Staff Working Paper/Document de travail du personnel—2023-21
Last updated: April 13, 2023

Trade and Diffusion of Embodied Technology: An Empirical Analysis

by Stephen Ayerst,[1] Faisal Ibrahim,[2] Gaelan MacKenzie[3] and Swapnika Rachapalli[4]

[1] International Monetary Fund
[2] University of Toronto
[3] International Economic Analysis Department, Bank of Canada
GMacKenzie@bankofcanada.ca
[4] University of British Columbia, Sauder School of Business

DOI: https://doi.org/10.34989/swp-2023-21 | **ISSN** 1701-9397

Acknowledgements

We thank our discussant, Anna Ignatenko, and participants at many conferences and seminars for helpful comments on this paper. We have also benefited from feedback from Murat Celik, Kevin Lim, Peter Morrow, Diego Restuccia, and Daniel Trefler. The views expressed herein are those of the authors and should not be attributed to the Bank of Canada or its Governing Council, nor the International Monetary Fund, its Executive Board, or its management.

Abstract

Using data from patents, citations, inter-sectoral sales and customs, we examine the international diffusion of technology through imports of sectoral knowledge and production inputs. We construct measures of the flow of technology embodied in imports. These measures are weighted by inter-sectoral knowledge and production input-output linkages that capture the relevance of this technology for generating new innovations in different sectors in importing countries. We develop an instrumental variable strategy to identify the causal effects of technology embodied in imports on innovation and diffusion outcomes. For sectors in importing countries, increases in both knowledge- and production-weighted embodied technology imports lead to technology diffusion (measured using backward citations in new patent applications) and increases in the rate of new innovations (measured using the forward citations those patents receive). Effects are substantially larger for knowledge-weighted imports of embodied technology, which also lead to improvements in the average quality of new innovations.

Topics: Trade integration; International topics; Productivity; Development economics
JEL codes: O33, F14, O31, O19, F61

Résumé

À l'aide de données tirées des brevets, des citations de brevets, des ventes intersectorielles et des services douaniers, nous examinons la diffusion internationale de la technologie par l'intermédiaire des importations d'intrants sectoriels de connaissances et de production. Nous construisons des mesures du flux de la technologie incorporée aux importations. Ces mesures sont pondérées par les liens entre les intrants et les extrants de connaissances et de production intersectoriels, deux éléments qui servent à comprendre la pertinence de la technologie incorporée pour la création de nouvelles innovations dans différents secteurs dans les pays importateurs. Nous élaborons une stratégie de variables instrumentales pour identifier les effets causals de la technologie incorporée aux importations sur l'innovation et la diffusion. Pour certains secteurs dans les pays importateurs, les hausses des importations des technologies pondérées à la fois par les connaissances et par la production entraînent une diffusion de la technologie (mesurée au moyen des citations antérieures dans les nouvelles demandes de brevets) et des accélérations du rythme des innovations (mesurées à l'aide des citations postérieures reçues par ces brevets). Les effets sont substantiellement plus larges pour les importations de technologie incorporée pondérées par les connaissances, ce qui se traduit aussi par des améliorations dans la qualité moyenne des innovations.

Sujets : Intégration des échanges; Questions internationales; Productivité; Économie du développement
Codes JEL : O33, F14, O31, O19, F61

1 Introduction

Innovation and R&D activity are concentrated in a relatively small number of advanced economies. Recent work demonstrates the quantitative importance of international technology diffusion for the gains from trade and aggregate growth (see, for example, Buera and Oberfield, 2020; Sampson, 2020; Cai et al., 2022). However, little direct empirical evidence exists on the significance of specific channels through which ideas spread across borders. In this paper, we examine the diffusion of technology across countries and sectors through technology embodied in imports of goods from the US using evidence from patents and citations data.

We focus on this channel for three reasons.[1] First, new innovations often manifest themselves as new products or enhancements to existing products, and many of these new or enhanced products are then traded between countries. These product flows potentially convey information about the innovations embodied within them to the users of the products. Second, the foundational knowledge on which new innovations are based originates from many distinct sectors; these sources vary across sectors and need not be related to sectors' sources of production inputs. Since countries' patterns of trade depend in part on patterns of comparative advantage, their imports of technology embodied in trade flows affect innovation in different sectors in those countries in different ways. Incorporating variation in sectors' sources of knowledge and production inputs is necessary to assess the impacts of a given amount of technology embodied in a set of trade flows on different sectors.[2] Third, the effects of trade policies go beyond the well-studied impacts of tariffs on, for example, static intermediate and final goods prices, since they can also affect the flow of information and technology across countries and sectors. Because new innovations often build on existing knowledge, changes in technology flows due to changes in trade policy can have effects on innovation activities not accounted for by the policy-induced responses of innovation to import competition and market access.[3]

The first contribution of our paper is to estimate the extent to which trade is a channel of international technology diffusion by investigating the effects of technology embodied in

[1]Other channels include technology licensing, foreign direct investment, knowledge transfers within multinational firms, immigration, trade in services, and cross-border scientific or technical collaborations (see Keller (2004, 2010, 2021) for surveys of empirical evidence of different channels).

[2]Knowledge inputs are the ideas and technologies that are built upon to generate new innovations while production inputs are the products that are used up in the production of goods and services.

[3]Shu and Steinwender (2019) survey the empirical literature examining evidence of the effects of import competition and market access on innovation. Existing work that, like us, focuses on effects that are present in patents data includes Bloom et al. (2016), Bombardini et al. (2018), Autor et al. (2020) for import competition, and Coelli et al. (2020) and Aghion et al. (2021b) for market access.

imports on innovation and diffusion outcomes. This channel underlies many theoretical and quantitative models of international technology spillovers (e.g., Grossman and Helpman, 1991; Alvarez et al., 2013; Buera and Oberfield, 2020). We start by developing a conceptual framework to guide our empirical analysis. In this framework, the rate at which firms in each sector innovate depends on investment in R&D, domestic technology spillovers, and international spillovers from the technological frontier. International spillovers depend on embodied technology imports—the import-weighted stock of frontier technology in a sector—and the relevance of technology in input sectors for generating new innovations in a sector.

We use patents and citations data to measure innovation and diffusion outcomes. Patents document innovations that result in new products, new components of existing products, or new methods of producing products. Moreover, patent citations provide an explicit account of the sources of technologies upon which new patentable technologies are built that can be used to measure technology diffusion. The second contribution of our paper is to construct a novel dataset on country-sector level innovation activity and technology diffusion. We leverage the Google Patents database to construct detailed patent and citations outcomes for a wide range of countries. The database contains information on the locations of inventors that allows us to construct measures of patenting activity and cross-country citation flows. We also use imports from the Centre d'Études Prospectives et d'Informations Internationales (CEPII) database of international trade flows and cross-sector sales from the Bureau of Economic Analysis (BEA) in our analysis. Finally, we map data into a consistent classification of sectors using a series of concordances.

The third contribution of our paper is to construct measures of cross-sector inputs in embodied technology imports. As a first step, we construct measures of the inter-sectoral relevance of technology inputs. We use inter-sectoral citations and sales data to construct knowledge and production input-output (IO) tables. We construct knowledge IO linkages for a country using the share of citations from each sectors' patents assigned to that country to each other sectors' patents. We similarly construct production IO linkages using the share of sales between sectors in total input purchases. Unlike with patents, data to construct the production IO table at the level of sectoral aggregation we use is only available for the US, which we use as the frontier economy in our analysis. Since knowledge and production IO linkages could, in principle, be similar for many sectors, we demonstrate that the US knowledge and production IO tables are distinct.[4] In particular, we document that knowledge

[4]Though not the focus of our paper, we are among the first to provide a descriptive comparison of the knowledge and production IO tables of an economy. Concurrent work in Hötte (2021) constructs similar knowledge and production IO tables and compares them.

and production IO linkages are not highly correlated on average, that knowledge IO linkages are less concentrated than production IO linkages for the average sector, and that the sectors that are key economy-wide sources of inputs differ between the knowledge and production IO tables.

We combine the knowledge and production IO linkages with data on imports from the US and sector-level US technology stocks that are constructed using quality-adjusted flows of new patents to develop two measures of inputs in embodied technology imports. We refer to these measures as the knowledge-weighted and production-weighted embodied technology imports. Specifically, we aggregate import-weighted US technology stocks using a Cobb-Douglas function where inter-sectoral weights are taken from the knowledge and production IO tables. We exclude own-sector embodied technology imports from these measures in our empirical specifications both to avoid potential endogeneity concerns arising from within-country-sector demand shocks and because own-sector imports can affect innovation activity due to import competition effects. The knowledge-weighted measure is directly related to our channel of interest since it is based on knowledge flows across sectors that are derived from inter-sectoral citations. The production-weighted measure is also included as potentially important transfers of technology can occur through production interactions between sectors. A key outcome of our analysis is to compare the strength of spillovers from knowledge- and production-weighted embodied technology imports.

Our main empirical specification involves regressing measures of innovation and diffusion outcomes on knowledge-weighted and production-weighted embodied technology imports. The main innovation outcomes are patents, forward citations, and forward citations per patent while our main diffusion outcomes are US backward citations, US backward citations per patent, and the share of US backward citations in total backward citations of foreign patents.[5] We also include controls for a measure of the stock of domestic technology inputs for each sector, which uses country-specific knowledge IO linkages, and own-sector imports. Additionally, our long panel of data, spanning from 1995 through 2015, allows us to control for high-dimensional fixed effects. We include country-sector fixed effects to account for persistent differences across sectors in different countries in their patenting outcomes, and country-year fixed effects and aggregate sector-year fixed effects to control for common trends within countries and groups of sectors.

A potential concern with estimating the effects of the trade channel of technology diffusion

[5]Forward citations are measured over the five-year period following the initial filing year of a patent application. US (foreign) backward citations are measured as the cross-sector citations of US (foreign) patents by all patents applied for in a given country-sector-year.

on domestic patenting outcomes is that domestic shocks that affect the outcomes may also lead to changes in demand for US imports. For instance, if domestic R&D and embodied technology in cross-sector imports are substitutes in the production of new innovations, then shocks to domestic R&D productivity in a sector will reduce demand for embodied technology imports and ordinary least squares (OLS) estimates of the effects would suffer from a negative bias. On the other hand, if imported embodied technology and R&D are complements in the production of new innovations, there would be a positive bias in the OLS estimates.[6] To address this concern, we use an instrumental variable (IV) strategy to isolate the effects of embodied technology imports on innovation and diffusion outcomes. For each country, we construct a cluster of related countries that fall into the same quintiles of the distributions of the total trade (exports plus imports) to GDP ratio and GDP per capita. We then construct instruments that replace US imports in a given sector for each country with US exports in the sector to all countries outside of the country's cluster. The IV strategy isolates US supply shocks by excluding countries that are likely to experience correlated demand shocks.

Using our IV strategy, we find that embodied technology imports have positive effects on all three main innovation outcomes. A 1% increase in knowledge-weighted embodied technology imports leads to statistically significant increases of 0.041% for the count of patents, 0.059% for forward citations, and 0.024% for the per-patent rate of forward citations. Estimates of the effects of production-weighted embodied technology imports are smaller: a 1% increase leads to a 0.006% increase in both the count of patents and forward citations, but has no statistically significant effect on the per-patent rate of forward citations. To quantify the magnitude of the impacts, we compute the effects on each outcome variable of a one standard deviation increase in the residual variation of the embodied technology measures that remains after removing the variation explained by the empirical model's other controls and fixed effects. For patent counts, forward citations, and per-patent forward citations, respectively, the effects of knowledge-weighted (production-weighted) embodied technology imports account for 8.1% (1.1%), 6.8% (0.7%), and 1.8% (0%) of one standard deviation of the residual variation in the outcome variable remaining after removing the variation in the outcome explained by the fixed effects. The considerably larger effects of the knowledge-weighted measure is consistent with our expectation that the knowledge IO table better captures the relevance of technology in cross-sector inputs for generating new innovations.

For diffusion outcomes, we find that a 1% increase in knowledge-weighted embodied

[6] Data on R&D spending at the level of industry disaggregation used in our analysis is unavailable for most countries in our sample.

technology imports leads to a 0.081% increase in the flow of backward citations to US patents, while a 1% increase in the production-weighted measure leads to a 0.011% increase. In terms of the effects of a one residual standard deviation increase in these measures, the former effect accounts for 7.0% and the latter effect accounts for 0.9% of a residual standard deviation of the outcome variable. Despite estimating larger elasticities for US backward citations compared to those for the count of patents, we do not find consistent evidence that either measure of embodied technology imports increases the per-patent rate of US backward citations. Similarly, effects on the share of US backward citations in total backward citations of foreign patents are statistically insignificant.[7]

The magnitudes and statistical significance of the coefficient estimates are robust to a variety of alternative specifications. These include specifications with different lags of the explanatory variables relative to the outcome variables; alternative IV strategies, including the traditional leave-one-out approach; alternative transformations of the outcome variables to address observations for which outcomes are zero in level; variations in the data included in the construction of the outcome variables; and different measures of technology stocks. We also find effects that are larger in magnitude when we restrict the sample to the 40 countries with the most patenting activity over our sample period. The robustness of the results reinforces our conclusion that trade is an important channel through which technology diffuses across countries and that this primarily operates through imports of knowledge inputs rather than production inputs.

Related Literature. Our work contributes to the literature on the channels of international technology diffusion (most recently surveyed by Keller, 2021), particularly those papers that examine the trade channel. This includes the pioneering work by Coe and Helpman (1995) and the within-sector analysis of R&D diffusion across borders through both trade and non-trade channels in Acharya and Keller (2009). Our focus on direct evidence for diffusion using citations in new patents is closely related to MacGarvie (2006) and the concurrent and complementary study in Aghion et al. (2021a), both of which use French firm-level data on the extensive margins of trade participation to show that citations to firms' patents increase in foreign markets with which firms interact through trade. We add to this body of evidence by showing with a sector-level analysis that technology embodied in imports of knowledge

[7]Although we find that embodied technology imports from the US lead to increased rates of innovation, this does not necessitate a shift in the cross-country composition of knowledge inputs that are used in the generation of the new innovations. For example, domestic innovators may learn about technologies developed in other foreign countries that are embodied in US imports.

inputs is a source of technology diffusion through the trade of goods.

In doing so, our paper provides evidence for the international technology diffusion that underlies recent growth models with trade, diffusion, and innovation (Buera and Oberfield, 2020; Sampson, 2020; Cai et al., 2022). Most closely related among these is Cai et al. (2022), which allows for inter-sector technology diffusion (both within and across borders). Unlike this work, we provide evidence that imports of embodied technology are a specific channel through which technology diffuses across countries.

The empirical approach we take to evaluate the effects of diffusion of technology across countries is complementary to recent work that uses patents data to measure international technology diffusion through inter-sectoral networks, including Fons-Rosen et al. (2019), Berkes et al. (2022), and Liu and Ma (2022). To the best of our knowledge, ours is the first paper to include inter-sectoral knowledge IO measures based on these data to estimate the trade channel of technology diffusion. Fons-Rosen et al. (2019) use patents-based sector-pair measures of technological similarity adapted from Bloom et al. (2013), which are distinct from our citations-based IO measures, to investigate the foreign direct investment channel of technology diffusion. Berkes et al. (2022) show that there has been a large increase in international knowledge spillovers since the 1990s as measured by cross-country patent citations and that the innovations induced by this increase in diffusion lead to an increase in the growth rates of sectoral output per worker and total factor productivity. Closely related is the empirical exercise in Liu and Ma (2022) that documents that global spillovers from past patenting activity that depend on the network of patent citations across countries and sectors lead to increases in innovation.

Our paper is also related to the branch of the trade literature examining the effects of changes in access to intermediate production inputs due to trade policy on many dimensions of firm performance. This line of research includes work that shows that increased openness to trade of production inputs leads to increases in productivity (Amiti and Konings, 2007; Topalova and Khandelwal, 2011), product scope and new product introduction (Goldberg et al., 2010), and reductions in marginal costs (De Loecker et al., 2016).[8] Though our analysis is conducted at the sector level rather than the firm level and does not focus on the effects of trade policy, our results speak directly to the mechanisms through which trade in inputs leads to improvements in performance and suggest that technology diffusion and increases in the generation of new patented technology follow from increases in embodied technology imports, and particularly so for trade in knowledge inputs.

[8] See also the other relevant works surveyed in Shu and Steinwender (2019).

We also build on work that examines the inter-sectoral patterns of knowledge flows and the implications of these flows in single country settings. Acemoglu et al. (2016) document the patterns of citations across technology classes in US patents and use them to construct innovation IO networks to show how inventions developed in one class spill over to other classes and characterize the degree of localization in the innovation network. Cai and Li (2019) also develop a citations-based IO network and use it to describe patterns in how the direction of firms' innovations evolves along knowledge IO linkages and the aggregate growth implications of these patterns. Our work contributes to this literature by showing how inter-sectoral knowledge IO linkages are important mediators of the diffusion of technology across countries through the trade of goods.

Outline. The remainder of this paper proceeds as follows. Section 2 describes the data used in our analysis. Section 3 presents the conceptual framework used to guide our empirical analysis. Section 4 describes the constructions of the knowledge and production IO tables. Section 5 describes our empirical strategy and baseline specification. Section 6 discusses the estimation results and robustness checks. Section 7 concludes.

2 Data

In this section, we provide an overview of the data used for the main analysis. We use data on patent applications and citations, inter-sectoral purchases of inputs by US sectors, and bilateral product-level trade flows from the US into other countries. These data come from a variety of sources and are provided in a range of distinct classifications that compel us to use concordance tables to translate all the data into a consistent classification system. We briefly describe the data and concordances we use below and leave the remaining details of the data collection and variable construction to Appendix B.

Patents and citations data. We draw on data collected by Google Patents from a wide range of patent offices around the world. For each distinct patent family, which comprises the set of patent applications for a given innovation filed at one or more patent offices, we identify the earliest date a patent was applied for at any patent office and treat this as the filing date for the patent family. Each application in a patent family contains the following information that we use in our analysis: the technology categories to which the innovation is relevant, which are represented by International Patent Classification (IPC) codes; the set of inventors of the patent application and their countries of residence; and citations to other patents listed in the patent application.[9] Throughout our analysis, we focus on patent applications rather than patent grants as grant dates are unavailable in the Google Patents database for patents applied for at many national patent offices, whereas application dates are available.[10] Furthermore, as we examine technology diffusion and its effects, patent application events better reflect the timing of diffusion than do patent grant events.

We calculate the number of initial applications of patent families filed in each year between 1995 and 2015 in each country and technology subclass (a 4-character IPC code) and refer to these as patent counts.[11] Patents are assigned to countries using fractional counts by computing the share of inventors of each patent from each country.[12] For a subset of patent families, applications are submitted to the three patent offices that throughout our sample period are of global significance, including the European Patent Office (EPO), the Japan Patent Office (JPO), and the United States Patent and Trademark Office (USPTO). We count the number of such triadic patent applications.[13]

In addition to counts of patent families, we use information on citations between patents. To measure the quality of patents filed in each year and each country and technology subclass, we compute the number of citations received by these patents across citing patents applied for each year from 1995 to 2021 in all countries and technology classes and define these as the forward citations of the patents in each year. Backward citations data are used for two purposes. First, as described in Section 4.1, we use backward citations to measure knowledge linkages between sectors. Second, for patents filed each year and in each non-US country and technology subclass, we calculate the number of backward citations to US patents, domestic patents, and other foreign patents filed in any technology subclass in each year.

Inter-sectoral input purchases. To measure production input-output relationships, we employ the Bureau of Economic Analysis (BEA) Supplementary Use Tables. These tables are available at five-year intervals and provide the value of purchases by input sector made

[9]We focus our analysis on those patent families with non-missing data for each of these three sets of information. Appendix B explains how we select information on these attributes from among the patent applications in a family.

[10]For instance, there are no grant dates available for patents filed at the Israel Patent Office.

[11]For families with multiple IPC codes, we count these patents once for each technology subclass.

[12]Using information on the countries of the inventors rather than the patent office of the initial application of a patent family allows us to account for innovations developed in one country for which patent protection is first sought in another country. The sample used in our baseline analysis includes data from 82 countries.

[13]We also include patents applied for at the JPO, the USPTO, and the patent offices of France, Germany, and the United Kingdom. These definitions of a triadic patent family are consistent with the methodology described by Dernis (2003).

by US output sectors based on the most up-to-date US industrial classification in use at the time. We use tables that span from 1992 to 2007. Sector classifications are based on US Standard Industrial Classification (SIC) codes for the 1992 Use table, while in more recent vintages they are based on the North American Industry Classification System (NAICS). We describe how we convert the data based on the various SIC and NAICS classifications into a consistent classification in Appendix B. The BEA Use tables not only cover a long period of time, they are available at a high level of disaggregation compared to alternative sources of inter-sectoral sales data. Moreover, using US data enables us to examine how sectors in importing countries are affected by the technology embodied in imports of production inputs from the US based on the patterns of how those inputs are used in the US.

Bilateral trade data. Import data from CEPII's Base pour L'Analyse du Commerce International (BACI) database provide the value of imports of different goods from the US into each country. Our analysis uses annual data from 1995 to 2015. Import values are denominated in current US dollars that we convert to constant 2010 US dollars using CPI deflators taken from the OECD. Goods are classified using 1992 Harmonized System (HS) codes at the 6-digit level of disaggregation.

Concordances between classifications. Because the raw data underlying our analysis are categorized using different classification systems, we employ multiple concordances between these classifications to provide a coherent framework for analysis. We choose the most disaggregated sectors in the 2002 BEA data as our endpoint classification system. This classification, in which sectors are defined similarly to those in the 2002 US 6-digit NAICS classification, allows us to retain a high degree of disaggregation in our analysis while avoiding the potential problems that would arise in a crosswalk of our inter-sectoral input purchase data from the BEA sectors into the more numerous HS goods categories.[14]

We implement a concordance methodology that enables us to first construct measures of technology embodied in goods at the same level of disaggregation as the imports data and second to measure the flow of technology embodied in goods imported from different US sectors. The data downloaded from the Google Patents database are classified into different IPC version 8 4-character technology subclasses.

For the first stage, we convert the data on patent counts, forward citations, stocks of technology (the measurement of which we describe in Section 5.1.3), and backward citations

[14]There are no publicly available sources of data on input-output relationships across goods categorized by disaggregated HS codes. The analysis sample used in our baseline specification includes 292 sectors.

between technology subclasses into categories of goods.[15] To do this, we use the concordance developed by Lybbert and Zolas (2014) between technology subclasses and 2002 6-digit HS codes and then crosswalk this data to 1992 6-digit HS codes. This first concordance is based on an algorithm that uses keywords extracted from the 1992 HS code descriptions that are matched with the text of patent titles and abstracts to construct probabilistic links between the IPC technology subclasses of the matched patents and the HS goods categories.[16]

In the second stage, a series of crosswalks between 1992 HS codes and our endpoint 2002 BEA classification that provide us with weights used to map goods into sectors is overlaid on the technology stocks, patents, citations, and trade data. The crosswalks used are the following: first from 1992 6-digit HS codes to 1987 4-digit Standard Industrial Classification (SIC) codes, second from 1987 4-digit SIC codes to 2002 6-digit NAICS codes, and third from these NAICS codes into the 2002 BEA classification. In applying the first two of these crosswalks, mappings from 1992 HS codes to 2002 NAICS codes use weights derived from the earliest available breakdown of employment by 2002 6-digit NAICS sector from County Business Patterns (CBP) data.[17] Similar procedures that leverage CBP-based employment weights are used to crosswalk the data underlying the different vintages of the BEA Use tables into the 2002 BEA sector categories.

3 Conceptual Framework

Before turning to our empirical analysis, we describe a stylized conceptual framework to guide our analysis. Time is discrete and indexed by t. The economy is populated by a mass of identical firms within each sector of each country. Because firms are identical, we refer to them by their country-sector-year (i, h, t) to simplify notation. We define three levels of sectoral aggregation that align with our data and the empirical approach described in Section 5: denote n as a summary sector (the highest aggregation), h as a sector (the focus

[15]See Appendix B for the procedure we use to calculate citations between technology categories.

[16]Related papers that use the concordances introduced by Lybbert and Zolas (2014) and extended to other classifications in Goldschlag et al. (2020) include Kukharskyy (2020) and Hötte (2021), among others. Kukharskyy (2020) uses the concordances with citations data to construct inter-sectoral knowledge linkages but applies these linkages to investigating how the applicability of multinational parent firms' knowledge capital for a foreign affiliate affects the ownership stake (the degree of integration) of the parent firm in its affiliate. Hötte (2021) also constructs inter-sectoral knowledge linkages and combines them with production linkages to explore how different network characteristics of the knowledge and production IO tables are associated with the level and growth of US sector-level output and patenting.

[17]The details of this procedure and links to the sources of all concordances used in this paper are provided in Appendix B.

of our analysis), and p as a subsector (or product). We also define $\mathcal{P}^h$ as the set of subsectors p in sector h and $n(h)$ as the summary sector n that contains sector h.

Firms in each country produce innovations by investing in R&D, denoted by $R^h_{i,t}$, to earn future profits $\pi^h_{i,t+1}$ per innovation in the following period. Expected profits in period $t+1$ are given by $\mathbb{E}_t[\pi^h_{i,t+1}] = \pi_{i,t} \times \pi^{n(h)}_t \times \pi^h_i \times e^{u^h_{i,t}}$, where $u^h_{i,t}$ is the realization of an independent and identically distributed random variable that is known to firms in period t.[18] To be consistent with our empirical analysis, expected profits depend on a country-year term common to all sectors within a country, $\pi_{i,t}$, a factor that is common to all countries and to all sectors within a summary sector, $\pi^{n(h)}_t$, and a time-invariant country-sector-specific component, π^h_i. A firm (i, h, t) that invests $R^h_{i,t}$ into R&D produces innovations in the next period at rate

$$X^h_{i,t+1} = \left(\frac{R^h_{i,t}}{\psi^h_{i,t}}\right)^{\frac{1}{\zeta}} \left(Z^h_{i,t} S^h_{i,t}\right)^{1-\frac{1}{\zeta}},$$

where $\psi^h_{i,t}$ governs the relative cost of R&D across country-sector-years, $Z^h_{i,t}$ is the amount of domestic technology that is relevant for sector h, and $S^h_{i,t}$ is a spillover from the technology frontier (described below). We assume that the R&D cost parameter is $\psi^h_{i,t} = \psi_{i,t} \times \psi^{n(h)}_t \times \psi^h_i \times e^{v^h_{i,t}}$, where $v^h_{i,t}$ is the realization of an independent and identically distributed random variable that, like $u^h_{i,t}$, is known to firms in period t.[19] While we refer to $X^h_{i,t+1}$ as the rate of innovations, this variable could be relabeled and interpreted as the quality of a given rate of innovations or the quality-adjusted rate of innovations. We explore all three interpretations in the empirical analysis.

Domestic technology input $Z^h_{i,t}$ depends on the stocks of technology in different sectors of the domestic economy and the relevance of those technology stocks as inputs into innovation for firms in the innovating sector h. Domestic technology is given by

$$Z^h_{i,t} = \prod_l \left(G_Z\left(\textstyle\sum_{p\in\mathcal{P}^l} K^p_{i,t}\right)\right)^{\alpha^{l,h}_{i,t}},$$

where $G_Z(\cdot)$ is a monotonic function that dictates the strength of spillovers from domestic technology stocks in an input sector. We set this function to $G_Z(x) = (1+x)^{\eta_Z}$.[20] Domestic

[18]We simplify the environment by assuming that firms only earn profits in the next period, but the qualitative predictions of the model that we test would be equivalent if firms earned a stream of profits where the expected value was proportional to expected profits.

[19]We assume that both $u^h_{i,t}$ and $v^h_{i,t}$ are unobserved by the econometrician in our empirical application.

[20]This specification of $G_Z(x)$ is consistent with our treatment of zeros for technology stocks in the empirical analysis.

technology spillovers from input sector l to sector h depend on the total stock of technology in country-sector-year (i, l, t), $\sum_{p\in\mathcal{P}^l} K^p_{i,t}$, where the sum is taken over the subsectors within sector l, and a measure of the relevance of technology from sector l for producing innovations in sector h in country i in year t denoted by $\alpha^{l,h}_{i,t}$.

Spillovers from the frontier economy depend on the stocks of technology embodied in goods imported from there. The flow of technology coming into sector h depends on a Cobb-Douglas aggregator of spillovers from input sectors l given by

$$S^h_{i,t} = \prod_l \left(G_S\left(\sum_{p\in\mathcal{P}^l} m^p_{F,i,t} \times K^p_{F,t}\right)\right)^{\gamma^{l,h}_{F,t}},$$

where $G_S(\cdot)$ is a monotonic function that dictates the strength of spillovers from the technology embodied in imports of goods from the frontier economy in a sector. We set this function to $G_S(x) = (1 + x)^{\eta_S}$. The parameter $\gamma^{l,h}_{F,t}$ captures the relevance of technology from sector l for innovating in sector h in the frontier economy in period t.

The flow of technology embodied in imports of sector l goods from the frontier depends on the stocks of technology $K^p_{F,t}$ in subsectors $p \in \mathcal{P}^l$ and a measure of the relative abundance of imported goods that embody those technology stocks in country-year (i, t), $m^p_{F,i,t}$. The relative abundance of imported goods is $m^p_{F,i,t} = M^p_{F,i,t}/Y^p_{i,t}$, or the imports of subsector p goods from the frontier economy to the domestic economy $M^p_{F,i,t}$ divided by the subsector's domestic absorption $Y^p_{i,t}$ (output minus net exports). Unlike with spillovers from domestic technology, we scale frontier technology stocks by the relative abundance of frontier goods in the domestic economy so that frontier technology spillovers depend on the extent to which goods that embody frontier technology are available in the domestic economy. Our measure of embodied technology can be thought of as capturing the probability that a domestic innovator encounters a frontier good in sector l and, given the encounter, the technology embodied in that good can be used as an input into producing new innovations in sector h (as in, for example, Bloom et al. (2013), Lucas Jr. and Moll (2014), Perla and Tonetti (2014), and Buera and Oberfield (2020)). Whenever $\gamma^{l,h}_{F,t} > 0$ and $\eta_S > 0$, increased domestic abundance of frontier goods in subsectors $p \in \mathcal{P}^l$ (higher $m^p_{F,i,t}$) and increased technology embodied in those goods (higher $K^p_{F,t}$) both increase domestic innovation in sector h.

We allow both $\alpha^{l,h}_{i,t}$ and $\gamma^{l,h}_{F,t}$ to be time dependent to account for changes in the sourcing of sector l inputs into innovation by sector h over time. Moreover, we assume that the former parameter is specific to the domestic economy while the latter is specific to the frontier economy.

The problem of a firm is to maximize expected profits net of R&D expenditures. The optimal innovation rate for firms in country-sector-year (i, h, t) is determined by

$$X_{i,t+1}^h = \arg\max_X X\pi_{i,t+1}^h - \psi_{i,t}^h X^\zeta \left(Z_{i,t}^h S_{i,t}^h\right)^{1-\zeta},$$

where the second term is the R&D cost paid by the firm for a given innovation rate. Solving the problem implies that firms innovate at rate

$$X_{i,t+1}^h = \tilde{\zeta} \times S_{i,t}^h \times Z_{i,t}^h \times \left[\frac{\pi_{i,t}}{\psi_{i,t}} \times \frac{\pi_t^{n(h)}}{\psi_t^{n(h)}} \times \frac{\pi_i^h}{\psi_i^h} \times e^{u_{i,t}^h - v_{i,t}^h}\right]^{\frac{1}{\zeta-1}}, \tag{1}$$

where $\tilde{\zeta} = \zeta^{-1/(\zeta-1)}$. Taking the log of (1) and grouping variables implies that the log of the innovation rate is

$$\ln X_{i,t+1}^h = \ln S_{i,t}^h + \ln Z_{i,t}^h + f_{i,t} + f_t^{n(h)} + f_i^h + \epsilon_{i,t}^h,$$

where $f_{i,t} = (\ln \pi_{i,t} - \ln \psi_{i,t})/(\zeta - 1)$, $f_t^{n(h)} = (\ln \pi_t^{n(h)} - \ln \psi_t^{n(h)})/(\zeta - 1)$, $f_i^h = (\ln \pi_i^h - \ln \psi_i^h)/(\zeta - 1)$, and $\epsilon_{i,t}^h = (u_{i,t}^h - v_{i,t}^h)/(\zeta - 1)$. Substituting in the spillovers from domestic and imported frontier technology using the assumed functional forms for $G_Z(\cdot)$ and $G_S(\cdot)$ implies that

$$\ln X_{i,t+1}^h = \eta_S \ln EmbTech_{i,t}^h + \eta_Z \ln OwnTech_{i,t}^h + f_{i,t} + f_t^{n(h)} + f_i^h + \epsilon_{i,t}^h, \tag{2}$$

where

$$EmbTech_{i,t}^h = \prod_l \left(1 + \sum\nolimits_{p \in \mathcal{P}^l} m_{F,i,t}^p \times K_{F,t}^p\right)^{\gamma_{F,t}^{l,h}}$$

and

$$OwnTech_{i,t}^h = \prod_l \left(1 + \sum\nolimits_{p \in \mathcal{P}^l} K_{F,t}^p\right)^{\alpha_{F,t}^{l,h}}.$$

The expression in (2) provides the foundation for our empirical strategy, which seeks to identify and estimate the technology spillover elasticity parameters η_S and η_Z. In the next two sections, we describe the construction of the variables underlying $EmbTech_{i,t}^h$ and $OwnTech_{i,t}^h$, which are shorthand labels for frontier and domestic technology spillovers, respectively.

The conceptual framework highlights the relationship between technology embodied in imported goods and innovation outcomes. The assumptions on the nature of expected profits and R&D investment costs are flexible and capture many macroeconomic differences across countries and sectors that may otherwise be of concern in estimating the relationship implied by (2). These include, for example, country-specific business cycles, trends that are common to all countries and to all sectors within a summary sector, and time-invariant differences in the patterns of comparative advantage of countries across different sectors. However, difficulties in estimating the relationship may arise if there are persistent country-sector-specific shocks that drive both an increase in imports from the frontier and domestic innovation. To address these issues, we separate the own-sector and cross-sector effects of imports of embodied technology from the frontier, since we expect these issues to be most prevalent within sectors, and we develop an instrumental variable estimation strategy. These remedies are discussed in detail in Section 5.

4 Inter-Sectoral Technology Relevance

Motivated by the conceptual framework, we begin our empirical analysis by developing measures of the inter-sectoral relevance of technology inputs. For $\alpha_{i,t}^{l,h}$, which parameterizes domestic technology spillovers in country i and year t from sector l to sector h, we base this on patent citation relationships between those sectors. Spillovers from technology embodied in imports from the frontier may depend on both the inter-sectoral knowledge flows captured by these citations as well as patterns of inter-sectoral input purchases, which we measure using production relationships between sectors, so we allow $\gamma_{F,t}^{l,h}$ to depend on both types of relationships. We also highlight key differences between the knowledge and production IO tables that are constructed using these relationships to shed light on how we separately identify the effects of imported embodied technology that operate through these two channels.

4.1 Knowledge and Production Input-Output Linkages

Our analysis estimates the effects of technology embodied in imported goods on patenting outcomes. We focus on two natural candidates to describe the relevance of technology in each sector for generating innovations in other sectors. The first is *knowledge input-output linkages*, which describe the relative flow of patent citations across sectors. This measure is tightly linked with our focus on innovation outcomes since patent citations represent a direct report of the flows of technology that underlie the generation of new innovations for which the protection of rents conferred by patent rights are sought. The second is *production* *input-output linkages*, which describe the relative flow of intermediate inputs across sectors. While less directly linked to innovation outcomes, the use of intermediate inputs captures another channel through which technology can diffuse around the economy. We collect these measures of technology relevance into separate knowledge and production IO tables of the economy and document patterns of inter-sectoral technology flows.

Denote the number of citations of country-sector-year (j, l, s) patents by country-sector-year (i, h, t) patents as $Cites^{l,h}_{j,i,s,t}$. This variable captures the reported flow of technology, or knowledge, from (j, l, s) to (i, h, t).[21] We denote the set of sectors by $\mathcal{H}$ and the set of countries by $\mathcal{I}$.

Knowledge IO linkages, which measure the relevance of knowledge produced in each input (cited) sector for each output (citing) sector, are constructed using the backward citations made by patents. More specifically, let $\alpha^{l,h}_{i,t}$ denote the knowledge IO linkage between sectors l and h for patents filed in country i in year t. We allow for this measure to change over time and base the relationship in year t on patents filed between years $t - \bar{\tau}$ and t for some chosen lag $\bar{\tau}$. The knowledge IO linkage is given by

$$\alpha^{l,h}_{i,t} = \frac{\sum_{j\in\mathcal{I}} \sum_{\tau=0}^{\bar{\tau}} \sum_{s\leq t-\tau} Cites^{l,h}_{j,i,s,t-\tau}}{\sum_{k\in\mathcal{H}} \sum_{j\in\mathcal{I}} \sum_{\tau=0}^{\bar{\tau}} \sum_{s\leq t-\tau} Cites^{k,h}_{j,i,s,t-\tau}}. \tag{3}$$

In our analysis, we set the maximum lag used in the construction of the knowledge IO linkages to a ten-year window ($\bar{\tau} = 9$) to account for slow-moving technological transitions.[22] The knowledge IO linkages capture the country-sector (i, h) citations made by patents filed over a ten-year window to all prior sector l patents from all countries as a share of total citations made by country-sector (i, h) patents filed over the ten-year window.

Similarly, we measure production IO linkages as the importance of goods produced in each input sector as intermediate inputs into production in each output sector. Because the availability of highly disaggregated data on inter-sectoral sales is comparatively limited, we focus on within-country transactions in the US. We define $\beta^{l,h}_{i,t}$ as the analog to $\alpha^{l,h}_{i,t}$ for the

[21] Similarly to the allocation of patents to countries, we weight each citation by the product of the cited and citing patents' fractional country weights based on their respective inventor country compositions. In this notation, each year refers to the filing year of the relevant patent families.

[22] For example, Berkes et al. (2022) find relatively gradual structural transformation in key patenting sectors over a 100-year period. Similarly, Baslandze (2018) and Ayerst (2022) find that ICT diffusion affected patterns of patent citations over this period, which highlights the need for dynamic knowledge IO linkages.

production IO table. The production IO linkage is given by

$$\beta_{i,t}^{l,h} = \frac{Sales_{i,t}^{l,h}}{\sum_{k \in \mathcal{H}} Sales_{i,t}^{k,h}}, \tag{4}$$

where $Sales_{i,t}^{l,h}$ is the total value of sector l goods sold to sector h in country i and year t. Production IO linkages measure, for year t, the share of sales from sector l to sector h in the total sales from all sectors to sector h.

In our analysis, the production IO linkages are based on the US data from the BEA Use tables that was described in Section 2. Since the BEA Use tables are only available at five-year intervals, we use the linkages constructed from the data in each table for multiple years. For consistency with the measurement of production IO linkages, we also use only knowledge IO linkages from the same years for which there is a BEA Use table. In addition, to allow sectoral variation in exposure to domestic and imported frontier technology inputs to be determined in advance of exposure in a given year, we use IO linkages that are lagged relative to the years in which exposure is measured. This lag in exposure variation is applied to both knowledge and production IO linkages.[23]

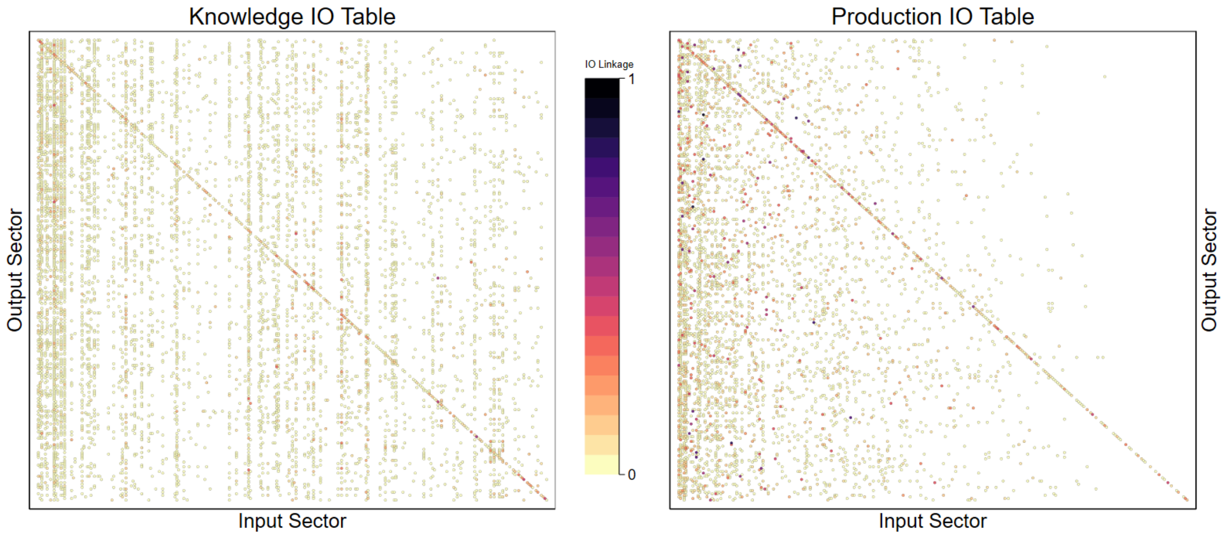
4.2 Description of Knowledge and Production IO Tables

The construction of the knowledge and production IO tables relies on different data. However, there is little point in examining the effects of imports of embodied technology in knowledge and production inputs separately if the two IO tables are identical. We now turn to illustrating some stylized observations regarding the two IO tables to demonstrate that they are different potential measures of inter-sectoral technology relevance.

Both knowledge and production IO linkages take on values between zero and one. Values closer to one indicate stronger relationships between sectors whereas values closer to zero indicate weaker relationships. In Figure 1, we depict the knowledge and production IO tables for the US economy in 2002, with values of $\alpha_{US,2002}^{l,h}$ represented in the left panel and $\beta_{US,2002}^{l,h}$ in the right panel. In each table, rows correspond to input sector l and columns correspond to output sector h. The color of each cell depends on the size of the IO linkage between the input and output sectors. We plot only those IO linkages for which the input sector accounts for at least 1% of the inputs used by the output sector. We also sort sectors in the IO tables

[23] To be more precise, we use $\alpha_{i,1992}^{l,h}$ and $\beta_{i,1992}^{l,h}$ for exposure measured between 1995 and 2000, $\alpha_{i,1997}^{l,h}$ and $\beta_{i,1997}^{l,h}$ between 2001 and 2005, $\alpha_{i,2002}^{l,h}$ and $\beta_{i,2002}^{l,h}$ between 2006 and 2010, and $\alpha_{i,2007}^{l,h}$ and $\beta_{i,2007}^{l,h}$ between 2011 and 2015.

Figure 1: Input-Output Tables



Notes: Figure displays the knowledge and production IO tables where each point represents an IO linkage. The row position of each output sector and column position of each input sector are held constant across both IO tables to facilitate visual comparisons across tables. Sectors are sorted based on their economy-wide importance as suppliers of production inputs by summing up the production IO linkages of each input sector over off-diagonal output sectors. The plots include the 291 2002 BEA sectors in agriculture, forestry, fishing and hunting, manufacturing, and mining with a non-zero sum of knowledge IO linkages across input sectors. Knowledge (production) IO linkages are defined in Equation (3) (Equation (4)). Knowledge IO linkages are based on backward citations of patents assigned to the US filed between 1993–2002 while production IO linkages are based on the 2002 BEA Use table. Both plots only display IO linkages that account for at least 1% of the inputs used by an output sector while all other IO linkages are visually suppressed.

based on their relative importance as sources of production inputs across output sectors to visually highlight the differences in the IO tables.

An immediate insight one can draw from Figure 1 is that there are clear differences in the patterns of knowledge and production IO linkages for many sectors. We formalize this visual intuition through three observations that highlight the differences between the knowledge and production IO tables.[24]

Observation 1: *The sources of knowledge and production inputs are not highly correlated for the average sector.*

Observation 2: *The sources of production inputs are more highly concentrated than the sources of knowledge inputs for the average sector.*

[24]One can also clearly see that own-sector IO linkages along the diagonal are, in general, large relative to off-diagonal IO linkages in both the knowledge and production IO tables. We discuss the importance of own-sector versus cross-sector (off-diagonal) linkages both for the presentation of these observations in Appendix A and for our empirical results in Section 6.

Observation 3: *The key input-supplying sectors are distinct in the knowledge and production IO tables.*

We relegate the documentation and further discussion of these observations to Appendix A as a comparison of the IO tables is tangential to our main objectives. That said, a key implication of the observations is that the knowledge and production IO structures of the economy capture different relationships between sectors and, consequently, may capture different potential sources of technology spillovers. Given this, in our baseline analysis we explore the diffusion of technology through imports of embodied technology weighted in two ways: using knowledge IO linkages and using production IO linkages.

5 Empirical Specification

In this section, we describe the main empirical specification used in our analysis and the construction of its key variables. We start by specifying our main outcome variables and then use the knowledge and production IO tables described in the previous section to develop the main explanatory variables. Finally, we outline the empirical counterpart of Equation (2) and an instrumental variable (IV) approach that we use to identify the effects of spillovers from technology embodied in imports.

5.1 Variable Construction

We now describe the main outcome and input variables in our analysis. Throughout the analysis, we focus on the effects of imports from the US. We make this assumption for two main reasons. First, the US is both the most innovative country and the largest originator of cross-country citations over the time period we analyze (see Berkes et al. (2022) for evidence). In this regard, the US best captures what we think of as the frontier economy. Second, because the US has data available to construct our measures of inter-sectoral knowledge and production IO linkages that are consistently defined across time, choosing the US to be the frontier economy ensures these two measures are comparable with each other.[25]

[25]Measuring production IO linkages for other countries at the level of sectoral disaggregation available for the US and across our sample period is not possible due to data limitations.

5.1.1 Sample

The unit of observation in the analysis is a country-sector-year. We limit our final panel of data to the years 1995 to 2015. We restrict ourselves to this time span because in earlier years there is a lack of detailed trade data for many countries and including later years would cause truncation issues for patents and forward citations, which are the main data used to measure innovation outcomes.

We also limit the set of countries in our final sample based on the following criteria. First, we drop countries if they have no triadic patents in any sector in any of the 21 years of the panel. Second, we drop those that had a population of less than one million in 1995 to avoid the inclusion of countries where patenting outcomes may be too noisy. Third, we drop those countries that have exports to GDP or imports to GDP ratios in 2015 above the 98th percentile or below the 2nd percentile of those statistics among the remaining set of countries. Fourth, we drop countries that have imports to GDP or exports to GDP ratios in 2015 that are larger than one. These previous two conditions exclude from our sample countries that trade for reasons unrelated to production or consumption, such as countries that primarily act as trade intermediaries. Finally, we keep only those countries that are above the 25th percentile of total triadic patents across all years among the remaining countries, which corresponds to a cutoff of just under ten triadic patents over the sample. This restriction excludes countries where patented innovations are either infrequent or of relatively low quality from a global perspective. We restrict based on triadic patents because it is a measure of quality that is unrelated to citations, which may be influenced by country-specific factors. Additionally, while the 25th percentile is a restrictive cutoff, many countries report zero or near zero triadic patents. Including these countries would tend to bias our estimates downwards, since it would increase instances of zero or near-zero patenting in a country-sector-year, and would generate noise in our outcomes.[26]

5.1.2 Outcome Variables

We divide our outcome variables into innovation outcomes and diffusion outcomes. For both groups, our baseline results include three sets of outcome variables.

Innovation Outcomes. The conceptual framework highlights the relationship between frontier technology spillovers and innovation outcomes. There, the focus is on the rate of

[26] Consistent with this expectation, we find substantially larger coefficient estimates for our main specification when we restrict our sample to the top 40 countries in terms of total patents across all sample years.

innovation as the main outcome variable, which we measure in the data using both the rate of patenting ($Patents$) and the citation-weighted rate of patenting ($FwdCites$). We also use the average quality of patents, measured by the number of forward citations per patent ($FwdRate$), to compare the intensive and extensive margin effects of frontier technology spillovers on innovation outcomes. We discuss the construction of each variable below.

1. **Patent Counts (*Patents*).** Our first variable, $Patents^h_{i,t}$, is the count of patent applications in country-sector-year (i, h, t). We take this measure from the Google Patents database following the allocation rules described in Section 2.

2. **Forward Citations (*FwdCites*).** For patents filed by country-sector-year (i, h, t), the number of forward citations received from patents applied for in the five years following the filing year of the cited patents can be computed as

$$FwdCites^h_{i,t} = \sum\nolimits_{l\in\mathcal{H}} \sum\nolimits_{j\in\mathcal{I}} \sum\nolimits_{s=t}^{t+5} Cites^{h,l}_{i,j,t,s}. \tag{5}$$

   We focus on forward citations received in the first five years of a patent's life to mitigate truncation issues that would arise in later periods of the sample if citations received in any year were used instead.[27] We interpret forward citation-weighted patent counts as a measure of quality-adjusted patenting activity. We do not take a stance on whether the expected effect of technology embodied in imports on quality-adjusted patenting should be larger or smaller than the effect on the count of patents. The effect may be larger for quality-adjusted patenting if high-quality innovators benefit more from technology spillovers and push low-quality innovators out of the market. Conversely, it may be smaller if domestic producers use frivolous patent filings to protect their market share or to extract rents from foreign entrants.

3. **Forward Citation Rate (*FwdRate*).** For patents filed by country-sector-year (i, h, t), the rate of forward citations received per patent application is

$$FwdRate^h_{i,t} = \frac{FwdCites^h_{i,t}}{Patents^h_{i,t}}. \tag{6}$$

[27]To simplify the construction of our data, we focus on the five-year period measured using the calendar year in which a patent is applied for. For example, for a patent filed in June 2000 we include forward citations received up to December 31, 2005, in measuring $FwdCites$ for the year 2000 for the pertinent country-sector. We do not expect this choice to affect our results since our unit of measurement is a year.

The forward citation rate is a measure of the average quality of patent applications and is used to capture the intensive margin response to a shock to imported frontier technology. As the denominator of this measure will be zero for (i, h, t) observations with no patent applications, we must take a stance on the treatment of such zeros. We exclude these observations from the estimation sample.

Diffusion Outcomes. Patents and citations data also provide direct evidence on the extent to which trade of embodied technology is a source of technology diffusion and leads to higher flows of knowledge from the US. We use the backward citation information underlying the knowledge IO table as a measure of cross-country knowledge flows. We construct three outcome variables that measure the cross-sector flow of backward citations to US patents ($USBackCites$), the per-patent rate of cross-sector backward citations to US patents ($USBackRate$), and the share of backward citations to US patents in the total cross-sector backward citations to foreign patents ($USBackShare$). The measures are defined below.

1. **US Backward Citations (*USBackCites*).** The number of backward citations made by patents in country-sector-year (i, h, t) to US patents filed in any year up to and including year t in sectors other than sector h is

$$USBackCites_{i,t}^{h} = \sum_{l \neq h} \sum_{s \leq t} Cites_{US,i,s,t}^{l,h}. \tag{7}$$

   We exclude the own-sector backward citations from this outcome variable to be consistent with our focus on cross-sector imports of embodied technology as described below.[28] That is, since our main explanatory variable of interest is spillovers from cross-sector imports of embodied technology, we do not want to include own-sector citations in measuring diffusion directly.

2. **US Backward Citations Rate (*USBackRate*).** For patents filed by country-sector-year (i, h, t), the average number of cross-sector backward citations to US patents per patent application is

$$USBackRate_{i,t}^{h} = \frac{USBackCites_{i,t}^{h}}{Patents_{i,t}^{h}}.$$

[28] We look at the entire history of backward citations since backward citations do not suffer from the truncation issues in more recent years, as is the case with forward citations.

Similarly to the forward citation rate, we think of the backward citation rate as a measure of the intensive margin of technology diffusion. Whereas the first diffusion outcome measures the total number of cross-sector knowledge inputs that flow from the US to sector h in country i, the second measures the intensity with which the typical sector h patent uses those cross-sector US knowledge inputs.

3. **Backward Citation Share (*USBackShare*).** Our final outcome variable is the share of cross-sector foreign backward citations that are made to US patents by patents filed in (i, h, t). Specifically, we construct the US backward citation share as

$$USBackShare_{i,t}^{h} = \frac{USBackCites_{i,t}^{h}}{\sum_{l \neq h} \sum_{j \neq i} \sum_{s \leq t} Cites_{i,j,s,t}^{l,h}}.$$

   Relative to the other two outcomes, the backward citation share informs us on whether knowledge inputs are substituted towards technology patented in the US in response to larger embodied technology flows from the US. It is also possible that sectors in importing countries cite more non-US foreign patents in response to those flows as they learn from those patents as well as the US patents underlying our measure of imported embodied technology. This would lead to estimates of the effects of imports of embodied technology on this outcome to be small relative to the estimates of effects on $USBackCites$.

Summary Statistics. In our baseline specification described below, we measure outcome variables using the average of each variable over the three-year window between year t and $t + 2$. Table 1 presents summary statistics of the main outcome variables used in that specification. The counts of observations for $FwdRate$, $USBackRate$, and $USBackShare$ are smaller than for the other outcomes. For $FwdRate$ and $USBackRate$, the denominators of these rates ($Patents$) are zero for some country-sector-years, while for $USBackShare$, some country-sector-years have no cross-sector citations to foreign patents. The distributions of $Patents$, $FwdCites$, and $USBackCites$ are highly skewed, with the median country-sector-year having values of these variables close to zero. By contrast, the distributions of $FwdRate$, $USBackRate$, and $USBackShare$ show little skewness.

Table 1: Summary Statistics for Outcome Variables

	N	Median	Mean	SD
$Patents^h_{i,t}$	478,880	0.066	0.742	1.345
$FwdCites^h_{i,t}$	478,880	0.140	1.159	1.832
$FwdRate^h_{i,t}$	361,290	1.430	1.434	0.746
$USBackCites^h_{i,t}$	478,880	0.295	1.478	2.089
$USBackRate^h_{i,t}$	361,290	2.128	2.065	0.990
$USBackShare^h_{i,t}$	356,457	0.498	0.486	0.197

Notes: All outcome variables are averaged over the three-year window t to $t+2$. Except for $USBackShare$, all statistics are calculated on the log of one plus the outcome variable. SD is the standard deviation.

5.1.3 Embodied Technology Imports and Other Controls

Our main variable of interest is the spillovers from imported embodied frontier technology ($S^h_{i,t}$ in the conceptual framework). Here, we also describe the construction of country-subsector-level technology stocks ($K^p_{i,t}$), which are important determinants of technology spillovers, and the domestic stock of technology ($Z^h_{i,t}$), which is used as a control in our main specification.

Technology Stocks ($\boldsymbol{K^p_{i,t}}$). Before turning to our main variables of interest, we discuss the construction of technology stocks $K^p_{i,t}$ since this is used as an input for the main variables. We measure the technological content of a subsector's goods using patents data and, following Hall et al. (2001), use citation weights to adjust for the relative quality of patents. Specifically, we use forward citations in the first five years after a patent's application as our preferred measure of patent quality since this avoids issues with truncation of citations in later periods of the data and imposes minimal structure in constructing a comparable measure of quality across countries, subsectors, and years. Technology stocks are given by

$$K^p_{i,t} = (1-\delta)K^p_{i,t-1} + FwdCites^p_{i,t}, \tag{8}$$

where $FwdCites^p_{i,t}$ is the count of five-year forwards citations received by patents filed in country-subsector-year (i,p,t) and δ is the depreciation rate of technology that we set to 5% to be consistent with commonly used values. For each country and subsector, we initialize the stock of technology in 1940 with the value $K^p_{i,1940} = FwdCites^p_{i,1940}/\delta$. The initial value has little influence on the technology stocks used in the period of our analysis since it is measured over 50 years prior to the start of this period.

Domestic Technology Stock (*OwnTech*). The stock of domestic technology that is relevant as an input into innovation in sector h, $Z^h_{i,t}$, enters into the expression for innovation in Equation (2) because it captures domestic spillovers.[29] The empirical counterpart of the domestic technology input is

$$OwnTech^h_{i,t} = \prod_l \left(1 + \sum_{p \in \mathcal{P}^l} w^l(p) K^p_{i,t}\right)^{\alpha^{l,h}_{i,t}}, \tag{9}$$

where $w^l(p)$ is the concordance weight of subsector (product) p in sector l discussed in Section 2. Higher values of $OwnTech$ reflect higher capability in country-sector-year (i, h, t) in generating new technologies based on past domestic innovation activity in sectors that its patents have a tendency to cite. We use the domestic knowledge IO linkage $\alpha^{l,h}_{i,t}$ in the construction of domestic technology inputs because it provides the best available measure of the relevance of sector l technology for innovation in sector h in country i and year t.[30]

Embodied Technology (*EmbTech*). Our main variable of interest reflects the amount of technology embodied in imported goods that is relevant as an input into innovation in importing country-sectors. We do not impose structure on whether technology flows between sectors are better captured by the knowledge or production IO linkages, so we set the linkages $\gamma^{l,h}_{i,t}$ in the conceptual framework to be a combination of knowledge ($\alpha^{l,h}_{i,t}$) and production ($\beta^{l,h}_{i,t}$) IO linkages. In doing so, our analysis is informative of the relative importance of knowledge and production linkages for innovation and diffusion outcomes.

Following the conceptual framework, we measure the frontier technology spillover in two steps. We first construct the imported embodied technology flow in each subsector-year (p, t) as the product of the US technology stock $K^p_{US,t}$ and US imports $M^p_{US,i,t}$. Then, we aggregate across subsectors using concordance weights and use the upstream knowledge and production IO linkages to weight the resulting imported embodied technology flows in each sector-year.

For each country-sector-year (i, h, t), our measure of knowledge-weighted embodied technology imports is given by

$$EmbTechK^h_{i,t} = \prod_{l \neq h} \left(1 + \sum_{p \in \mathcal{P}^l} w^l(p) K^p_{US,t} M^p_{US,i,t}\right)^{\alpha^{l,h}_{US,t}} \tag{10}$$

[29] This variable can also be thought of as capturing past spillovers from technology imported from the frontier to the extent that they were built upon by domestic innovators in the past and thus enter into the domestic technology stocks measured in year t.

[30] We do not have consistent measures of production IO linkages for most countries and years, which prevents us from constructing a comparable production IO linkage $\beta^{l,h}_{i,t}$.

Table 2: Summary Statistics for Embodied Technology Imports

	N	Median	Mean	SD
$\ln(EmbTechK_{i,t}^{h})$	478,880	16.057	15.718	3.243
$\ln(EmbTechP_{i,t}^{h})$	478,880	13.129	12.604	4.088

Note: SD is the standard deviation.

and production-weighted embodied technology imports is given by

$$EmbTechP_{i,t}^{h} = \prod\nolimits_{l \neq h} \left(1 + \sum\nolimits_{p \in \mathcal{P}^l} w^l(p) K_{US,t}^{p} M_{US,i,t}^{p}\right)^{\beta_{US,t}^{l,h}}. \tag{11}$$

The amount of imported embodied technology depends on the flow of technology into country i from the US in every sector l. This flow is increasing in the volume of imports and the stocks of technology in subsectors within sector l. Countries that spend more on sector l goods from the United States have a higher flow of technology into them from that sector. For example, a larger volume of imports could reflect more varieties of a sector's goods being imported. Our measure reflects the idea that as a country imports more, ideas upon which domestic innovators can build become more abundant and readily available. The effect of a flow of technology from a given sector l is weighted by the tendency of that sector's technology to be used in sector h in the US. Table 2 provides summary statistics of the knowledge- and production-weighted measures of embodied technology imports.

We construct the measures of embodied technology from the value of US imports ($M_{US,i,t}^{p}$), rather than imports scaled by absorption as in the conceptual framework ($m_{US,i,t}^{p} = M_{US,i,t}^{p}/Y_{i,t}^{p}$), due to limitations on the availability of output data at the level of aggregation we examine. A potential issue with this approach is that higher imports could simply reflect that the importing country has a larger population or economy.[31] We include granular fixed effects as a best attempt to deal with this. An alternative would be to use US import shares (i.e., US imports to country i over all imports to country i for each subsector). However, this could lead to misleading conclusions because trends in trade patterns during the period of our analysis (i.e., all countries have increased trade flows over this period) have led to declining US import shares for many subsectors and countries.

We omit the own-sector component in the embodied technology spillover terms as within-sector imports and innovation outcomes can potentially be related to each other for multiple

[31]This is not a concern for the US technology stocks $K_{US,t}^{p}$ since they are common to all importers and capture the abundance of technology embodied within imports, meaning their levels are important for the interpretation of our results.

reasons. Within-sector demand shocks can lead to countries importing more foreign products to satiate demand while at the same time investing more in innovation activities in the sector due to increased returns to innovation. Moreover, own-sector imports can also affect innovation outcomes in a country through import competition effects, since firms may invest more in innovation in order to escape foreign competition. Finally, R&D productivity shocks and profitability shocks to a country-sector can lead to comovements of imports and innovations in the country-sector. We discuss endogeneity concerns further in Section 5.3.

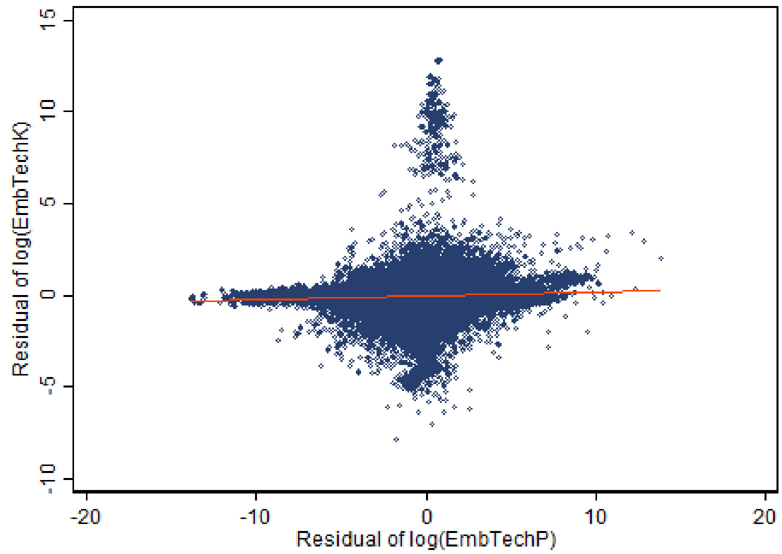
Identifying and estimating the separate effects of knowledge- and production-weighted imports of embodied technology requires that there is sufficient variation across observations in our sample in these two measures. To assess this, we regress the logs of both $EmbTechK$ and $EmbTechP$ on the set of fixed effects included in the baseline specification discussed in Section 5.2. In Figure 2, we plot the fitted residuals from these regressions on top of which we overlay a line of best fit from a regression of one set of residuals on the other. This figure demonstrates that for much of the support of either of the residualized input measures, there is considerable variation in the other residualized measure. The R-squared of the overlayed regression is 0.0039, while the correlation of the two residualized input measures is 0.062. The substantial amount of variation in our embodied knowledge and production input measures gives us confidence that our results provide a comparison of the importance of imported knowledge and production inputs from a frontier economy on our patenting outcomes.

Own-Sector Embodied Technology (*EmbTechDiag*). The Input-Output tables developed in Section 4 show that own-sector inputs (the diagonals of the IO tables) tend to be important in both the knowledge and production IO tables. Given that imports of own-sector embodied technology inputs are a likely source of technology diffusion, we also include them as a control in our empirical specification. Specifically, we construct own-sector embodied technology as

$$EmbTechDiag^h_{i,t} = 1 + \sum\nolimits_{p \in \mathcal{P}^h} w^h(p) K^p_{US,t} M^p_{US,i,t}. \tag{12}$$

We do not scale these technology inputs by the IO weights $\alpha^{h,h}_{US,t}$ or $\beta^{h,h}_{US,t}$ since, as mentioned above, we expect that this variable captures factors not directly related to the effects of technology diffusion, such as import competition. For example, Bloom et al. (2016) find that increased trade with China between 2000 and 2005 led to an increase in patenting activity in European firms that were more exposed to that competition (which was also the case for

Figure 2: Residualized Embodied Technology in Imports



Notes: Figure plots residuals of $\log(EmbTechP)$ and $\log(EmbTechK)$ and the line of best fit from the regression of the latter measure on the former. Residuals are computed by regressing each measure on the set of fixed effects included in the baseline specification discussed in Section 5.

increased exposure to trade from other low-wage countries).[32]

5.2 Estimation Equation

We now present the empirical counterpart of Equation (2) in terms of our constructed variables that serves as our baseline specification:

$$\begin{aligned}\ln(1 + Outcome^h_{i,t}) = {} & \theta_1 \ln EmbTechK^h_{i,t-\tau} + \theta_2 \ln EmbTechP^h_{i,t-\tau} \\ & + \theta_3 \ln OwnTech^h_{i,t-\tau} + \theta_4 \ln EmbTechDiag^h_{i,t-\tau} \\ & + V^h_{i,t}\rho + f_{i,t} + f^{n(h)}_t + f^h_i + \epsilon^h_{i,t}, \end{aligned} \tag{13}$$

where $V^h_{i,t}$ is a vector of controls that includes own-sector imports from the world and exports to the world and $f_{i,t}$, $f^{n(h)}_t$, and f^h_i are country-year, (summary) sector-year, and country-sector fixed effects. In the baseline regressions we average outcomes over a three-year window

[32]However, they find that changes in import penetration of high-wage countries like the US had no effect on patenting. In contrast to Bloom et al. (2016), Autor et al. (2020) find that import competition due to increased trade with China decreased patenting activity in publicly listed US firms and technology classes more exposed to that competition. We do not estimate effects of import competition from low-wage countries such as China in this paper.

from t to $t+2$ to reduce year-to-year noise in the outcome variables and to allow for a more gradual diffusion of technology. With the exception of $USBackShare$, we also transform each outcome variable as $\ln(1+Outcome^h_{i,t})$ to retain observations that have innovation or diffusion outcomes that are zero in level in our sample. We show that our results are robust to other transformations in Appendix C. Our baseline estimates use a lag of $\tau = 1$ years for the explanatory variables. We also present results for the model where outcomes are for period t only and the input variables are measured at lags $\tau \in \{1, 2, \ldots, 5\}$ to examine the dynamic response of each outcome to changes in exposure to technology embodied in imported goods.

In all regressions, we allow for the possibility that the residuals are correlated across years within a country-sector pair (due to serial correlation) and across countries in each year within a sector (since much of the variation in our variables of interest is at the sector-year level). To do so, we estimate multi-way clustered standard errors at the country-sector and sector-year levels (Cameron et al. (2011)).

The conceptual framework in Section 3 offers some predictions on the expected sign of the coefficients in Equation (13). Our hypothesis is that $\eta_S > 0$ and $\eta_Z > 0$ such that higher levels of relevant technology inputs that are either embodied in imported goods from the US or present in domestic technology stocks due to past domestic innovations lead to higher levels of innovation outcomes. Consequently, we expect that θ_1, θ_2, and θ_3 are positive for $Patents$ and $FwdCites$. Although imported own-sector inputs captured in $EmbTechDiag$ should have a positive effect on innovation rates, this measure also incorporates import competition effects that may reduce innovation activity, so the expected sign of θ_4 for these innovation outcomes is ambiguous. The stylized model provides less guidance on the anticipated effects of our variables of interest on the average quality of new innovations captured by $FwdRate$ or on the diffusion outcomes. As discussed above, effects of each variable on $FwdRate$ could be positive or negative. Furthermore, although we expect that the signs of the coefficients are the same for $Patents$ and $USBackCites$, since we evaluate the responses of innovation and diffusion to shocks to embodied technology imported from the US, the framework is silent on how the per-patent rate of backward citations to US patents, $USBackRate$, and the share of citations of foreign patents that go to US patents, $USBackShare$, should respond.

5.3 Endogeneity Concerns

The fixed effects in Equation (13) control for time-invariant characteristics of country-sector pairs, factors that vary at the country level over time, and sector-year shocks that are common to sectors within a summary sector. Despite the inclusion of these fixed effects, there remain potential endogeneity concerns with our regressors of interest.

One possibility is that variation across country-sector-years in the amount of relevant technology inputs embodied in a country's imports in prior years could reflect demand shocks for those inputs that also directly affect patenting outcomes. For example, shocks to expected profits, captured by $u^h_{i,t}$ in the conceptual framework, would both increase R&D investment but also the imports of intermediate inputs used in the production of goods in (i, h, t).[33] If these shocks were serially correlated, there would be a spurious positive correlation between innovation output and imports of embodied technology in past years arising from the profitability shocks. Since there is no data available on R&D expenditures at the level of sectoral disaggregation used in our analysis, we cannot control for these innovation inputs which may cause an omitted variable bias to affect our estimates.

To address these concerns, we use an instrumental variable strategy that focuses on variation in US imports that is a function of supply shocks to US exports. For each country-sector-year, we instrument each regressor that includes US imports with a variable that uses US exports to all countries outside of a cluster of similar countries to which that country is assigned (discussed below). For example, we instrument the knowledge- and production-weighted imports of embodied technology as

$$IVEmbTechK^h_{i,t} = \prod\nolimits_{l \neq h} \left(1 + \sum\nolimits_{p \in \mathcal{P}^l} w^l(p) K^p_{US,t} \left(\sum\nolimits_{j \notin \mathcal{G}_i} M^p_{US,j,t}\right)\right)^{\alpha^{l,h}_{US,t}} \tag{14}$$

and

$$IVEmbTechP^h_{i,t} = \prod\nolimits_{l \neq h} \left(1 + \sum\nolimits_{p \in \mathcal{P}^l} w^l(p) K^p_{US,t} \left(\sum\nolimits_{j \notin \mathcal{G}_i} M^p_{US,j,t}\right)\right)^{\beta^{l,h}_{US,t}}, \tag{15}$$

respectively, where $\mathcal{G}_i$ is a cluster of countries with similar characteristics to country i.[34]

For each country i, the cluster $\mathcal{G}_i$ is the set of countries that fall into the same quintiles of both GDP per capita and the ratio of total trade (imports plus exports) to GDP as country i.[35] We group countries based on GDP per capita to capture similarities in technological development across countries and the ratio of total trade to GDP to capture similarities in trade patterns across countries.

Our instrumental variable strategy isolates variation in subsector-level trade flows to each

[33]We do not explicitly model demand for production inputs from different sectors and instead implicitly subsume them into the expected profit function.

[34]The instrument for own-sector imports of embodied technology is $IVEmbTechDiag^h_{i,t} = 1 + \sum_{p \in \mathcal{P}^h} w^h(p) K^p_{US,t} (\sum_{j \notin \mathcal{G}_i} M^p_{US,j,t})$.

[35]These quintiles are computed over all countries included in the BACI database using data for 1995.

country that stems from shocks to US export supply in a subsector. A standard leave-one-out instrument would exclude only the domestic economy to discount changes in trade that result from domestic demand shocks in the subsector. We extend this intuition by not only excluding the domestic economy but also countries that share similar characteristics and, consequently, may face demand shocks that are correlated with those facing the domestic country.

6 Results

In this section, we discuss the results from estimating the effects of knowledge- and production-weighted imports of embodied technology on the innovation and diffusion outcomes. We begin by discussing estimates using the baseline specification described in Equation (13). Then, we show that our main conclusions hold using different lags of the explanatory variables. We also use the empirical model to provide a quantification of the magnitude of the estimated effects. This section closes with a discussion of the robustness of the results.

6.1 Baseline Results

Table 3 presents the estimates for the innovation outcomes using ordinary least squares (OLS) in columns (1) to (3) and two-stage least squares (2SLS), where the IV approach described in Section 5.3 is applied, in columns (4) to (6).

The OLS results suggest that knowledge-weighted imports of embodied technology $EmbTechK$ have a positive impact on the innovation rate as measured by both $Patents$ and $FwdCites$. Despite there being a larger point estimate for $FwdCites$ than for $Patents$ and a positive point estimate for $FwdRate$, the OLS results do not point to a statistically significant effect of $EmbTechK$ on $FwdRate$. Qualitatively, production-weighted embodied technology imports $EmbTechP$ have similar effects on the innovation rate measures as $EmbTechK$. However, the estimated elasticities are substantially smaller, which suggests that spillovers from technology embodied in imports from the US primarily operate through knowledge IO linkages rather than production IO linkages. We show in Section 6.3 that these differences are quantitatively important for our innovation outcomes after accounting for differences in the variation in $EmbTechK$ and $EmbTechP$ in our sample.

The 2SLS estimates of the effects of $EmbTechK$ are larger than the OLS results and statistically significant for each innovation rate measure. For $Patents$ and $FwdCites$, the 2SLS coefficient estimates are around twice as large as the OLS estimates for the knowledge-weighted embodied technology imports. The coefficient estimates are also larger for production-weighted

Table 3: Innovation Outcomes

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
	Patents	*FwdCites*	*FwdRate*	*Patents*	*FwdCites*	*FwdRate*
ln($EmbTechK$)	0.018***	0.027***	0.007	0.041***	0.059***	0.024***
	(0.003)	(0.004)	(0.005)	(0.008)	(0.011)	(0.009)
ln($EmbTechP$)	0.004***	0.004***	0.000	0.006***	0.006***	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
ln($OwnTech$)	0.011***	0.031***	-0.011***	0.011***	0.030***	-0.011***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
ln($EmbTechDiag$)	0.000	-0.000	0.002	0.061***	0.078***	0.009
	(0.000)	(0.001)	(0.001)	(0.010)	(0.014)	(0.011)
Observations	478,880	478,880	361,290	478,880	478,880	361,290
F-Stats						
EmbTechK				4,468	4,468	10,890
EmbTechP				15,030	15,030	20,754
EmbTechDiag				397	397	484

Notes: Table reports coefficient estimates for innovation outcomes using Equation (13). All dependent variables are first averaged over a three-year window and then transformed using $\ln(1 + Outcome)$, where $Outcome$ is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the log of total exports to the world and log of total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats in columns (4) to (6) are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

embodied technology imports. Comparing the effects of $EmbTechK$ and $EmbTechP$, the coefficient estimates remain larger for $EmbTechK$, which is consistent with our expectation that patent citations better reflect the patterns of inter-sectoral relevance of technology inputs for generating new innovations. That being said, the estimated effects of $EmbTechP$ are positive and statistically significant for $Patents$ and $FwdCites$, which indicates that technology diffusion also operates through inter-sectoral production relationships.

The remaining 2SLS estimates are in line with our expectations. The coefficients on $OwnTech$ are positive and significant for $Patents$ and $FwdCites$, which suggests that domestic technology in input sectors (which reflects past domestic innovation activity) contributes positively to the generation of new innovations. The effects of $EmbTechDiag$ are also positive and statistically significant for these two outcomes, and relatively large compared to those of the other explanatory variables. Although we do not focus on the own-sector effects of technology embodied in imports due to the aforementioned difficulties with interpreting the mechanisms underlying these effects, the estimates imply that this channel has large impacts on innovation rates and suggest that the spillover effects captured

Table 4: Diffusion Outcomes

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
	$USBackCites$	$USBackRate$	$USBackShare$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln(EmbTechK)$	0.034***	0.006	0.000	0.081***	0.013	-0.000
	(0.006)	(0.006)	(0.001)	(0.016)	(0.010)	(0.002)
$\ln(EmbTechP)$	0.007***	0.002**	0.001*	0.011***	0.001	0.000
	(0.002)	(0.001)	(0.000)	(0.003)	(0.001)	(0.000)
$\ln(OwnTech)$	0.045***	-0.014***	-0.000	0.044***	-0.014***	-0.000
	(0.003)	(0.003)	(0.001)	(0.003)	(0.003)	(0.001)
$\ln(EmbTechDiag)$	0.000	0.003**	0.001**	0.127***	0.037***	0.008***
	(0.001)	(0.001)	(0.000)	(0.021)	(0.013)	(0.003)
Observations	478,880	361,290	356,457	478,880	361,290	356,457
F-Stats						
$EmbTechK$				4,468	10,890	11,433
$EmbTechP$				15,030	20,754	21,441
$EmbTechDiag$				397	484	473

Notes: Table reports coefficient estimates for diffusion outcomes using Equation (13). All dependent variables are averaged over a three-year window and then, except for $USBackShare$, transformed using $\ln(1 + Outcome)$, where $Outcome$ is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the log of total exports to the world and log of total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats in columns (4) to (6) are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

in the estimated coefficients on $EmbTechK$ and $EmbTechP$ may understate the total effects of diffusion of technology arising from imports of goods.

Table 4 presents the main results for the diffusion outcomes for both the OLS and 2SLS estimates. Since the signs of the 2SLS estimates are the same as those of the OLS estimates, we discuss the two sets of results together. The coefficient estimates for $USBackCites$ are positive for both knowledge- and production-weighted embodied technology imports. This may reflect in part the results in column (4) of Table 3, since higher imports of embodied technology from the US lead to higher rates of new patent applications and, holding fixed the inventor country and sectoral composition of the backward citations in those new applications, this would lead to higher rates of cross-sector backward citations to patents with US inventors. Moreover, the estimates in column (4) of both Tables 3 and 4 suggest that increased imports of embodied technology from the US lead to an increased intensity of cross-sector citations to US patents.

However, when the sample is restricted to country-sector-year observations with strictly positive $Patents$ over the three-year period between t and $t + 2$, as is the case in column (5) of Table 4 for $USBackRate$ as an outcome variable, we find statistically insignificant 2SLS estimates of the effects of both knowledge- and production-weighted imports of embodied technology on this measure of the intensity with which cross-sector US knowledge inputs are used in the generation of new innovations. Similarly, we find little evidence of an impact of either measure of embodied technology inputs in imports on $USBackShare$ in column (6). While the framework in Section 3 has no implications for how these two outcomes should respond, these results suggest that although diffusion of embodied technology from the US leads to an increase in the rate of innovation, it has little to no impact on the intensity with which that technology is used in new innovations to the extent that this is captured by patent citations as we measure them.[36] One possible reason for this is that our measure of backward citations includes citations to patents filed in any past year. While per-patent citations to more recently generated innovations may increase in response to higher imports of embodied technology, citations to older technology may decrease and this composition effect may lead to an overall estimate that is zero on average. Another possibility, particularly for the effects on $USBackShare$, is that domestic innovators learn from and cite both technology developed by US-based inventors as well as innovations embodied in imported products that were developed in non-US foreign countries.

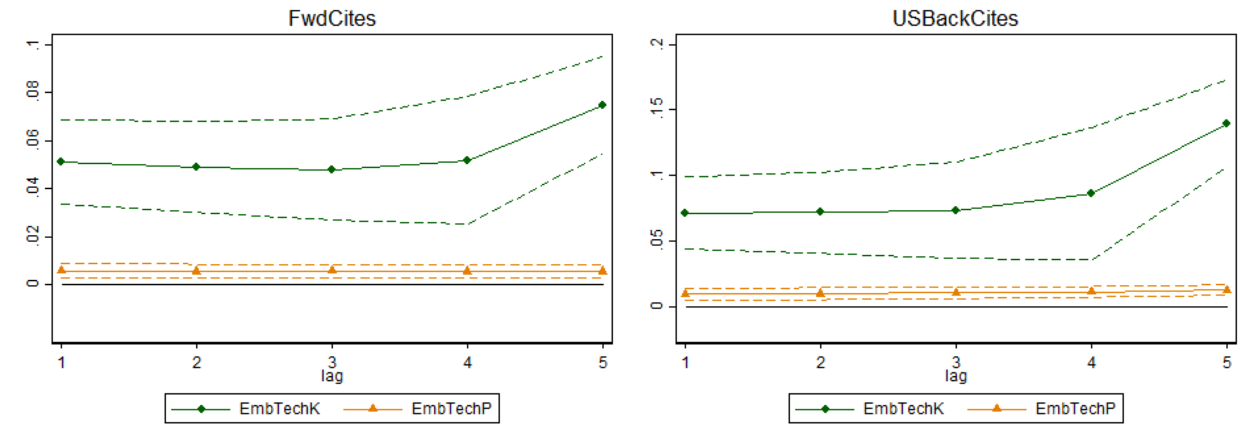
6.2 Alternative Lagged Effects

The results in Tables 3 and 4 average the outcome variables over a three-year window, in part to allow for the gradual diffusion of technology. Technology diffusion is a gradual process, and it may take several years before spillovers from imports of embodied technology are reflected in patentable innovations. To examine these diffusion dynamics, we separately estimate Equation (13) using 2SLS with the outcome variable measured at period t (rather than a three-year window) and the explanatory variables entering with lags $\tau \in \{1, 2, \dots, 5\}$. Figure 3 summarizes the coefficient estimates for the knowledge- and production-weighted measures of embodied technology imports. In all regressions the set of controls are the same as in the baseline regressions. We present the results for $FwdCites$ and $USBackCites$ here and display estimates for the remaining outcomes in Appendix Figure C.1.

The results at different lags are similar in magnitude to the baseline results. At longer lags, the effects of $EmbTechK$ are stronger while they are relatively stable across all five models

[36] As discussed in Section 6.4, the estimated effect of knowledge-weighted imports of embodied technology inputs on $USBackRate$ is positive and statistically significant at conventional levels in some alternative specifications, which is consistent with the relative size of the effects in column (4) of both Tables 3 and 4.

Figure 3: Main Outcomes Estimated at Different Lags



Notes: Figure plots 2SLS coefficient estimates of the effects of $EmbTechK$ and $EmbTechP$ on $FwdCites$ (left panel) and $USBackCites$ (right panel) for five separate models using Equation (13) with the outcome variable measured in period t. The five models include the explanatory variables lagged by τ years relative to year t for each of $\tau \in \{1, 2, \ldots, 5\}$. The dashed lines are the 95% confidence intervals of the effects. All models include the same controls and fixed effects as in the baseline specification.

for $EmbTechP$. The comparisons between the baseline results and the alternative lagged effects for both the magnitudes of the effects and their statistical significance take on similar patterns for the other innovation and diffusion outcomes. We view this as supportive evidence for averaging the outcome variables over a three-year window in the baseline results since these alternative results point to diffusion of imported embodied technology having gradual impacts on the outcomes. Consistent with this view, we find larger, albeit quantitatively similar, point estimates for the effects of $EmbTechK$ and $EmbTechP$ if instead the outcomes are averaged over a five-year window (see Appendix Table C.1).

6.3 Quantitative Significance

To characterize the quantitative magnitude of the results, we compare the relative amount of variation in the outcome variables and the implied variation of the outcomes attributable to the key model variables. We focus on the residualized standard deviation of the explanatory variables, RSDE, which is calculated as the standard deviation of each variable after removing the estimated effect of all other regressors as well as the fixed effects used in the baseline specification. The residual variation in the outcome variables, RSDO, is the standard deviation of each outcome variable after removing the estimated fixed effects. We net out the estimated fixed effects from both types of variables to remove variable trends as well as cross-country and cross-summary-sector variation in the variables. These differences are important for

Table 5: Quantitative Significance

Outcome	RSDO	Coefficient Estimate: *EmbTechK*	Coefficient Estimate: *EmbTechP*	Relative Implied RSD (%): *EmbTechK* (RSDE = 0.368)	Relative Implied RSD (%): *EmbTechP* (RSDE = 1.201)
Patents	0.187	0.041	0.006	8.1	1.1
FwdCites	0.317	0.059	0.006	6.8	0.7
FwdRate	0.481	0.024	0	1.8	0
USBackCites	0.425	0.081	0.011	7.0	0.9
USBackRate	0.606	0.013	0.002	0.7	0.1
USBackShare	0.142	-0.001	0	-0.2	0

Notes: RSDO is calculated as the standard deviation of the outcome variable after controlling for the fixed effects used in the baseline specification (Equation (13)). RSDE is calculated as the standard deviation of the explanatory variable after controlling for the other regressors and fixed effects used in that specification. For each of *EmbTechK* and *EmbTechP*, relative implied RSD refers to the product of the coefficient estimate and the the RSDE divided by the RSDO. Coefficient estimates are taken from Tables 3 and 4.

both the outcomes (e.g., increases in patenting over time) and embodied technology imports (e.g., increases in trade flows over time). However, the fixed effects are not important for understanding the economic significance of the key variables.[37] We estimate the effect of the RSDE of the explanatory variables implied by the model and scale them by the RSDO of the outcome variables. The results are summarized in Table 5.

The table shows that a one RSDE increase of *EmbTechK* would generate an increase measured as a percentage of one RSDO of 8.1% for *Patents*, 6.8% for *FwdCites*, 1.8% for *FwdRate*, and 7.0% for *USBackCites*. Consistent with the earlier discussion of the measures of embodied technology imports, there is more residualized variation in *EmbTechP*, which increases its relative quantitative importance, but this gap is not large enough to offset the differences in coefficient estimates found in Table 3 and Table 4. Therefore, the overall impact of production-weighted embodied technology imports is marginal compared with knowledge-weighted embodied technology imports.

6.4 Robustness of Results

We conclude this section with a discussion of the robustness of the main results to alternative specifications. Throughout, we focus on estimates of the effects of knowledge- and production-weighted embodied technology imports *EmbTechK* and *EmbTechP*. Overall, estimates are similar to those found using the baseline specification. Results are provided in Appendix C.

[37] For example, the inability of the empirical model to explain a secular trend in patenting over time is not informative for understanding the importance of embodied technology imports.

Alternative Instruments. Our baseline 2SLS results use instrumental variables that isolate US supply shocks by using variation in US exports to all countries outside of a country's cluster of similar countries. We also construct alternative instruments using both the traditional leave-one-out instrument, which can be viewed as having a single country in each cluster, and an instrument using US exports to all other countries within a country's cluster of similar countries (Appendix Tables C.2 and C.3, respectively). In both cases, the results are similar in sign, magnitude, and statistical significance to the baseline results. In the latter case, the coefficient on $EmbTechK$ for $USBackRate$ is also positive and statistically significant.

Transformation of Outcome Variables. In the baseline specification, we transform the outcome variables using $\ln(1 + Outcome^h_{i,t})$ to avoid excluding observations for which $Outcome^h_{i,t} = 0$. We find similar results using both the $\ln(Outcome^h_{i,t})$, which excludes zeros, and $\text{arsinh}(Outcome^h_{i,t})$ transformations (see Appendix Tables C.4 and C.5, respectively).[38] In the former case, we also find that the coefficient estimate for $USBackRate$ is positive and statistically significant. As an additional check, we estimate Equation (13) using the baseline transformation but drop the observations for which $Patents^h_{i,t} = 0$ for all years between t and $t+2$ (Appendix Table C.6). Estimates of the effects of $EmbTechK$ on $Patents$, $FwdCites$, and $USBackCites$ are larger in magnitude than in the baseline, but the effects of $EmbTechP$ are essentially unchanged.

Technology Stocks. The country-subsector-year-level technology stocks $K^p_{i,t}$ defined in Equation (8) use the flow of five-year forward citations received by new patent applications, and these serve as the basis for our measures of domestic technology stocks and technology embodied in imports from the US. This approach controls for the relative quality of different patents in the measurement of technology stocks. We show in Appendix Table C.7 that the results are robust to an alternative approach that uses the flow of new patent applications (unadjusted for quality) to measure technology stocks.

Country Sample. We also consider an alternative specification wherein we restrict the sample to observations from the top 40 countries based on the total number of patents applied for across 1995–2015. The point estimates of the effects of $EmbTechK$ on $Patents$, $FwdCites$, $FwdRate$, and $USBackCites$ are larger with the more restricted sample, while

[38] The second transformation is $\text{arsinh}(Outcome) = \ln(Outcome + \sqrt{Outcome^2 + 1})$.

the effects of $EmbTechP$ are more or less unchanged. That the effects of $EmbTechK$ are larger could suggest that technology diffusion from imports of knowledge inputs generates more innovations in countries that are already relatively innovative in the sense that inventors located there file more and higher-quality patents. Alternatively, the effects of imports of embodied technology in less innovative countries may be smaller because the innovative activity that is generated by those embodied technology imports is not well captured by patent applications.

Other Results and Controls. We find similar results when we examine effects of embodied technology imports on various alternative outcomes (see Appendix Tables C.9 and C.10). For example, when we restrict the sample of patents used to construct the outcome variables to triadic patents (though still including forward citations received from any patent and cross-sector backwards citations made to any patent) we estimate effects of $EmbTechK$ on $Patents$, $FwdCites$, and $USBackCites$ that are about half the size of the baseline estimates, though effects on $FwdRate$ are essentially unchanged. The effects of $EmbTechP$ on these first three outcomes are also smaller but remain statistically significant. These estimates reinforce the finding that diffusion of technology through imports leads to increases in the rate and average quality of new innovations. As an alternative measure of patent quality, we examine effects on the forward citations received from patents assigned to foreign countries and find results that are unchanged from those in the baseline. Lastly, we include both cross-sector as well as own-sector backward citations in our diffusion outcome variables. When these additional citations are included, increased technology embodied in imports of knowledge inputs $EmbTechK$, which remains constructed using only cross-sector knowledge inputs, leads to a statistically significant increase in the total per-patent rate of backward citations to US patents.

7 Conclusion

Innovation activities are highly concentrated in a small number of countries, but new technology eventually diffuses to other countries. One potentially important channel through which technology diffuses across borders is international trade of goods, since importers can learn about the technology embodied in those goods. This paper assesses the extent to which knowledge and production inputs in traded goods contribute to the diffusion of technology and to the amount and quality of innovations developed in importing country-sector pairs.

To do this, knowledge and production IO tables are constructed using data on inter-sectoral patent citations and sales. These measures of the relevance of goods from different input sectors as inputs into the creation of new innovations and the production of goods in different output sectors are combined with measures of the stocks of technology embodied within sectors' products and data on product-level trade flows between countries to construct measures of knowledge- and production-weighted technology embodied in cross-sector imports. We show that increases in both measures of technology embodied in imported goods lead to higher rates of innovation in an importing country-sector pair and that effects of knowledge-weighted embodied technology imports are substantially larger than those of production-weighted embodied technology imports.

Our results point to important directions for future research. For example, the estimated elasticities in this paper could be used to discipline a quantitative model of cross-country and cross-sector technology diffusion through trade. This would allow for an evaluation of the aggregate growth and welfare implications of accounting for this channel of diffusion and adding it to the potential effects of trade policy on innovation.

References

Acemoglu, Daron, Ufuk Akcigit, and William R. Kerr (2016) "Innovation Network," *Proceedings of the National Academy of Sciences*, 113 (41), 11483–11488.

Acharya, Ram C. and Wolfgang Keller (2009) "Technology Transfer through Imports," *Canadian Journal of Economics*, 42 (4), 1411–1448.

Aghion, Philippe, Antonin Bergeaud, Timothee Gigout, Matthieu Lequien, and Marc Melitz (2021a) "Exporting Ideas: Knowledge Flows from Expanding Trade in Goods," Working Paper.

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, and Marc J. Melitz (2021b) "The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports," NBER Working Paper 24600, National Bureau of Economic Research.

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas Jr. (2013) "Idea Flows, Economic Growth, and Trade," NBER Working Paper 19667, National Bureau of Economic Research.

Amiti, Mary and Jozef Konings (2007) "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, 97 (5), 1611–1638.

Autor, David, David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu (2020) "Foreign Competition and Domestic Innovation: Evidence from US Patents," *American Economic Review: Insights*, 2 (3), 357–374.

Ayerst, Stephen (2022) "The Diffusion of New General Purpose Technologies," Working Paper.

Baslandze, Salomé (2018) "The Role of the IT Revolution in Knowledge Diffusion, Innovation and Reallocation," Working Paper.

Berkes, Enrico, Kristina Manysheva, and Martí Mestieri (2022) "Global Innovation Spillovers and Productivity: Evidence from 100 Years of World Patent Data," Working Paper.

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

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

Bombardini, Matilde, Bingjing Li, and Ruoying Wang (2018) "Import Competition and Innovation: Evidence from China," Working Paper.

Buera, Francisco J. and Ezra Oberfield (2020) "The Global Diffusion of Ideas," *Econometrica*, 88 (1), 83–114.

Cai, Jie and Nan Li (2019) "Growth Through Inter-sectoral Knowledge Linkages," *The Review of Economic Studies*, 86 (5), 1827–1866.

Cai, Jie, Nan Li, and Ana Maria Santacreu (2022) "Knowledge Diffusion, Trade and Innovation Across Countries and Sectors," *American Economic Journal: Macroeconomics*, 14 (1), 104–145.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011) "Robust Inference with Multiway Clustering," *Journal of Business & Economic Statistics*, 29 (2), 238–249.

Coe, David T. and Elhanan Helpman (1995) "International R&D Spillovers," *European Economic Review*, 39 (5), 859–887.

Coelli, Federica, Andreas Moxnes, and Karen Helene Ulltveit-Moe (2020) "Better, Faster, Stronger: Global Innovation and Trade Liberalization," *The Review of Economics and Statistics*, 1–42.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik (2016) "Prices, Markups, and Trade Reform," *Econometrica*, 84 (2), 445–510.

Dernis, Hélène (2003) "*OECD Triadic Patent Families - OECD methodology: an overview*," `https://www.wipo.int/export/sites/www/meetings/en/2003/statistics_workshop/presentation/statistics_workshop_dernis.pdf`.

Fons-Rosen, Christian, Şebnem Kalemli-Özcan, Bent E. Sørensen, Carolina Villegas-Sanchez, and Vadym Volosovych (2019) "Foreign Investment and Domestic Productivity: Identifying Knowledge Spillovers and Competition Effects," Working Paper.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova (2010) "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India," *The Quarterly Journal of Economics*, 125 (4), 1727–1767.

Goldschlag, Nathan, Travis J. Lybbert, and Nikolas J. Zolas (2020) "Tracking the Technological Composition of Industries with Algorithmic Patent Concordances," *Economics of Innovation and New Technology*, 29 (6), 582–602.

Grossman, Gene M. and Elhanan Helpman (1991) "Trade, Knowledge Spillovers, and Growth," *European Economic Review*, 35 (2-3), 517–526.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg (2001) "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498, National Bureau of Economic Research.

Hötte, Kerstin (2021) "Demand-Pull and Technology-Push: What Drives the Direction of Technological Change?", Working Paper.

IFI CLAIMS Patent Services and Google (2022) "*Google Patents Public Data*," `https://console.cloud.google.com/marketplace/product/google_patents_public_datasets/google-patents-public-data/` [Accessed: March 12, 2022].

Keller, Wolfgang (2004) "International Technology Diffusion," *Journal of Economic Literature*, 42 (3), 752–782.

——— (2010) "International Trade, Foreign Direct Investment, and Technology Spillovers," in *Handbook of the Economics of Innovation*, 2, 793–829: Elsevier.

——— (2021) "Knowledge Spillovers, Trade, and Foreign Direct Investment," NBER Working Paper 28739, National Bureau of Economic Research.

Kleinberg, Jon M. (1999) "Authoritative Sources in a Hyperlinked Environment," *Journal of the Association for Computing Machinery*, 46 (5), 604–632.

Kukharskyy, Bohdan (2020) "A Tale of Two Property Rights: Knowledge, Physical Assets, and Multinational Firm Boundaries," *Journal of International Economics*, 122, 103262.

Liu, Ernest and Song Ma (2022) "Innovation Networks and R&D Allocation," NBER Working Paper 29607, National Bureau of Economic Research.

Lucas Jr., Robert E. and Benjamin Moll (2014) "Knowledge Growth and The Allocation of Time," *Journal of Political Economy*, 122 (1), 1–51.

Lybbert, Travis J. and Nikolas J. Zolas (2014) “Getting Patents and Economic Data to Speak to Each Other: An ‘Algorithmic Links with Probabilities’ Approach for Joint Analyses of Patenting and Economic Activity,” *Research Policy*, 43 (3), 530–542.

MacGarvie, Megan (2006) “Do Firms Learn from International Trade?” *Review of Economics and Statistics*, 88 (1), 46–60.

Perla, Jesse and Christopher Tonetti (2014) “Equilibrium Imitation and Growth,” *Journal of Political Economy*, 122 (1), 52–76.

Sampson, Thomas (2020) “Technology Gaps, Trade and Income,” Working Paper.

Sanderson, Eleanor and Frank Windmeijer (2016) “A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables,” *Journal of Econometrics*, 190 (2), 212–221.

Shu, Pian and Claudia Steinwender (2019) “The Impact of Trade Liberalization on Firm Productivity and Innovation,” *Innovation Policy and the Economy*, 19, 39–68.

Topalova, Petia and Amit Khandelwal (2011) “Trade Liberalization and Firm Productivity: The Case of India,” *Review of Economics and Statistics*, 93 (3), 995–1009.

WIPO (2016) “Standard ST.16: Recommended standard code for the identification of different kinds of patent documents. *Handbook on Industrial Property Information and Documentation.*,” `https://www.wipo.int/export/sites/www/standards/en/pdf/03-16-01.pdf` [Accessed: July 2, 2021].

Appendix A Comparison of IO Tables

In this appendix, we provide a descriptive comparison of the knowledge and production IO tables of the US economy and highlight three observations that emerge from the exercise. Throughout this analysis, we focus on the knowledge IO table constructed using the 1993–2002 window of US patent applications and the production IO table constructed using the 2002 BEA Use table as in Section 4.1.[39]

Appendix A.1 Correlations of IO Linkages

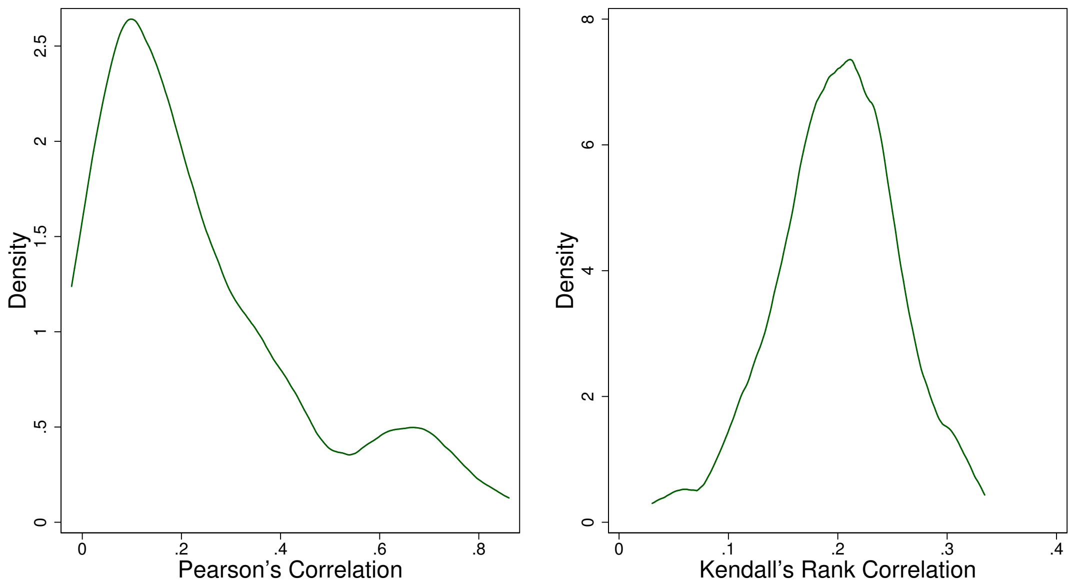
Our empirical analysis, which compares the effects of imports of technology embodied in knowledge and production inputs on patenting outcomes, depends to a large extent on there being distinct variation in the sources of those inputs for the average sector in order to draw the inferences that we do. That this is the case may seem immediate from visual inspection of Figure 1, but here we formalize this underpinning of our analysis. At a high level, the correlation of $\alpha^{l,h}_{US,2002}$ and $\beta^{l,h}_{US,2002}$ across all 84,681 sector-pair IO linkages (for the 291 sectors) is 0.211, while for the off-diagonal IO linkages it is 0.169.

While this is reassuring, we are primarily concerned with the potential that knowledge and production input sources are highly correlated on average within output sectors. To address this, we compute the linear (Pearson) and rank (Kendall adjusted for ties) correlations of $\alpha^{l,h}_{US,2002}$ and $\beta^{l,h}_{US,2002}$ across all input sectors l for each output sector h. The former of these measures evaluates the covariance between knowledge and production inputs and hence their cardinal relationship while the latter evaluates the similarity of the rankings of knowledge and production input sources and hence their ordinal relationship. In Appendix Figure A.1, we plot the distributions of these correlations. One can see that while there are some sectors for which knowledge and production input sources are highly correlated, this is not the case for the vast majority of sectors.

More formally, we display summary statistics of these distributions in Appendix Table A.1. We also include statistics for the distributions of correlation coefficients computed using only off-diagonal IO linkages to show that differences in the intensity of use of own-sector knowledge and production inputs are not driving these low average correlations. We now state our first observation regarding the comparison of the knowledge and production IO tables.

[39] Although we make use of dynamic knowledge IO tables as inputs into our regression analysis, the purpose of this appendix is not to describe the evolution of IO tables over time but instead to demonstrate that the sources of knowledge and production inputs are distinct for the average sector.

Appendix Figure A.1: Distributions of Correlation Coefficients of IO Linkages



Notes: Figure plots the distributions of correlation coefficients of IO Linkages. Coefficients are computed as the correlation of knowledge and production IO linkages across all input sectors for each output sector. The left panel displays the distribution of the Pearson's linear correlation coefficients while the right panel displays the distribution of the Kendall's rank correlation coefficients (adjusted for ties). IO linkages are defined in Section 4.1.

Appendix Table A.1: Summary Statistics of IO Linkage Correlation Coefficients

	Min	Max	Median	Mean	Std. Dev.
All Inputs					
Pearson	-0.022	0.861	0.171	0.236	0.212
Kendall	0.030	0.334	0.202	0.199	0.055
Off-Diagonal Inputs					
Pearson	-0.027	0.861	0.133	0.186	0.195
Kendall	0.021	0.329	0.195	0.193	0.056

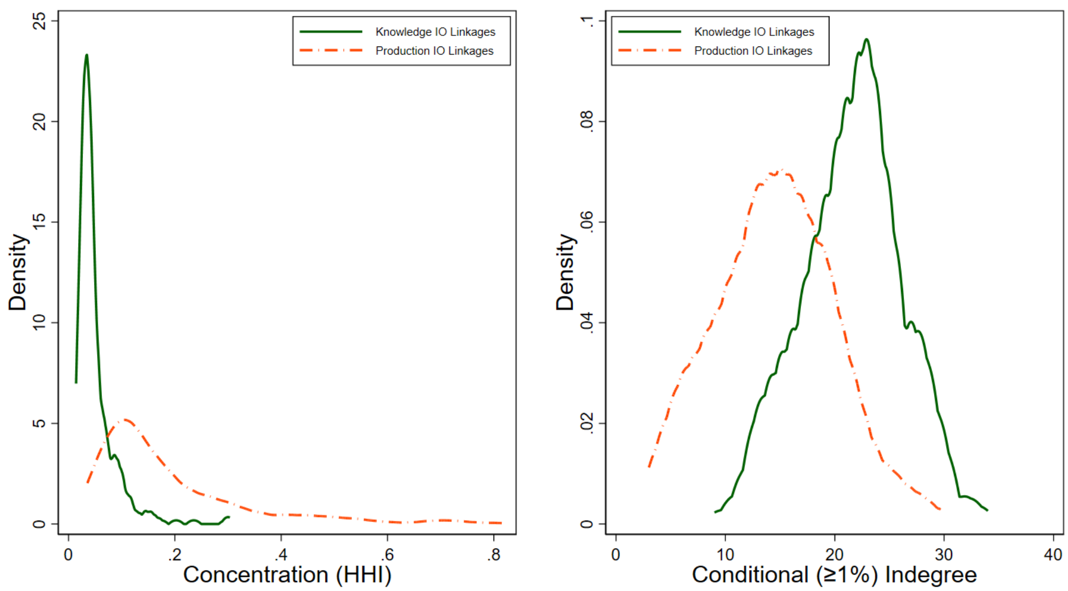
Notes: Table reports summary statistics of the distributions of correlation coefficients of IO linkages for the 291 output sectors plotted in Figure 1. Pearson is the linear correlation between knowledge and production IO linkages. Kendall is the rank correlation (adjusted for ties) of the knowledge and production IO linkages. Coefficients for off-diagonal sectors omit the own-sector IO linkage in the calculation. Std. Dev. is the standard deviation. IO linkages are defined in Section 4.1.

Observation 1: *The sources of knowledge and production inputs are not highly correlated for the average sector.*

Appendix A.2 Concentration and Sparsity of IO Linkages

Next, we investigate another major difference between the knowledge and production IO tables: knowledge inputs tend to be drawn from a wider range of sectors and are less concentrated across input sectors than are production inputs.

Appendix Figure A.2: Distributions of Concentration and Sparsity of IO Linkages



Notes: Figure plots the distributions of the concentration and conditional indegree measures of knowledge and production IO linkages across output sectors. The left panel displays the distributions of concentration measured by the HHI. The right panel displays the distributions of conditional indegrees for the condition $c = 1\%$. The HHI and conditional indegrees are defined in text. IO linkages are defined in Section 4.1.

To demonstrate this, we compute two measures of the concentration or sparsity of input sources for each output sector using the knowledge and production IO linkages. First, we calculate the Herfindahl-Hirschman Index (HHI) of knowledge and production IO linkages for each output sector. For output sector h, these indices are defined as $\text{HHI-K}^h_{US,2002} = \sum_{l\in\mathcal{H}}(\alpha^{l,h}_{US,2002})^2$ for knowledge IO linkages and $\text{HHI-P}^h_{US,2002} = \sum_{l\in\mathcal{H}}(\beta^{l,h}_{US,2002})^2$ for production IO linkages. Second, we construct conditional indegrees (CID) for both IO tables that measure the number of input sectors that have an IO linkage with an output sector that is larger than some threshold level c.[40] For output sector h, the conditional indegree for knowledge IO linkages is $\text{CID-K}^h_{US,2002}(c) = \sum_{i\in\mathcal{H}} \mathbb{1}(\alpha^{l,h}_{US,2002} \geq c)$ and for production IO linkages is $\text{CID-P}^h_{US,2002}(c) = \sum_{i\in\mathcal{H}} \mathbb{1}(\beta^{l,h}_{US,2002} \geq c)$, where $\mathbb{1}(\cdot)$ is the indicator function.

In Appendix Figure A.2, we depict the distributions of the HHI and CID measures for both knowledge and production IO linkages. These graphs show that the mass of the distribution of the concentration of knowledge IO linkages lies to the left of that of the distribution of

[40]As a matter of terminology, we align the meaning of indegree with that of an input sector. However, other authors such as Cai and Li (2019) refer to what we call indegrees as outdegrees in the context of knowledge IO linkages because citations, the data that underlie these measures, flow *from* an output sector (or technology subclass) *to* an input sector (technology subclass).

Appendix Table A.2: Summary Statistics of IO Linkage Concentration Measures

	Min	Max	Median	Mean	Std. Dev.
All Inputs					
HHI-$K^h_{US,2002}$	0.014	0.303	0.039	0.051	0.038
HHI-$P^h_{US,2002}$	0.035	0.823	0.141	0.186	0.142
CID-$K^h_{US,2002}(1\%)$	9	34	22	21.550	4.591
CID-$P^h_{US,2002}(1\%)$	3	30	15	14.509	5.474
Off-Diagonal Inputs					
HHI-$K^h_{US,2002}$	0.013	0.297	0.035	0.043	0.030
HHI-$P^h_{US,2002}$	0.037	0.888	0.142	0.197	0.168
CID-$K^h_{US,2002}(1\%)$	11	34	24	23.533	4.770
CID-$P^h_{US,2002}(1\%)$	3	30	15	15.447	5.597

Notes: Table reports summary statistics of the distributions of the Herfindahl-Hirschman Index (HHI) and conditional indegree (CID) of IO linkages for the 291 output sectors plotted in Figure 1 and for both knowledge and production inputs. For measures computed using off-diagonal sectors, own-sector IO linkages are omitted from the denominators of the IO linkages defined in Section 4.1. The HHI and CID measures are defined in text. The CID measures count IO linkages that are at least 1%. Std. Dev. is the standard deviation.

the concentration of production IO linkages while the reverse is true for the distributions of conditional indegree measures.

Appendix Table A.2 lists summary statistics of these distributions as well as the distributions of the HHI and CID statistics computed using only off-diagonal input sectors. For this latter group of distributions, we modify the definitions of the knowledge and production IO linkages such that, for output sector h, the denominators of Equation (3) and Equation (4) only sum over input sectors $l \neq h$.[41] Knowledge IO linkages are less concentrated than production IO linkages, in part because for the average output sector there are fewer significant knowledge input sectors than production input sectors (where significant means larger than 1% here). We interpret this contrast between the two IO tables as implying that the production IO table is more sparsely connected than the knowledge IO table. This figure and table lead us to our second observation on the differences between the knowledge and production IO tables.

Observation 2: *The sources of production inputs are more highly concentrated than the sources of knowledge inputs for the average sector.*

Appendix A.3 Key Input Sectors

The last major distinction between the knowledge and production IO tables that we explore is the difference between the input sectors that are important suppliers of inputs throughout

[41]This ensures that the shares used to compute the HHI sum to one.

the economy across the two tables. To do this, we consider alternative measures of the economy-wide importance of input sectors and show using each of these measures that the ranking of input sector importance varies across the knowledge and production IO tables.

In particular, we consider three network centrality measures that characterize input sector importance. First, we compute the conditional outdegree (COD) of each input sector analogously to the CID measures discussed in Appendix A.2. For input sector l, these outdegrees are $\text{COD-K}^{l}_{US,2002}(c) = \sum_{h\in\mathcal{H}} \mathbb{1}(\alpha^{l,h}_{US,2002} \geq c)$ for knowledge IO linkages and $\text{COD-P}^{l}_{US,2002}(c) = \sum_{h\in\mathcal{H}} \mathbb{1}(\beta^{l,h}_{US,2002} \geq c)$ for production IO linkages. Second, we use the (unconditional) weighted outdegree (WOD) of input sectors with $\text{WOD-K}^{l}_{US,2002} = \sum_{h\in\mathcal{H}} \alpha^{l,h}_{US,2002}$ for knowledge IO linkages and $\text{WOD-P}^{l}_{US,2002} = \sum_{h\in\mathcal{H}} \beta^{l,h}_{US,2002}$ for production IO linkages. Finally, we calculate the authority weight centrality (AWC) developed by Kleinberg (1999) that represents the contribution of each input sector to the entire knowledge or production IO table and is determined simultaneously with the hub weight centrality (HWC) that represents the absorption of inputs of each output sector from the knowledge or production IO table.[42] In our context, these measures are defined by

$$\begin{aligned}
\text{AWC-K}^{l}_{US,2002} &= \lambda_K \sum_{h\in\mathcal{H}} \alpha^{l,h}_{US,2002}\text{HWC-K}^{h}_{US,2002},\\
\text{HWC-K}^{l}_{US,2002} &= \mu_K \sum_{h\in\mathcal{H}} \alpha^{h,l}_{US,2002}\text{AWC-K}^{h}_{US,2002},\\
\text{AWC-P}^{l}_{US,2002} &= \lambda_P \sum_{h\in\mathcal{H}} \beta^{l,h}_{US,2002}\text{HWC-P}^{h}_{US,2002},\\
\text{HWC-P}^{l}_{US,2002} &= \mu_P \sum_{h\in\mathcal{H}} \beta^{h,l}_{US,2002}\text{AWC-P}^{h}_{US,2002},
\end{aligned}$$

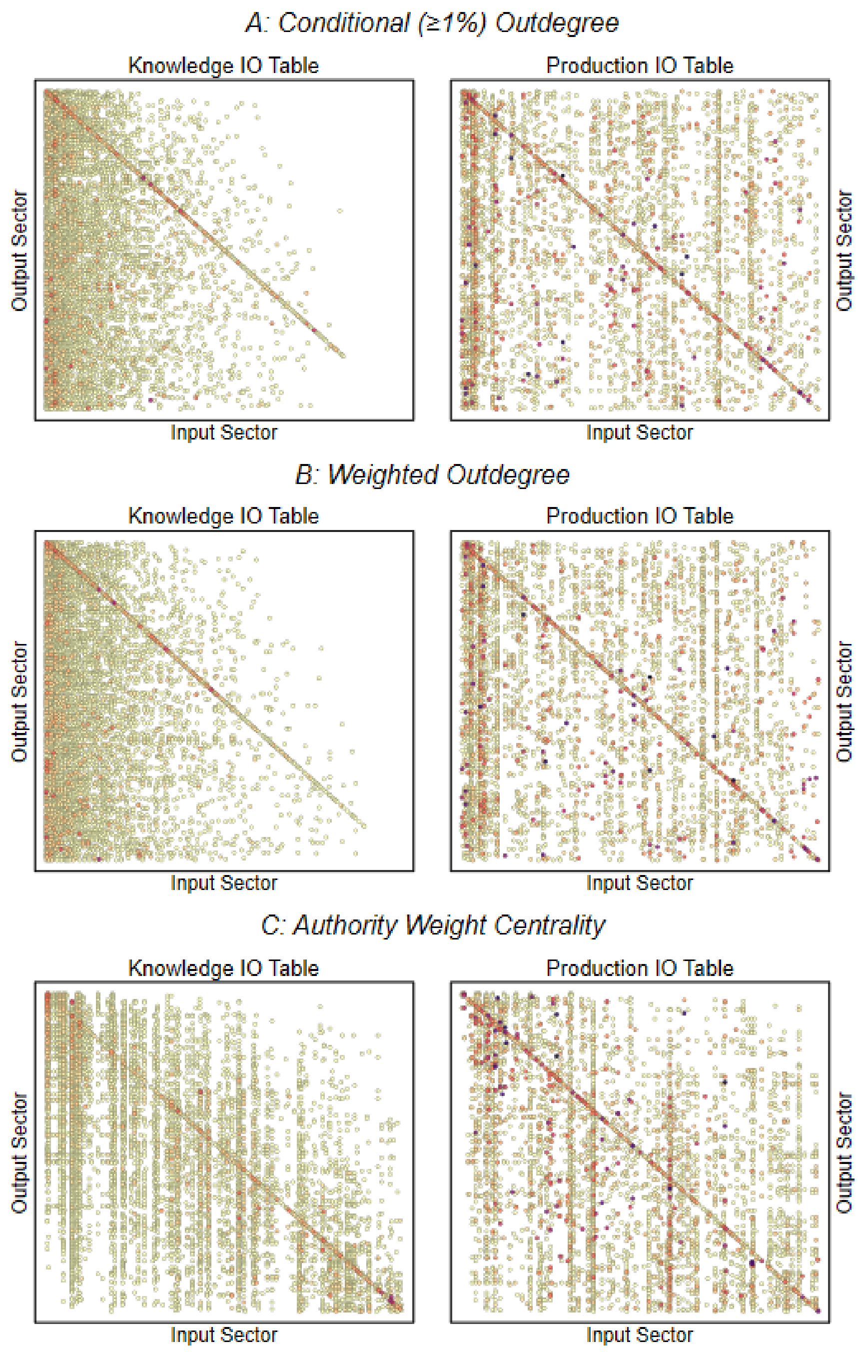
where λ_K (λ_P) and μ_K (μ_P) are the Euclidean norms of the vectors of $\{\text{AWC-K}^{l}_{US,2002}\}_{l\in\mathcal{H}}$ ($\{\text{AWC-P}^{l}_{US,2002}\}_{l\in\mathcal{H}}$) and $\{\text{HWC-K}^{l}_{US,2002}\}_{l\in\mathcal{H}}$ ($\{\text{HWC-P}^{l}_{US,2002}\}_{l\in\mathcal{H}}$), respectively.

To illustrate that the key input sectors are different across the IO tables, we reproduce versions of Figure 1 in which we reorder sectors according to the ranking of sectors by these three centrality measures. In Appendix Figure A.3, we order sectors in each panel by the rank of sectors of the corresponding centrality measure in the knowledge IO table. Sectors follow the *same* order in the plot of both the knowledge and production IO tables.

It is clear from Appendix Figure A.3 that the importance of a sector as a supplier of inputs in the knowledge IO table is not highly related to the importance of the sector as a supplier

[42] Cai and Li (2019) document that the authority weight centralities of sectors and patent technology classes are important determinants of sector-level and firm-level innovation activity.

Appendix Figure A.3: Key Input Sectors in the Knowledge IO Table



Notes: Figure plots the knowledge and production IO tables with sectors ordered by the rank of the centrality measures constructed using knowledge IO linkages. Within each panel, the row position of each output sector and column position of each input sector is held constant across both IO tables. Panel A ranks sectors by the conditional outdegrees for the condition $c = 1\%$. Panel B ranks sectors by the weighted outdegree. Panel C ranks sectors by the authority weight centrality. Each centrality measure is defind in text. IO linkages are defined in Section 4.1. Knowledge IO linkages are based on backward citations of US patents filed between 1993–2002 while production IO linkages are based on the 2002 BEA Use table. All plots only display IO linkages that account for at least 1% of the inputs used by an output sector.

of inputs in the production IO table.[43] We close this section by stating our third observation from comparing the US knowledge and production IO tables.

Observation 3: *The key input-supplying sectors are distinct in the knowledge and production IO tables.*

Appendix B Data Appendix

Google Patents Data. Our knowledge IO linkages, stocks of technology, and diffusion and innovation outcomes are constructed using data from the Google Patents Public Data available from IFI CLAIMS Patent Services and Google (2022). This paper uses the November 2021 version of the database, which includes patents applied for at 105 different national and regional patent offices between 1782 and 2021 with patent inventors located in 242 different countries and regions.[44] Each patent used in our analysis is *linked* to the patents it cites (from any year since 1782) and the patents that cite it (through 2021).

We draw data from Google Patents at the patent family level, where a patent family is the collection of all applications for a given innovation. A patent application to a patent office potentially comprises multiple patent documents submitted to that office or that are produced in the examination and granting process. Some of these documents include original and revised primary documents and some represent supplementary documents, such as non-patent literature and search reports.[45]

We begin by determining the *focal set of patent families* that are the object of our analysis. These families have non-missing data for IPC version 8 codes, filing dates, and inventor countries listed in their *primary series* documents as defined in point 11 of WIPO (2016) (i.e., those with letter groups 1–3).[46] We refer to these primary series documents as *primary publications* and to all other documents as *supplementary publications*. All of our analysis examines effects on the focal set of patent families for which data are collected solely from primary publications. For patent families that are linked to this focal set of patent families through forward and backward citations, we prioritize recording data from primary

[43]When sectors are instead ordered by the rankings of the centrality measures constructed using production IO linkages, the reverse implication is visually apparent. These graphs are available on request.

[44]The large number of locations is accounted for by the inclusion of sub-national regions, such as Hong Kong, which we keep as separate regions whenever trade data is also available for the sub-national region.

[45]The Google Patents database contains a total of 136.1 million different patent documents.

[46]92% of patent publications are primary series documents, and 98% of patent families have at least one primary series document filed.

publications but make use of information in supplementary publications if the relevant information (e.g., the IPC codes) is missing from all available primary publications of the linked families.

Out of a total of 74.8 million patent families in the Google Patents database, 67.9 million of them have at least one 4-character IPC code, which is a minimum requirement in order for them to be included in the data underlying the knowledge IO tables we construct. Meanwhile, 71.8 million patent families have filing dates, while only 20.9 million have inventor country information.[47] In total, 18.9 million patent families have all three sets of information. The focal set of patent families is the subset of 18.0 million patent families which derive all of this information from primary publications.

As there are potentially multiple sets of filing dates, inventor countries, and IPC codes coming from the different publications within a patent family, we aggregate all of this information up to the patent-family level using the following rules. The filing date is the earliest of the filing dates found in the family's primary publications. The list of inventor countries are those in the longest vector of inventor countries found in the family's primary publications.[48] The set of IPC codes for a patent family corresponds to the superset of all distinct 4-character IPC codes contained in the family's primary publications. For patent families that are linked to focal patent families, data for any of these fields that are missing from primary publications are then taken from supplementary publications to fill in data gaps.[49] We record whether or not a patent family is triadic using information on the patent offices to which the patent family's applications are submitted. In the rest of this section and throughout the paper, *patent* refers to the data associated with a patent family as measured according to this procedure.

Our knowledge IO table is constructed from the backward citations of focal patent families. To identify these citations, for each focal patent we record the list of distinct linked cited patents that appear in any of the primary publications of the citing focal patent.[50] In total,

[47]The number of patent families with assignee country information is only slightly higher at 23.1 million patent families covered. We do not use assignee country information to allocate patent families to countries as described below since the location of a patent assignee may not correspond to the location where innovation activity takes place, particularly for assignees that are multinational businesses.

[48]Note that the list of inventor countries may, by design, contain multiple instances of the same country, as different inventors can reside in the same country.

[49]By construction, this does not occur for our focal set of patent families.

[50]To compute the innovation outcome variables based on counts of forward citations received by focal patents from the linked patents that cite them, we additionally record the list of distinct cited (focal) patents that appear in the supplementary publications of the citing patents whenever a citing patent family has no citations in its primary publications. We do this to maximize the coverage of forward citations of focal patents in our data.

there are 10.8 million focal patent families with at least one such backward citation. Almost all of these have at least one backward citation in a primary publication that cites a patent that has a 4-character IPC code and are therefore included in the set of patents whose data underlie the technology subclass-to-technology subclass knowledge IO table.[51]

Using this data, we allocate focal patents to countries and technology categories to construct variables at the level of aggregation used in our analysis. We assign shares of each patent to countries in proportion to the share of inventors from each country listed in the patent application documents.

To produce a pre-concordance dataset at the country-technology subclass-filing year level for our innovation outcome variables, we treat each distinct technology subclass listed on a focal patent family as a separate patent. We add up the (fractional) count of each outcome for focal patents listing each technology class in each filing year and each country after applying the inventor-country weights to those patents. In particular, for a given country-technology subclass-year grouping of patents, we count the amounts of the following variables: total patents, total forward citations and five-year forward citations received by those patents, and total and five-year foreign forward citations (i.e., those citations received by the grouping of patents from patents in other countries, where we use inventor-country weights for both cited and citing patents).

For technology subclass-to-technology subclass backward citations, which are the data underlying our measurement of knowledge IO linkages, we additionally treat each distinct technology subclass listed on a linked cited patent as a separate patent. We calculate the number of backward citations of a given country-output technology subclass-filing year grouping to each input technology subclass of the patents cited by the grouping using inventor-country shares as weights and treating both input and output patents with multiple technology subclasses as multiple patents.[52] We use the counts contained in the cells of the resulting technology subclass-to-technology subclass input-output matrix to measure backward citations for our diffusion outcome variables.

Concordance Details and Sources. We use many concordances between data classification systems in this paper. Below, we describe the processes used to apply the concordances in more detail and provide the locations at which the concordance files can be accessed.

[51] Only around 17 thousand focal patents cite patents that do not have IPC code data.

[52] These counts are also computed for backward citations to each input technology subclass for cited US, domestic, and foreign patents by citing country-technology subclass-year patents (using inventor-country weights for both cited and citing patents).

We first crosswalk the Google Patents data on technology stocks, defined in Section 5.1.3, patent counts, and forward and backward citations, all of which are measured at the 4-character IPC version 8 level, to the 2002 BEA sector categories in two stages. The first stage uses the concordance weights between IPC technology subclasses and 2002 6-digit HS codes developed by Lybbert and Zolas (2014) and then takes these data from 2002 6-digit HS codes into 1992 6-digit HS codes.[53] This second concordance uses equal weights for each 1992 6-digit HS code into which a given 2002 6-digit HS code maps.[54]

The second stage, which is also applied to the BACI trade data that are categorized by 1992 6-digit HS codes, applies three distinct concordances to convert the data to the endpoint 2002 BEA classification. The first concordance identifies the 1987 4-digit SIC codes associated with each 1992 6-digit HS code using an unweighted mapping between the two classification systems.[55] The second concordance converts 1987 4-digit SIC codes into 2002 6-digit NAICS codes, again using an unweighted mapping between the classifications.[56] Combining these two concordances provides the set of 2002 6-digit NAICS codes associated with each 1992 6-digit HS code. We construct concordance weights to map the latter into the former using the share of employment of each NAICS code into which an HS code maps in the total employment of the NAICS codes associated with each HS code. Data on employment by 2002 NAICS code are taken from the 2003 County Business Patterns (CBP) dataset, which is the earliest available disaggregated source of employment data by NAICS code using the 2002 version of the NAICS codes.[57] The third concordance applies the mapping of 2002 6-digit NAICS codes into the endpoint 2002 BEA sector codes.[58] The composite weights between 1992 6-digit HS codes and our endpoint classification implied by combining the three concordances of

[53]There is no concordance between IPC technology subclasses and 1992 6-digit HS codes available. The first set of concordance weights can be accessed at `https://sites.google.com/site/nikolaszolas/PatentCrosswalk` (last accessed on August 3, 2022).

[54]These equal concordance weights are constructed from the unweighted crosswalk available from the World Bank's World Integrated Trade Solution (WITS) database accessible after creating an account at `https://wits.worldbank.org/product_concordance.html` (using the WITS classification labelling, this is the H2 to H0 concordance file; last accessed on August 5, 2022).

[55]This is taken from WITS at `https://wits.worldbank.org/product_concordance.html` (the H0 to SIC concordance file).

[56]This file is available from the US Census Bureau at `https://www.census.gov/naics/?68967` (last accessed on August 3, 2022).

[57]Using employment weights improves upon the alternative of using equal weights that arises due to the lack of weights in the files used in the first and second concordances of this stage. These data come from the US Census Bureau and are available at `https://www.census.gov/programs-surveys/cbp/data/datasets.html` (last accessed on August 3, 2022).

[58]The concordance file can be found in Appendix A of the BEA 2002 Standard Make and Use Tables available at `https://www.bea.gov/industry/benchmark-input-output-data` (last accessed on August 3, 2022).

this second stage are precisely the weights mapping subsectors into sectors referred to in Section 5.1.3.

For the backward citations data used to measure knowledge IO linkages, we apply these two crosswalk stages to both the cited and citing technology subclasses.

To measure production IO linkages in different years consistently in terms of our endpoint 2002 BEA classification, we apply concordances that are similar in nature to the second stage of the crosswalk of technology categories just described. We convert the inter-sectoral sales data in the 1992, 1997, and 2007 BEA Use tables.

For 1992, sector categories are based on the 1987 BEA classification system. We map categories from this system into the 1987 4-digit SIC sectors using a concordance provided by the BEA.[59] We then use the concordance between 1987 4-digit SIC sectors and 2002 6-digit NAICS sectors mentioned earlier to identify the 2002 NAICS sectors associated with each 1987 BEA sector. Using the same procedure as the second stage above, we compute as concordance weights the share of employment of each 2002 NAICS code into which a 1987 BEA sector maps in the total employment of those mapped-into 2002 NAICS codes with the 2003 CBP employment data. We combine these weights with the mapping of 2002 6-digit NAICS codes into the 2002 BEA classification to conduct the crosswalk.

In the 1997 table, the 1997 BEA classification of sectors is based on 1997 6-digit NAICS sectors. We use the BEA concordance between these classifications and the concordance between the 1997 6-digit NAICS sectors and 2002 6-digit NAICS sectors to identify the 2002 NAICS sectors associated with each 1997 BEA sector.[60] We proceed as before and construct weights for mapping 1997 BEA sectors into 2002 NAICS sectors using the 2003 CBP employment data and combine these weights with the mapping of 2002 6-digit NAICS codes into the 2002 BEA classification to conduct the crosswalk.

The data for the 2007 table are available only in terms of the 2012 BEA classification of sectors, which are themselves based on the 2012 6-digit NAICS sectors. In this case, we use three separate concordances to identify the 2002 NAICS sectors associated with each 2012 BEA sector. First, we use the crosswalk between the 2012 BEA classification and the 2012 NAICS sectors provided by the BEA.[61] The second and third concordances map

[59]This can be found at `https://www.bea.gov/industry/benchmark-input-output-data` using the 1987 Use table appendices.

[60]The first of these concordances is available at `https://www.bea.gov/industry/benchmark-input-output-data` using the appendices of the 1997 Use table (after redefinitions) while the second concordance is available at `https://www.census.gov/naics/?68967`.

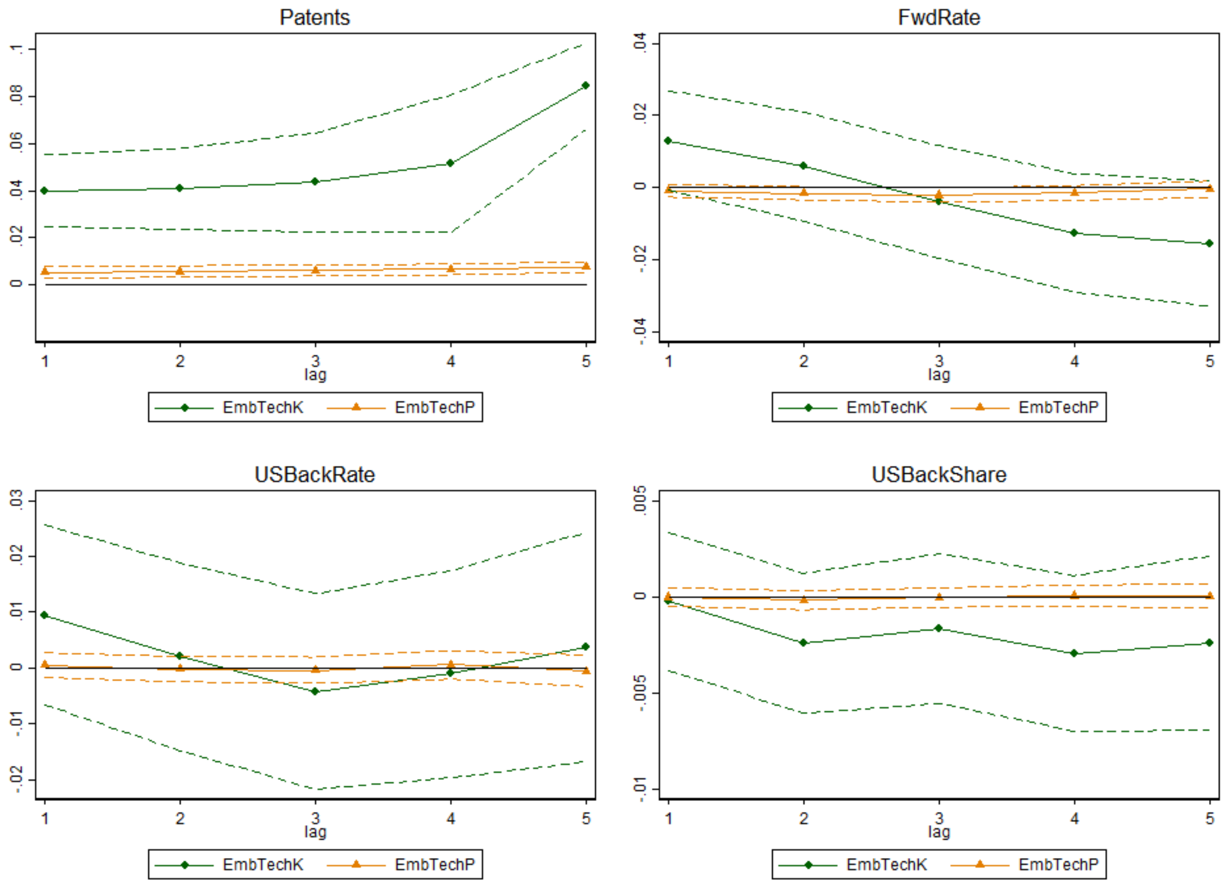
[61]This is available in the appendix of the 2007 Use table found at `https://www.bea.gov/industry/input-output-accounts-data` (last accessed on August 3, 2022).

2012 NAICS sectors into 2007 NAICS sectors and 2007 NAICS sectors into 2002 NAICS sectors, respectively.[62] Employment-based concordance weights for mapping between 2012 BEA sectors and 2002 NAICS sectors are constructed using the 2003 CBP employment data. We combine these weights with the mapping of 2002 NAICS sectors into the 2002 BEA sectors to complete the crosswalk.

[62] Both concordance files are available at `https://www.census.gov/naics/?68967`.

Appendix C Additional Figures and Tables

Appendix Figure C.1: Additional Outcomes Estimated at Different Lags



Notes: Figure plots coefficient estimates of the effects of $EmbTechK$ and $EmbTechP$ on $Patents$ (top left panel), $FwdRate$ (top right panel), $USBackRate$ (bottom left panel), and $USBackShare$ (bottom right panel) for five separate models using Equation (13) with the outcome variable measured in period t. The five models include the explanatory variables lagged by τ years relative to year t for each of $\tau \in \{1, 2, \ldots, 5\}$. The dashed lines are the 95% confidence intervals of the effects. All models include the same controls and fixed effects as in the baseline specification.

Appendix Table C.1: Five-Year Average for Outcome Variables

	(1) *Patents*	(2) *FwdCites*	(3) *FwdRate*	(4) *USBackCites*	(5) *USBackRate*	(6) *USBackShare*
ln($EmbTechK$)	0.041***	0.062***	0.029***	0.084***	0.018*	0.001
	(0.008)	(0.011)	(0.010)	(0.017)	(0.011)	(0.003)
ln($EmbTechP$)	0.006***	0.007***	-0.000	0.011***	0.002	0.000
	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.000)
Observations	478,880	478,880	377,096	478,880	377,096	372,451
F-Stats						
EmbTechK	4,468	4,468	8,225	4,468	8,225	8,673
EmbTechP	15,030	15,030	19,520	15,030	19,520	20,248
EmbTechDiag	397	397	469	397	469	464

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a five-year window and then, except for *USBackShare*, transformed using $\ln(1 + Outcome)$, where *Outcome* is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.2: Leave-One-Out Instrument for Imports

	(1) *Patents*	(2) *FwdCites*	(3) *FwdRate*	(4) *USBackCites*	(5) *USBackRate*	(6) *USBackShare*
ln($EmbTechK$)	0.041***	0.059***	0.024***	0.080***	0.013	-0.000
	(0.008)	(0.011)	(0.009)	(0.016)	(0.010)	(0.002)
ln($EmbTechP$)	0.006***	0.006***	-0.000	0.011***	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.000)
Observations	478,880	478,880	361,290	478,880	361,290	356,457
F-Stats						
EmbTechK	4,477	4,477	10,924	4,477	10,924	11,473
EmbTechP	15,031	15,031	20,787	15,031	20,787	21,476
EmbTechDiag	396	396	483	396	483	472

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a three-year window and then, except for *USBackShare*, transformed using $\ln(1 + Outcome)$, where *Outcome* is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.3: Leave-One-Out Within-Cluster Instrument for Imports

	(1)	(2)	(3)	(4)	(5)	(6)
	Patents	*FwdCites*	*FwdRate*	*USBackCites*	*USBackRate*	*USBackShare*
ln(*EmbTechK*)	0.029***	0.044***	0.023***	0.062***	0.024**	0.002
	(0.006)	(0.008)	(0.009)	(0.013)	(0.011)	(0.002)
ln(*EmbTechP*)	0.005***	0.006***	-0.000	0.010***	0.002	0.000
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.000)
Observations	478,880	478,880	361,290	478,880	361,290	356,457
F-Stats						
EmbTechK	1,195	1,195	1,106	1,195	1,106	1,111
EmbTechP	1,217	1,217	2,051	1,217	2,051	2,255
EmbTechDiag	101	101	78	101	78	78

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a three-year window and then, except for *USBackShare*, transformed using $\ln(1+Outcome)$, where *Outcome* is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.4: ln(*Outcome*) Transformation of Outcome Variables

	(1)	(2)	(3)	(4)	(5)	(6)
	Patents	*FwdCites*	*FwdRate*	*USBackCites*	*USBackRate*	*USBackShare*
ln(*EmbTechK*)	0.073***	0.099***	0.034***	0.103***	0.029**	-0.000
	(0.018)	(0.021)	(0.013)	(0.022)	(0.013)	(0.002)
ln(*EmbTechP*)	-0.002	-0.004	-0.001	-0.001	0.000	0.000
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.000)
Observations	361,290	342,782	342,782	344,145	344,145	356,457
F-Stats						
EmbTechK	10,890	11,853	11,853	11,759	11,759	11,433
EmbTechP	20,754	23,814	23,814	23,654	23,654	21,441
EmbTechDiag	484	470	470	466	466	473

Notes: Table reports 2SLS coefficient estimates using Equation (13) but with an alternative transformation of the outcome variable. All dependent variables are averaged over a three-year window and then, except for *USBackShare*, transformed using $\ln(Outcome)$, where *Outcome* is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.5: arsinh($Outcome$) Transformation of Outcome Variables

	(1)	(2)	(3)	(4)	(5)	(6)
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
ln($EmbTechK$)	0.048***	0.067***	0.030***	0.088***	0.015	-0.000
	(0.009)	(0.012)	(0.011)	(0.017)	(0.012)	(0.002)
ln($EmbTechP$)	0.007***	0.007***	-0.000	0.011***	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.000)
Observations	478,880	478,880	361,290	478,880	361,290	356,457
F-Stats						
$EmbTechK$	4,468	4,468	10,890	4,468	10,890	11,433
$EmbTechP$	15,030	15,030	20,754	15,030	20,754	21,441
$EmbTechDiag$	397	397	484	397	484	473

Notes: Table reports 2SLS coefficient estimates using Equation (13) but with an alternative transformation of the outcome variable. All dependent variables are averaged over a three-year window and then, except for $USBackShare$, transformed using arsinh($Outcome$), where $Outcome$ is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of $OwnTech$, $EmbTechDiag$, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.6: Non-Zero Patent Counts Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
ln($EmbTechK$)	0.158***	0.198***	0.024***	0.281***	0.013	-0.000
	(0.013)	(0.016)	(0.009)	(0.022)	(0.010)	(0.002)
ln($EmbTechP$)	0.007***	0.006***	-0.000	0.011***	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.000)
Observations	361,290	361,290	361,290	361,290	361,290	356,457
F-Stats						
$EmbTechK$	10,890	10,890	10,890	10,890	10,890	11,433
$EmbTechP$	20,754	20,754	20,754	20,754	20,754	21,441
$EmbTechDiag$	484	484	484	484	484	473

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a three-year window and then, except for $USBackShare$, transformed using $\ln(1 + Outcome)$, where $Outcome$ is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of $OwnTech$, $EmbTechDiag$, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.7: Technology Stocks Based on Patent Counts

	(1) *Patents*	(2) *FwdCites*	(3) *FwdRate*	(4) *USBackCites*	(5) *USBackRate*	(6) *USBackShare*
ln($EmbTechK$)	0.050***	0.069***	0.022**	0.097***	0.011	-0.001
	(0.010)	(0.013)	(0.010)	(0.019)	(0.011)	(0.003)
ln($EmbTechP$)	0.007***	0.007***	-0.001	0.013***	0.001	0.000
	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.000)
Observations	478,880	478,880	361,290	478,880	361,290	356,457
F-Stats						
EmbTechK	3,576	3,576	7,895	3,576	7,895	8,132
EmbTechP	11,514	11,514	16,367	11,514	16,367	16,736
EmbTechDiag	383	383	537	383	537	528

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a three-year window and then, except for *USBackShare*, transformed using $\ln(1 + Outcome)$, where *Outcome* is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.8: Top 40 Countries by Total Patents

	(1) *Patents*	(2) *FwdCites*	(3) *FwdRate*	(4) *USBackCites*	(5) *USBackRate*	(6) *USBackShare*
ln($EmbTechK$)	0.061***	0.082***	0.031***	0.097***	0.012	-0.001
	(0.012)	(0.015)	(0.009)	(0.020)	(0.010)	(0.002)
ln($EmbTechP$)	0.007***	0.006**	-0.002*	0.011***	-0.000	-0.000
	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)	(0.000)
Observations	233,600	233,600	223,624	233,600	223,624	222,510
F-Stats						
EmbTechK	5,753	5,753	14,377	5,753	14,377	15,861
EmbTechP	22,156	22,156	27,404	22,156	27,404	28,349
EmbTechDiag	366	366	418	366	418	416

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a three-year window and then, except for *USBackShare*, transformed using $\ln(1 + Outcome)$, where *Outcome* is the variable specified in the column title. All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.9: Other Innovation Outcomes

	Triadic Patents			Foreign Citations	
	(1)	(2)	(3)	(4)	(5)
	Patents	*FwdCites*	*FwdRate*	*FwdCites*	*FwdRate*
$\ln(EmbTechK)$	0.018***	0.033***	0.023**	0.058***	0.026***
	(0.004)	(0.006)	(0.010)	(0.010)	(0.009)
$\ln(EmbTechP)$	0.003***	0.004***	-0.002	0.006***	-0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Observations	478,880	478,880	259,847	478,880	361,290
F-Stats					
EmbTechK	4,468	4,468	13,556	4,468	10,890
EmbTechP	15,030	15,030	34,681	15,030	20,754
EmbTechDiag	397	397	398	397	484

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a three-year window and then transformed using $\ln(1 + Outcome)$, where *Outcome* is the variable specified in the column title. In columns (1) to (3), outcome variables are constructed using information from triadic patent applications, though forward citations from any patent to a triadic patent are included. In columns (4) and (5), the forward citations measures exclude citations where the citing patent is assigned to the domestic country and thus only include foreign citations (based on the inventor-country weights described in Section 2). All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Appendix Table C.10: Other Diffusion Outcomes

	Triadic Patents			Including Own-Sector Citations		
	(1)	(2)	(3)	(4)	(5)	(6)
	USBackCites	*USBackRate*	*USBackShare*	*USBackCites*	*USBackRate*	*USBackShare*
$\ln(EmbTechK)$	0.041***	0.016*	0.001	0.071***	0.018**	-0.002
	(0.008)	(0.010)	(0.002)	(0.013)	(0.008)	(0.002)
$\ln(EmbTechP)$	0.006***	-0.000	0.000	0.009***	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.000)
Observations	478,880	259,847	259,662	478,880	361,290	359,483
F-Stats						
EmbTechK	4,468	13,556	13,622	4,468	10,890	10,865
EmbTechP	15,030	34,681	34,783	15,030	20,754	20,936
EmbTechDiag	397	398	397	397	484	478

Notes: Table reports 2SLS coefficient estimates using Equation (13). All dependent variables are averaged over a three-year window and then, except for *USBackShare*, transformed using $\ln(1 + Outcome)$, where *Outcome* is the variable specified in the column title. In columns (1) to (3), outcome variables are constructed using information from triadic patent applications, though cross-sector backward citations to any patent by a triadic patent are included. In columns (4) to (6), the backward citations measures include both cross-sector and own-sector citations. All explanatory variables are lagged by one year. Other controls include the logs of *OwnTech*, *EmbTechDiag*, total exports to the world, and total imports to the world. All regressions include country-sector, summary-sector-year, and country-year fixed effects. Standard errors clustered at both the country-sector and sector-year level are in parentheses. The F-Stats are the Sanderson and Windmeijer (2016) F statistics for the test of the joint significance of the excluded instruments for each of the first-stage regressions of the endogenous variables that include imports in their construction. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.