015.

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.

Klapper, Leora, Luc Laeven, and Raghuram Rajan, "Entry regulation as a barrier to entrepreneurship," *Journal of Financial Economics*, December 2006, *82* (3), 591–629.

Kokko, Ari, "Technology, market characteristics, and spillovers," *Journal of Development Economics*, April 1994, *43* (2), 279–293.

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," *working paper*, 2014.

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.

Ronald, Findlay, "Relative Backwardness, Direct Foreign Investment, and the Transfer of Technology: A Simple Dynamic Model," *The Quarterly Journal of Economics*, 1978, *92* (1), 1–16.

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

Table 1: Horizontal Spillovers

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 $\mathrm{TFPR}_{\mathrm{i,t}}$). See section 2.4 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.

Table 2: Technology Spillovers and Vertical Linkages

Dependent Variable: log Firm Revenue TFP
Sample: Domestic Firms

	(1)	(2)	(3)	(4)	(5)
$\text{VERTICAL\_TEC}_{s4,t-1}$	0.297***		0.267**	0.282**	0.312***
	(0.074)		(0.089)	(0.089)	(0.074)
$\text{VERTICAL\_IO}_{s4,t-1}$		0.113**	0.042	0.049	
		(0.035)	(0.042)	(0.042)	
$\text{HORIZONTAL}_{s4,t-1}$				-0.336***	-0.336***
				(0.080)	(0.080)
$\text{HORIZONTAL\_TEC}_{s4,t-1}$				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, ($\log \text{TFPR}_{\text{i,t}}$). See section 2.4 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.

Table 3: Revenue TFP and Markups

Sample: Domestic Firms

Dependent Variable:	log TFPR (1)	log μ (2)	log MC (3)	Implied log TFPQ (4)
VERTICAL_TEC$_{s4,t-1}$	0.312***	0.267***	-0.267***	0.312***
	(0.074)	(0.074)	(0.045)	(0.114)
HORIZONTAL$_{s4,t-1}$	-0.336***	-0.192**	0.144***	-0.288***
	(0.080)	(0.064)	(0.032)	(0.107)
HORIZONTAL_TEC$_{s4,t-1}$	0.271***	0.076	-0.104**	0.299***
	(0.076)	(0.076)	(0.042)	(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 $\mathrm{TFPR}_{\mathrm{i,t}}$). 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 2.4 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.

Table 4: Direct Evidence on Technology Spillovers

DEPENDENT VARIABLE: log(1 + PATENTS)
SAMPLE: DOMESTIC FIRMS

	Full Sample (1)	Permanent Sample (2)
VERTICAL_TEC$_{s4,t-1}$	0.227**	0.213*
	(0.076)	(0.122)
HORIZONTAL$_{s4,t-1}$	-0.049	-0.216
	(0.098)	(0.166)
HORIZONTAL_TEC$_{s4,t-1}$	-0.071	0.153
	(0.085)	(0.139)
Observations	197,841	67,490
Firm FE	✓	✓
Country-Year FE	✓	✓
Sec4-Year FE	✓	✓
Cluster	cs4	cs4
s.d.(VERTICAL_TEC)	0.15	0.15
s.d.(HORIZONTAL)	0.16	0.16
s.d.(HORIZONTAL_TEC)	0.07	0.07

Notes: The table reports standardized coefficients and standard errors in parenthesis multiplied by 100. Column (1) uses the total sample of domestic firms while column (2) focuses on the sub-sample of domestic firms continuously observed during the full sample period. The dependent variable is the logarithm of the number of granted patents to firm i at time t, (log $(1 + \text{PATENTS})$). See section 2.4 for the details of the construction of the variables. Period 1999-2005. 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 OLS. *** denotes 1% significance; ** denotes 5% significance; * denotes 10% significance.

Table 5: Robustness

Dependent Variable: log Firm Revenue TFP
Sample: Domestic Firms

	(1) Benchmark	(2) OLS	(3) Permanent	(4) empl10	(5) sec2-year	(6) HHN	(7) Citation
$\text{VERTICAL\_TEC}_{s4,t-1}$	0.312***	0.342**	0.407**	0.442***	0.461***	0.282***	
	(0.074)	(0.163)	(0.136)	(0.091)	(0.089)	(0.074)	
$\text{HORIZONTAL}_{s4,t-1}$	-0.336***	-0.336*	-0.494***	-0.527***	-0.528***	-0.288***	-0.320***
	(0.080)	(0.176)	(0.132)	(0.099)	(0.080)	(0.080)	(0.080)
$\text{HORIZONTAL\_TEC}_{s4,t-1}$	0.271***	0.09	0.548***	0.362***	0.688***	0.341***	0.243**
	(0.076)	(0.195)	(0.153)	(0.094)	(0.083)	(0.076)	(0.076)
$\text{HERFIN}_{s4,t-1}$						-0.582***	
						(0.116)	
$\text{VERTICAL\_TEC}^{CIT}_{s4,t-1}$							0.333***
							(0.048)
Observations	322,698	322,698	101,424	221,907	322,705	322,698	322,698
Firm FE	✓	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	✓	✓	✓	✓	✓
Sec4-Year FE	✓	✓	✓	✓		✓	✓
Sec2-Year FE					✓		
Cluster	cs4	cs4	cs4	cs4	cs4	cs4	cs4
s.d.(HORIZONTAL)	0.16	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	0.07
s.d.(VERTICAL_TEC)	0.15	0.15	0.15	0.15	0.15	0.15	0.15
s.d.(HERFIN)						1.16	
s.d.($\text{VERTICAL\_TEC}^{CIT}$)							0.16

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 $\text{TFPR}_{\text{i,t}}$). See section 2.4 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.

Table 6: FDI Spillovers: Long-term effects

Sample: Domestic Firms

	(1)	(2)	(3)	(4)
Dependent Variable:	log TFPR	$\Delta^{j=1}$ log TFPR	$\Delta^{j=2}$ log TFPR	$\Delta^{j=4}$ log TFPR
Δ^jVERTICAL_TEC$_{s4}$	0.021***	0.001	0.032***	0.053***
	(0.005)	(0.005)	(0.004)	(0.004)
Δ^jHORIZONTAL$_{s4}$	-0.021***	-0.023***	-0.014**	-0.017***
	(0.005)	(0.006)	(0.006)	(0.005)
Δ^jHORIZONTAL_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 (Δ^j log, $\text{TFPR}_{\text{i,t}}$) where j indicates, first, second, and fourth differences. See section 2.4 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.

Table 7: Predicted Effects of Increases in FDI

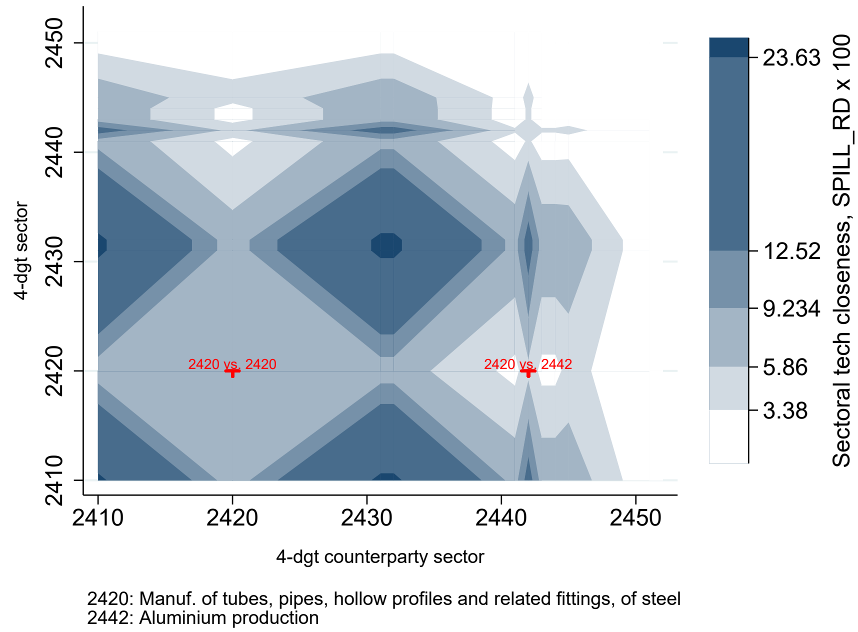
Sample: Domestic Firms

		(1)=(2)+(3)+(4)	(2)	(3)	(4)
Δ fo	Targeted Sectors	Δ log TFPR	Δhorizontal	Δhorizontal_tec	Δvertical_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

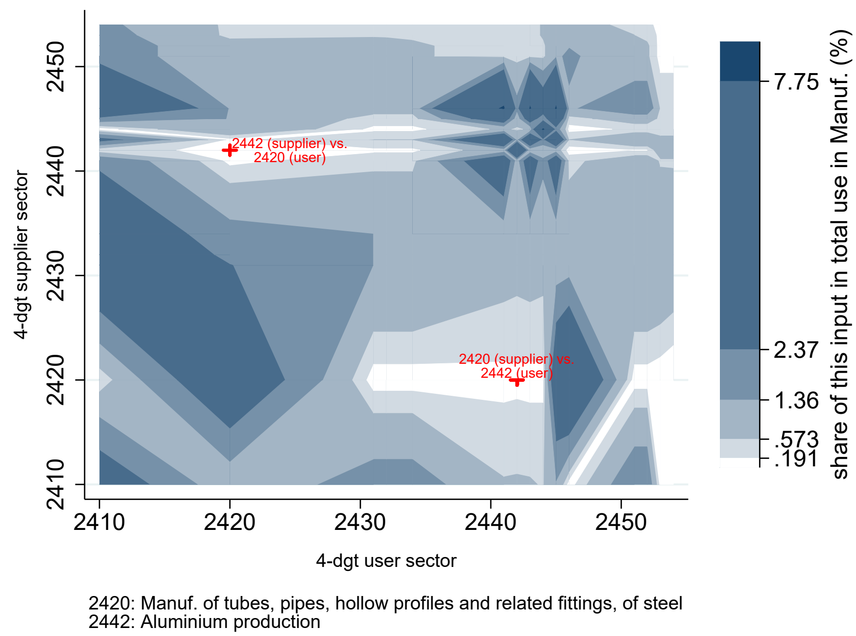
Notes: Row (1) displays the predicted average increase in TFPR in manufacturing as a result of a 10 percentage point increase in FDI, equally distributed across sectors. Row (2) displays the predicted average increase in TFPR in manufacturing as a result of a 10 percentage point increase in FDI, distributed 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 (19). 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.

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

A. Technological Closeness

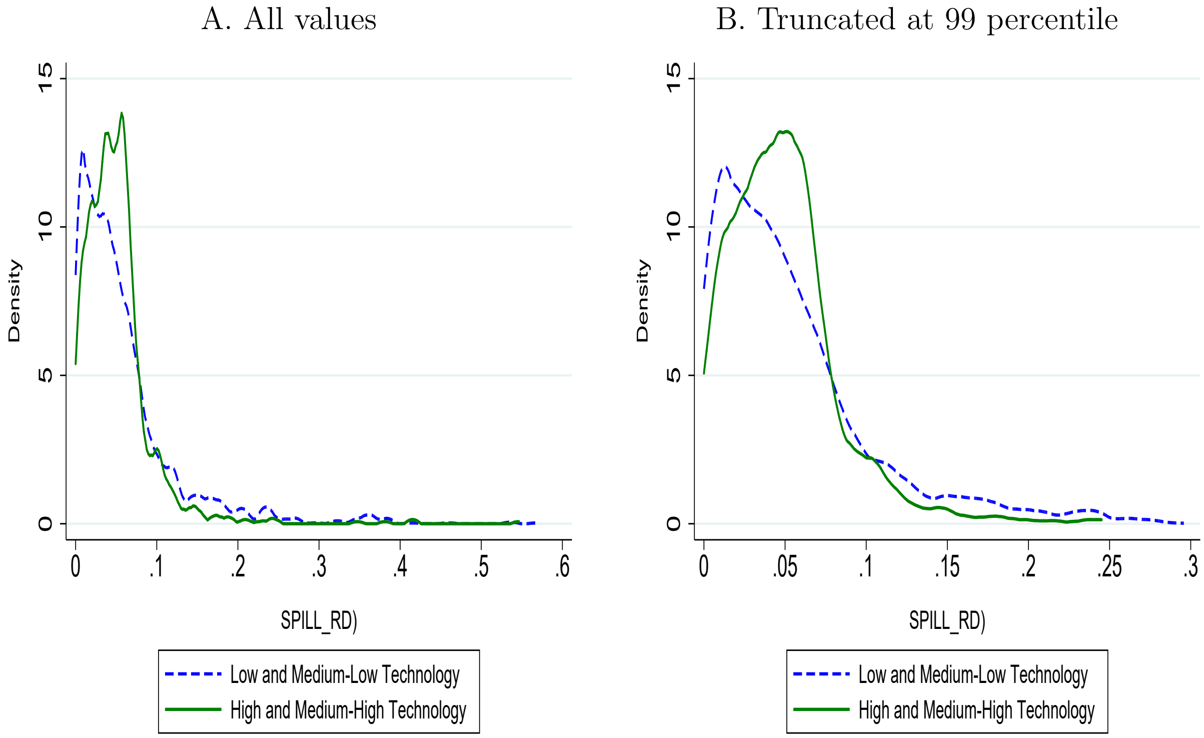


B. Input-Output Coefficients



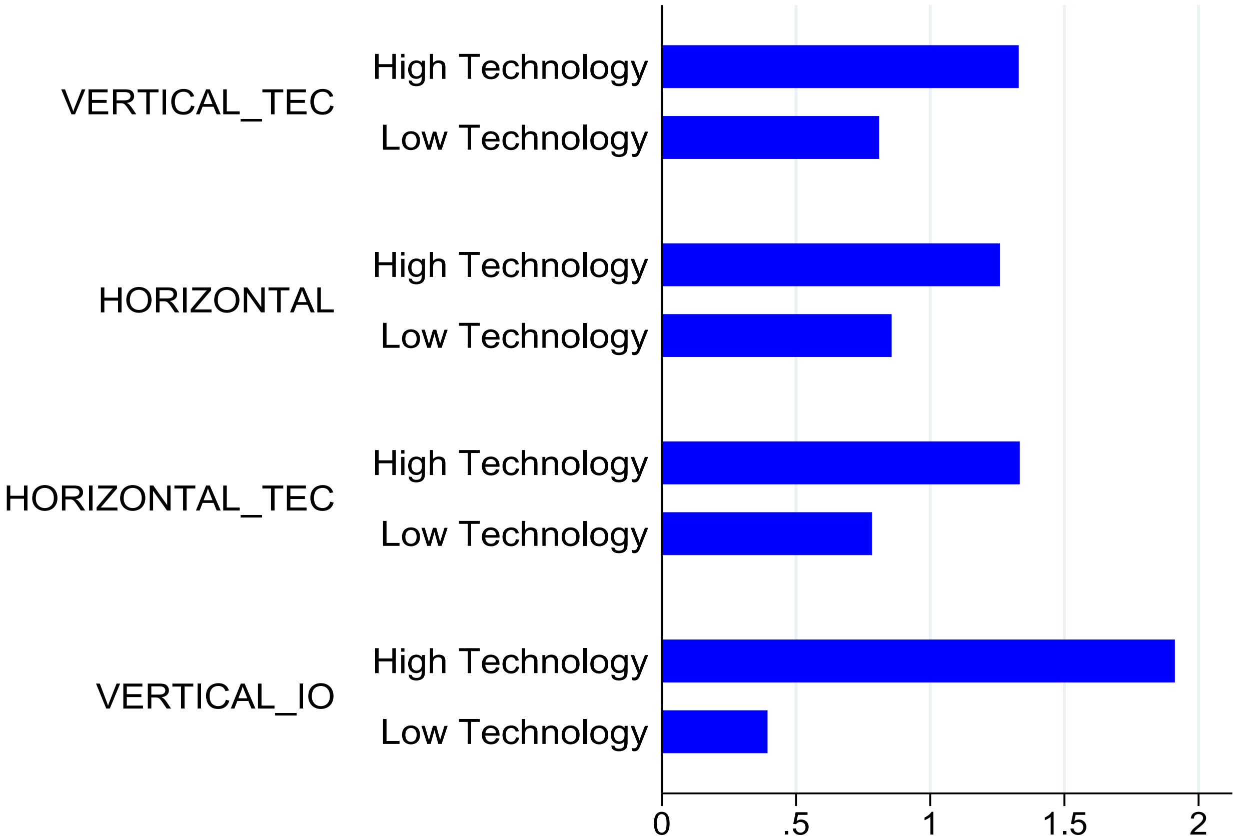
Notes: For the randomly-chosen sector 24, Panel A displays the four-digit technological closeness measure SPILL_RD, and Panel B displays the four-digit input-output coefficients. The input-output coefficients are α-parameters in equation (3) in the main text, and they record the fraction of its own output 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 Section **??** are marked.

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



Notes: The figure displays the frequency distribution of the four-digit technological 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.

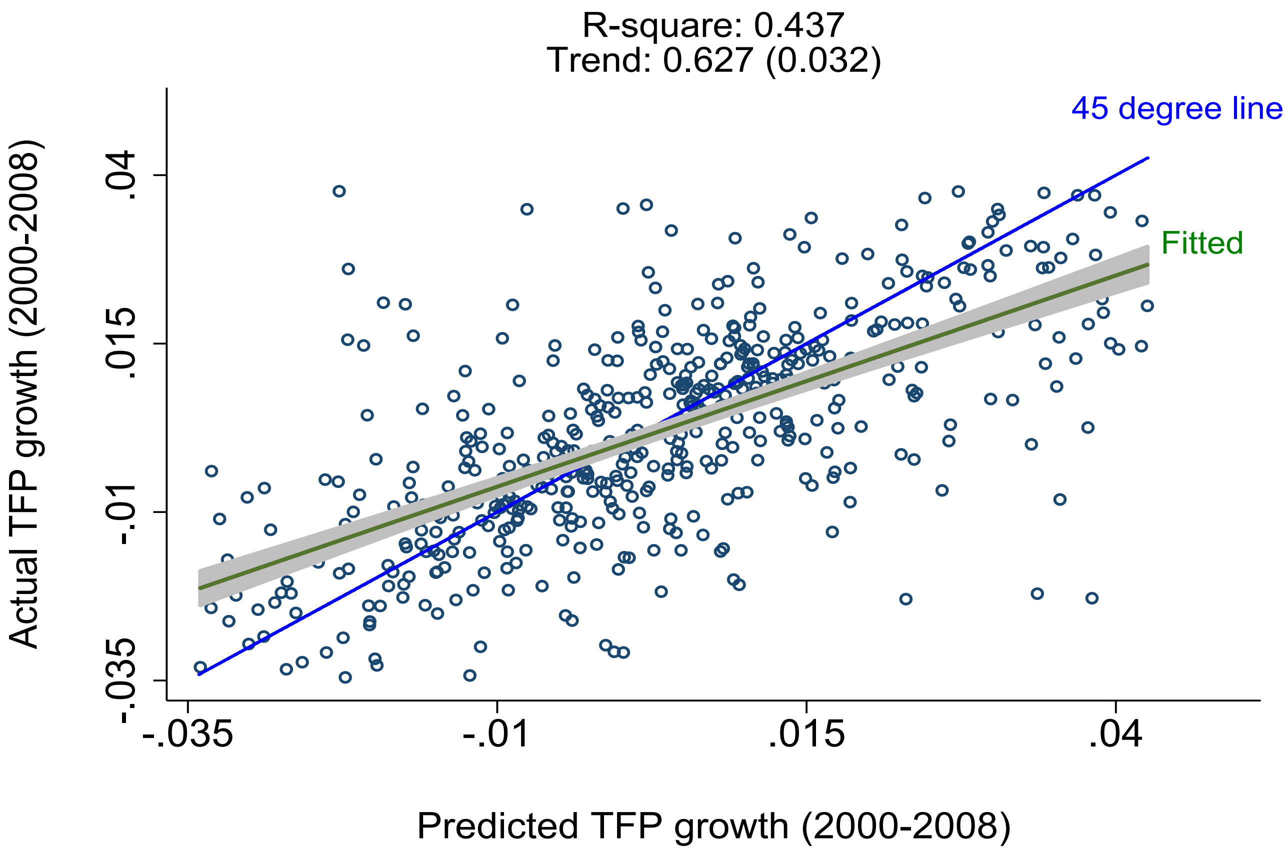
Figure 3: Spillover measures by technology sector in 2007.



.

Notes: The figure displays the average values of the main regressors, defined in Section 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)).

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.

Appendix

A Data

Kalemli-Ozcan, Sorensen, 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, Sorensen, Villegas-Sanchez, Volosovych and Yesiltas (2015) for more details.

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.[27] 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, Sorensen, 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.[28]

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.

[27] "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 (neither includes financial income) and if there are any "fine print" deviations, they are unlikely to be important for our comparison.

[28] The comparison by number of businesses is somewhat noisy because many firms in Eurostat have zero employment (self-employed), 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.

Foreign Ownership

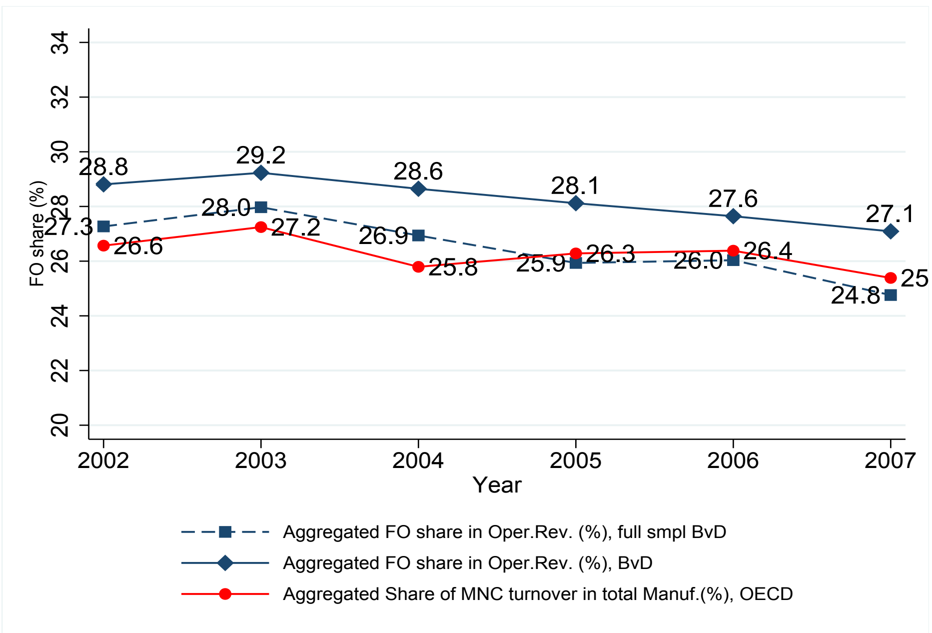
We validate our foreign ownership data by comparing it to the country-level *Activities of Foreign Affiliates* (AFA) database from the Organization for Economic Cooperation and Development (OECD), covering "affiliates under foreign control" (available at http://stats.oecd.org/Index.aspx?DataSetCode=AFA_IN3). 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 "permanent firms" sample).

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 permanent-firms-sample 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 $\frac{\sum_{i\in s4,c} \text{fo}_{i,c,t} \times \text{go}_{i,c,t}}{\sum_{i\in s4} \text{go}_{i,c,t}}$ over firms i, sectors $s4$, and countries c in the balanced (permanent) 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.

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.[29] 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:

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

[29] 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.

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}.

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

$$\omega_{it} = g(k_{it}, m_{it})\,, \tag{B.3}$$

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

$$E(e_{it}|l_{it}, k_{it}, m_{it}) = 0 \qquad t = 1, 2, ..., T\,, \tag{B.4}$$

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

$$\begin{aligned} E(y_{it}|l_{it}, k_{it}, m_{it}) &= \alpha + \beta_l l_{it} + \beta_k k_{it} + g(k_{it}, m_{it}) \\ &\equiv \beta_l l_{it} + h(k_{it}, m_{it})\,, \end{aligned} \tag{B.5}$$

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:

$$E(e_{it}|l_{it}, k_{it}, m_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, ..., l_{i1}, k_{i1}, m_{i1}) = 0 \qquad t = 1, 2, ..., T\,. \tag{B.6}$$

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

$$E(\omega_{it}|\omega_{i,t-1}, ..., \omega_{i1}) = E(\omega_{it}|\omega_{i,t-1}) \qquad t = 2, 3, ..., T, \tag{B.7}$$

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:

$$E(\omega_{it}|k_{it}, l_{i,t-1}, k_{i,t-1}, m_{i,t-1}, ..., l_{i1}, k_{i1}, m_{i1}) \\ = E(\omega_{it}|\omega_{i,t-1}) \equiv f[g(k_{i,t-1}, m_{i,t-1})]\,. \tag{B.8}$$

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:

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + f[g(k_{i,t-1}, m_{i,t-1})] + a_{it} + e_{it}\,. \tag{B.9}$$

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

$$y_{it} = \alpha + \beta_l l_{it} + \beta_k k_{it} + g(k_{i,t}, m_{i,t}) + e_{it} \tag{B.10}$$

and

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

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

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

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.[30]

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

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

C Technology Closeness Based on Citations

We construct a measure of technology closeness between sectors based on patent citations. To construct the measure, we merge the following databases:

1. The USPTO database, which lists the number of patents by cited-citing technology class. Technology classes are defined according to 37 different subcategories (for information on what they are; see page 41 of *http://www.nber.org/papers/w8498.pdf*). 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," and we use data on granted patents between 1976 and 2006. We use two files from the NBER patent database: first, *pat76_06_assg.dta*, which contains information on patent id, granted year, and technology class; second, *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 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:

    - 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: only the past 10 years prior to the starting data of our paper's dataset in 1999.
    - 1987-2006: the past 20 years with available information.

    Because 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 can define this dataset as the "Technological Input-Output Matrix."

2. The second dataset we use is Compustat. The aim is to match Compustat and the USPTO database and construct a sectoral measure which indicates the overlap in technology class citation based on cited and citing sector. 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.

   These are the details of the original Compustat file:

   - Sample: all firms in Compustat. This includes firms that are currently listed or have been listed in the past.
   - Period: the period is 1950 through 2016.
   - Observations: the raw data has 478,570 firm-year observations.

   This is the information once Compustat and USPTO datasets are matched:

   - Sample: Firms currently listed or listed in the past.
   - Period: 1976 through 2006.
   - Observations: 6,325 unique Compustat firms that are associated with a total of 1,124,198 patents.

To construct the measure of technology closeness between sectors based on citations, we follow these steps. First, we take the file including all the citations in the USPTO, which are uniquely identified at the citing patent and cited patent pair level. We then aggregate the data from the patent to the technology class pair level and make it a fully balanced panel. 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 can define this dataset as the "Technological Input-Output Matrix." Then, for each citing technology class, we generate the share of patents cited for each technology class.

Second, we use Compustat firms belonging to different four-digit sectors and identified by the variable "gvkey." Using the "bridge files" available online between Compustat and USPTO, we can make a link between "gvkey" and "pdpass." The variable "pdpass" is the firm identifier in USPTO. Therefore, after this bridge is done successfully, we know the stock of patents (and the technology class of each patent) for each of the Compustat firms. Suppose that firms operating in $SIC = 1$[31] in total have a stock of 1,000 patents, where 70% patents of those patents belong to TechClass=1 and the remaining 30% belong to TechClass=2.

The next step is to understand from which technology classes TechClass=1 and TechClass=2 learn. For illustration, let us focus on TechClass=1. Suppose in the USPTO there are 10,000 patents belonging to TechClass=1. Using the information on citing-cited, we can identify all the citations made by these 10,000 citing patents to previous patents (cited patents). For all these cited patents, we know the technology class. Suppose that 6,000 citations were made to patents of TechClass=1 and the remaining 4,000 citations were made to patents of TechClass=2. We find that TechClass=1 learns 60% from TechClass=1 and 40% from TechClass=2. So, as recap:

- SIC=1 produces 70% in TechClass=1. This technology class learns 60% from TechClass=1 and 40% from TechClass=2.
- SIC=1 produces 30% in TechClass=2. This technology class learns 5% from TechClass=1 and 95% from TechClass=2.
- How much does SIC=1 learn from TechClass=1? 70%*60% + 30%*5% = 42% + 1.5% = 43.5%
- How much does SIC=1 learn from TechClass=2? 70%*40% + 30%*95% = 28% + 28.5% = 56.5%

What remains to be known is: conditional on patenting in TechClass=1, which SIC codes produce these patents? and, conditional on patenting in TechClass=2, which SIC codes produce

[31] Notice each SIC sector corresponds to a four-digit industry code.

these patents?

From the initial bridge between Compustat and USPTO, we found the stock of patents produced by each Compustat firm, and we know both the technology class of the patent and the product market sector of the firm. Therefore, we take all the patents produced in TechClass=1 and ask to which SIC codes belong the firms that have patented those. For example, suppose that 20,000 patents belong to TechClass=1, and we find that 2,000 of these patents are produced by Compustat firms with SIC=1, while the remaining 18,000 patents are produced by Compustat firms with SIC=2. Therefore, conditional on the patents produced by TechClass=1, we find that 10% are from SIC=1 and 90% are from SIC=2. Similarly, conditional on the patents produced by TechClass=2, we find that 50% are from SIC=1 and 50% are from SIC=2.

So, our recap was that SIC=1 learned 43.5% from TechClass=1 and 56.5% from TechClass=2. Therefore,

How much does SIC=1 learn from SIC=1? 43.5%*10% + 56.5%*50% = 4.35% + 28.25% = 32.6%

How much does SIC=1 learn from SIC=2? 43.5%*90% + 56.5%*50% = 39.15% + 28.25% = 67.3%

Our final measure of knowledge flows based on citations between sectors takes these conditional probabilities on how much sectors learn from each other and weight by the $R\&D$ expenditure of the "cited" sector.

Once we have the $R\&D$ weighted technology closeness coefficients ($\text{SPILL\_RD}^{CIT}_{s4,\tilde{s}4}$), we compute the spillover variable following the same steps used for our main variable in the text:

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

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

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
$\log \mathrm{TFPR}$	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

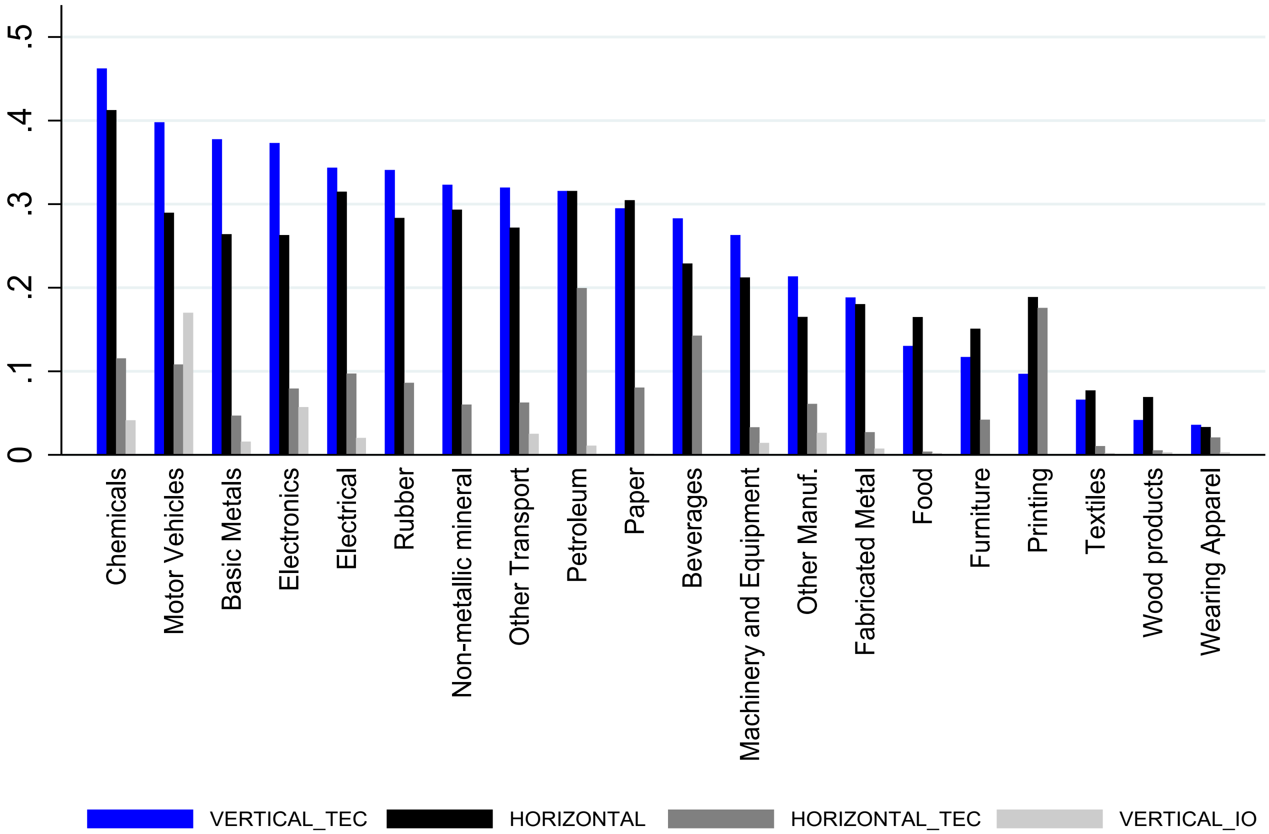
Notes: The table displays sample size, means, and standard deviations for our main variables for the sample of domestic firms used in the regressions. $\log \mathrm{TFPR}$ is the logarithm of revenue total factor productivity; μ is the estimated markup; MC is the estimated marginal cost; #PATENTS is 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

	$\log$ TFPR	$\log \mu$	$\log MC$	HORIZONTAL	HORIZONTAL_TEC	VERTICAL_IO	VERTICAL_TEC	VERTICAL_TEC^{CIT}
$\log$ TFPR	1							
$\log \mu$	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; $\log \mu$ 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;" 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 3. The data in these figures spans all countries in our sample but, unlike in the main paper, is for a single year 2007. We average over countries for that year.