Trade and Diffusion of Embodied Technology: An Empirical Analysis*

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

[1]International Monetary Fund
[2]University of Toronto
[3]Bank of Canada
[4]University of British Columbia

October 2022

Abstract

We examine knowledge diffusion through embodied technology trade using global patents and citation data. We use inter-sectoral citation and sales data to characterize knowledge and production input-output (IO) tables for individual countries. Using these IO tables we construct a measure of the knowledge-weighted and production-weighted embodied technology flows imported from the US. We then develop an instrumental variable strategy to identify the causal effect of embodied technology imports on innovation and diffusion. Increases of embodied technology imports lead to increased innovation (measured by forward citations) and knowledge diffusion (backward citations).

Keywords: Research Spillovers, Technology Diffusion, Trade, Patents, Innovation
JEL classification: O33, F14, O31, O19, F61

*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 IMF, its Executive Board, or its management.

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 global 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 and 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. Third, accounting for technology embodied in traded inputs has important policy implications. 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.[2]

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

[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]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 patent 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.

technology imports on innovation and diffusion outcomes. The 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 the conceptual framework, firm innovations depend on a combination of R&D investment, domestic knowledge spillovers, and international spillovers from the technological frontier. International spillovers depend on embodied technology imports—the import-weighted stock of frontier knowledge—and the relevance of cross-sector knowledge to the innovating sector.

We use patent data as our primary measure of innovation in our analysis. Patents document innovations that result in new products, new components of existing products, or new methods of producing products. The second contribution of our paper is to construct a novel dataset on country-sector level innovations and trade. We leverage the Google Patents database to construct detailed patent outcomes for a wide range of countries. In particular, the database allows us to construct measures of patenting based on the locations of innovators and measures of 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 consistent sector definitions using a series of concordances.

The third contribution of our paper is to construct measures of embodied technology imports. As a first step, we construct measures of the cross-sector relevance of knowledge. We use cross-sector citations and sales data to construct knowledge and production input-output (IO) tables as a measure of the relevance of cross-sector knowledge. Within a country, we construct the knowledge IO table using the relative share of citations from each sectors' patents to each other sectors' patents. We similarly construct the production IO table using the share of sales between sectors. Unlike with patents, data to construct the production IO table is only available for the US, which we take 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 knowledge and production IO tables are distinct.[3] In particular, we document that knowledge 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.

[3]Though not the focus of our paper, we are amongst 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.

Along with US imports, we use the IO tables to develop measures of embodied technology imports. We develop two measures based on the knowledge and production IO tables that we refer to as the knowledge-weighted and production-weighted embodied technology imports. Specifically, we aggregate US import-weighted knowledge stocks using Cobb-Douglas weights from the knowledge and production IO tables. We also exclude the own-sector component in the construction of embodied technology imports to avoid potential endogeneity concerns arising from within country-sector demand shocks that may increase both innovation activity and imports within the sector. The knowledge-weighted measure is directly related to our mechanism of interest since it relies on knowledge flows across sectors. The production-weighted measure is also included as potentially important transfers of technology can occur through production interactions. A key outcome of our analysis is then to measure the relative strength of spillovers from embodied technology weighted by knowledge and production linkages.

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 patents while our main diffusion outcomes are US backward citations, US backward citations per patent, and the US backward citation share.[4] We also include controls for each country's own knowledge stock, using country-specific cross-sector 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 patenting outcomes, and country-year fixed effects and aggregate sector-year fixed effects to control for common trends to countries and groups of sectors.

A potential concern with estimating the effects of the trade channel of technology diffusion on domestic innovation is that domestic shocks that affect innovation 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 the domestic R&D productivity will reduce demand for embodied technology imports and ordinary least squares estimates of the effects of embodied technology imports 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, then there would be a positive

[4]Forward citations is measured over a five-year period to reduce truncation issues. US backward citations is measured as the total citations of US patents by all patents applied for in a given country-sector-year.

bias on the OLS estimates.[5] 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 total trade (exports plus imports) to GDP ratio and GDP per capita. We then construct the instrument for each country as US exports to all countries outside of the country's cluster. The instrument isolates US supply shocks by excluding countries that are likely to experience correlated demand shocks.[6]

Using our IV strategy, we find that a 1% increase in the knowledge-weighted embodied technology imports increases citation-weighted patenting by around 0.059%. In comparison, a 1% increase in the production-weighted embodied technology imports increases citation-weighted patenting by 0.006%. To quantify the size of the of the estimated coefficients, we find that a one standard deviation increase of the residualized of knowledge-weighted and production-weighted embodied technology imports account for a 7.0% and 0.7% standard deviation of the residualized citation-weighted patenting respectively. The considerably larger estimate using the knowledge-weighted measures is consistent with our expectation that the knowledge IO table more closely approximates the relevance of knowledge across sectors.

For diffusion outcomes, we find that a 1% increase in embodied technology imports increases US backward citations by 0.081% for the knowledge-weighted measure and 0.011% for the production-weighted measure. Together, embodied technology imports account for just under 10% of the standard deviation of residualized US backward citations. Despite the elasticity for US backward citation being larger than that for patenting, we do not find consistent evidence that either measure of embodied technology imports increases the rate of US backward citations (US backward citations per patent) or the share of US backward citations (out of total backward citations in a country-sector-year). We expect that foreign backward citations are a noisier measures than patenting outcomes and find that the rate of US backward citations becomes positive and statistically significant in many of our robustness exercises.[7]

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

[6] The commonly-used leave-one-out instrument can be viewed as a case of this strategy in which each cluster includes only a single country.

[7] We expect backward citations, especially to foreign patents, to be a nosier measure than patenting outcomes due to a number of factors. First, we expect that our estimate for US backward citations are likely biased downwards due to US backward citations being measured on all previous patents, which is consistent with our relative coefficient estimates. Second, patent laws or customs may not require domestic innovators to cite foreign patents or domestic firms may try to use the domestic market to exclude foreign competitors. Third, while we find that US embodied technology imports lead to new innovations, this does not necessitate that imports affect the composition of knowledge being built on. For example, domestic innovators may learn

Our estimated coefficients are robust to a variety of alternative specifications. We find similar estimates for coefficients at different lags for regressors. We consider alternative instruments constructed using a traditional leave-one-out approach and constructed using all other countries within a cluster (as opposed to all countries outside of the cluster in our baseline). We find economically more significant results when we restrict the sample to the 40 countries with the most patenting activity. Finally, we find similar results using alternative transformations of knowledge stocks, alternative constructions of the knowledge stock, alternative innovation and diffusion outcomes, or other controls. In many of the robustness exercises we also find that the estimate of the US backward citation rate becomes statistically significant, consistent with our view that this variable is imprecisely measured.

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 work pioneered 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 embodied technology imports 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 amongst these is Cai et al. (2022), which allows for inter-sector technology diffusion (both within and across borders). Unlike this work, we provide evidence on embodied technology imports as 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); 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

about third-country innovations from US imports.

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, 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 technology embodied imports.

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) documents the patterns of citations across technology classes in US patents and uses them to construct innovation IO networks to show how inventions developed in one class spillover to other classes and 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 evolve 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

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

analysis. Section 4 describes the constructions of the knowledge and production IO tables. Section 5 describes our empirical strategy and baseline specifications. 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

[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 amongst 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.

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 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.

[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 at 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).

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 is 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 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]

[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 specifications includes 292 sectors.

[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), amongst others. Kukharskyy (2020) uses the concordances with citations data to construct cross-sector 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 cross-sector knowledge linkages and combines them with production

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, backward 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 $t = \{0, 1, 2, ...\infty\}$. The economy is populated by a unit mass of identical and perfectly competitive firms in each sector of each country. Because firms are identical, we refer to them by their country-sector-year (i, h, t) to simplify notation. To be consistent with our data structure and the empirical approach described in Section 5, we define three levels of sectoral aggregation—denote n as a summary sector (the highest aggregation), h as a sector (the focus of our analysis), and p as a sub-sector (or product). We also define the sets $\mathcal{P}^h$ as the set of sub-sectors 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.[18] Expected profits in period $t+1$ can be written as $\mathbb{E}_t[\pi^h_{i,t+1}] = \bar{\pi}_{i,t} \times \bar{\pi}^{n(h)}_t \times \bar{\pi}^h_i \times e^{u^h_{i,t}}$ where $u^h_{i,t}$ is an independent and identically distributed random variable that is known to firms in period t. We use a broader sector aggregation for sector-time expected profits to be consistent with our empirical strategy.

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.

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

A firm (i, h, t) that invests $R^h_{i,t}$ into R&D produces innovations in the next period at rate

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

where $\psi^h_{i,t}$ governs the relative cost of R&D across country-sector-years, $Z^h_{i,t}$ is the domestic stock of relevant knowledge for sector h, and $S^h_{i,t}$ is a spillover from the frontier of knowledge (described below). We assume that the R&D cost parameter is equal to $\psi^h_{i,t} = \psi_{i,t} \times \psi^{n(h)}_t \times \psi^h_i \times e^{v^h_{i,t}}$, where $v^h_{i,t}$ is an independent and identically distributed random variable that, like $u^h_{i,t}$, is known to firms in period t. 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 knowledge $Z^h_{i,t}$ depends on the stocks of knowledge in different sectors of the domestic economy and the relevance of those stocks of knowledge as inputs into innovation for the innovating sector h. Domestic knowledge is given by

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

where $G_Z(\cdot)$ is a monotonic function that dictates the strength of spillovers from domestic knowledge in an input sector. We set this function equal to $G_Z(x) = \omega_Z(1+x)^{\eta_Z}$.[19] The other variables are the knowledge stock $K^p_{i,t}$ of country-sub-sector-year (i, p, t) and the relevance of knowledge from country-sector-year (i, l, t) for producing innovations in sector h, denoted by $\alpha^{l,h}_{i,t}$.

Spillovers from the frontier economy depend on the stocks of knowledge embodied in traded goods coming from the frontier economy. The sectoral flow of knowledge coming into sector h from other sectors l depends on a Cobb-Douglas aggregator given by

$$S^h_{i,t} = \prod_l G_S\left(EmbTech^l_{F,i,t}\right)^{\gamma^{l,h}_{F,t}},$$

where $G_S(\cdot)$ is a monotonic function that dictates the strength of spillovers from the embodied frontier technology. We set this function equal to $G_S(x) = \omega_S(1 + x)^{\eta_S}$ in the empirical specification. The frontier spillovers into sector h depend on embodied technology flows

[19]This specification of $G_Z(x)$ is consistent with our treatment of zeros in the empirical analysis. We show that our results are robust to alternative specifications in Appendix C.

from all other sectors l. The value of $\gamma_{F,t}^{l,h}$ captures the strength of knowledge spillovers from sector l to sector h in period t. We allow the strength of knowledge spillovers to be time dependent to account for changes in technologies over this period. We also assume that the relevance of frontier knowledge in sector l for innovating in sector h is determined in the frontier economy, whereas the relevance of domestic knowledge for innovating is specific to the domestic economy. This could be thought of as reflecting how differences in the types of goods produced differences in the relevance of knowledge spillovers from embodied technology.

The flow of embodied technology in sector h is given by

$$EmbTech_{i,t}^{h} = \sum_{p \in \mathcal{P}^h} \left(m_{F,i,t}^{p} \times K_{F,t}^{p} \right),$$

where $K_{F,t}^{p}$ is the frontier stock of knowledge in sub-sector p and $m_{F,i,t}^{p} = M_{F,i,t}^{p}/Y_{i,t}^{h(p)}$ is the imports from the frontier economy to the domestic economy $M_{F,i,t}^{p}$ divided by the sub-sector output $Y_{i,t}^{h}$. Unlike with domestic knowledge, we scale frontier knowledge by the relative abundance of frontier goods in the domestic economy, as measured by $m_{F,i,t}^{p}$. That is, frontier knowledge spillovers depend on the extent to which knowledge is embodied within goods that are available in the domestic economy ($m_{F,i,t}^{p}$) and the amount of that knowledge ($K_{F,t}^{p}$). Intuitively, our measure of embodied technology can be thought of as capturing the probability that a domestic innovator encounters, through chance, a frontier good and, given the encounter, the probability that the innovator realizes a new innovation (as in, for example, Bloom et al., 2013; Lucas Jr. and Moll, 2014; Perla and Tonetti, 2014; Buera and Oberfield, 2020). In this regard, more abundant (higher $m_{F,i,t}^{p}$) or more knowledge intensive (higher $K_{F,t}^{p}$) products both increase domestic innovation.

The problem of a firm is to maximize expected profits net of R&D expenditures by choosing R&D expenditure. Equivalently, the firm's problem can be written as choosing the innovation rate

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

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

$$X_{i,t+1}^{h}=\tilde{\zeta}\prod_{j}G\left(\left(\sum_{p\in\mathcal{P}^{h}}\left(m_{F,i,t}^{p}\times K_{F,i,t}^{p}\right)\right)^{\gamma_{t}^{h,l}}\right)\times Z_{i,t}^{h}\times\left[\frac{\bar{\pi}_{i,t}}{\psi_{i,t}}\times\frac{\bar{\pi}_{t}^{n(h)}}{\psi_{t}^{n(h)}}\times\frac{\bar{\pi}_{i}^{h}}{\psi_{i}^{h}}\times e^{u_{i,t}^{h}-v_{i,t}^{h}}\right]^{\frac{1}{\zeta-1}}, \tag{1}$$

where $\tilde{\zeta}=\zeta^{-1/(\zeta-1)}$.

Taking the log of (1) and grouping variables implies that the innovation rate is given by

$$\ln X_{i,t+1}^{h}=\ln S_{i,t}^{h}+\ln Z_{i,t}^{h}+f_{i,t}+f_{t}^{n(h)}+f_{i}^{h}+\epsilon_{i,t}^{h}, \tag{2}$$

where $f_{i,t}=(\ln\bar{\pi}_{i,t}-\ln\psi_{i,t})/(\zeta-1)$, $f_{t}^{n(h)}=(\ln\bar{\pi}_{t}^{n(h)}-\ln\psi_{t}^{n(h)})/(\zeta-1)$, $f_{i}^{h}=(\ln\bar{\pi}_{i}^{h}-\ln\psi_{i}^{h})/(\zeta-1)$, and $\epsilon_{i,t}^{h}=(u_{i,t}^{h}-v_{i,t}^{h})/(\zeta-1)$. The expression in (2) provides the foundation for our empirical strategy. In the next two sections, we describe the construction of the variables that correspond to the values of $S_{i,t}^{h}$ and $Z_{i,t}^{h}$.

The conceptual framework highlights the relationship between imported embodied technology and innovation outcomes. In the framework, the assumptions on the nature of expected profits and investment costs are relatively flexible and capture many macroeconomic differences across countries and sectors that may otherwise be of concern in estimating the relationship. This would include, for example, country-specific business cycles, sector-specific trends, such as digitalization, and time-invariant differences in the comparative advantage of countries across different sectors. However, difficulties may arise if there are persistent country-sector-specific shocks that drive both an increase in imports and innovation. To deal with these issues, we separate the own-sector and cross-sector effects, since we expect these issues to be most severe within sectors, and we develop an IV strategy. These remedies are discussed in detail in Section 5.

4 Cross-Sector Knowledge Linkages

We use the conceptual framework as a roadmap for the empirical analysis. In this section, we start by developing measures of the cross-sector relevance of knowledge $\alpha_{i,t}^{l,h}$ and $\gamma_{F,t}^{l,h}$ using citation and production relationships between sectors. The linkages, together with import data ($M_{j,i,t}^{p}$) and the stocks of knowledge ($K_{i,t}^{p}$) constructed using patent data, comprise the main variables in the conceptual framework. We also use this section to highlight key differences between the knowledge and production IO tables to shed light on our empirical identification.

4.1 Production and Knowledge Input-Output Tables

Our analysis estimates the effects of traded embodied technology on patenting outcomes. We focus on two natural candidates to describe the relevance of knowledge 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 clearly linked with our focus on innovation outcomes since patent citations represent a direct report of the flow of knowledge. 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 a channel through which knowledge can be shared and diffused around the economy. We construct the measures of relevance using the knowledge and production IO tables that characterize the strength of cross-sector linkages.

Denote the country-sector-year (j, l, s) patents cited by country-sector-year (i, h, t) patents as $Cites^{l,h}_{j,i,s,t}$. This variable captures the reported flow of knowledge from (j, l, s) to (i, h, t).[20] The set of sectors is denoted 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 in country i. We allow for this relationship 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

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

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 allow for slow moving technological transitions.[21] 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.

[20]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 patents.

[21]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 patent citations over this period, highlighting the need for dynamic knowledge IO linkages.

Similarly, we measure production IO linkages as the importance of inputs produced in each input sector for each output sector. Because the availability of highly disaggregated data on cross-sector 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 production IO table. The production IO linkage is given by

$$\beta^{l,h}_{i,t} = \frac{\text{Sales}^{l,h}_{j,i,t}}{\sum_{k \in \mathcal{H}} \text{Sales}^{k,h}_{i,t}}, \tag{4}$$

where $\text{Sales}^{l,h}_{i,t}$ 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 linkages are based on US data from the BEA Use tables as described in Section 2. Since the BEA Use tables are only available at five year intervals, we use the production IO 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 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.[22]

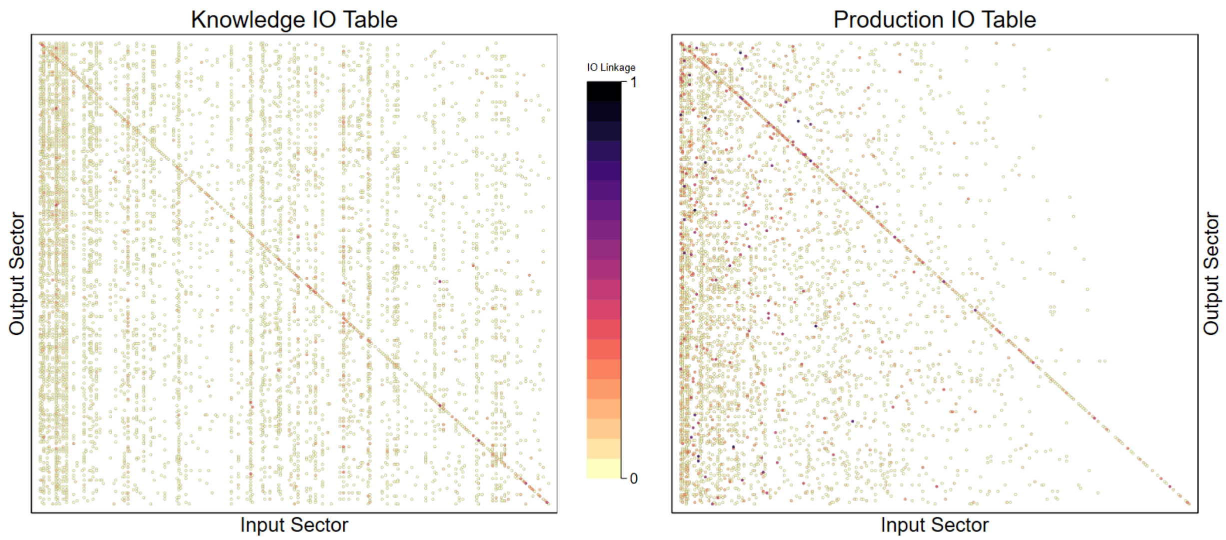
4.2 Description of Knowledge and Production IO Tables

The construction of the knowledge and production IO tables relies on different data. However, there is little point in examining the effects of 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 demonstrate that they are different potential sources of knowledge.

Both knowledge and production IO linkages take on values between zero and one. Values closer to one indicate stronger relationships whereas values further from one and 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 $\alpha^{l,h}_{US,2002}$ represented in the left panel and $\beta^{l,h}_{US,2002}$ 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

[22]To be more precise, we use $\alpha^{l,h}_{i,1992}$ and $\beta^{l,h}_{i,1992}$ for exposure measured between 1995 and 2000, $\alpha^{l,h}_{i,1997}$ and $\beta^{l,h}_{i,1997}$ when we examine exposure between 2001 and 2005, $\alpha^{l,h}_{i,2002}$ and $\beta^{l,h}_{i,2002}$ for years between 2006 and 2010, and $\alpha^{l,h}_{i,2007}$ and $\beta^{l,h}_{i,2007}$ 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.

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 based on their relative importance as a source 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.[23]

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

[23]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 2**: The sources of production inputs are more highly concentrated than the sources of knowledge inputs for the average sector.*

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

We relegate the construction 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 knowledge spillovers. Given this, in our baseline analysis we explore the diffusion of knowledge through 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 of our analysis and the construction of key variables. Following Equation (2), our main regressions involve regressing innovation outcomes on measures of embodied technology imports. We start by specifying our main outcomes of interest. We then use the knowledge and production IO tables constructed in the previous section to develop the main explanatory variables. Finally, we outline the empirical analog of Equation (2) and an instrumental variable (IV) approach that we use to identify the effects of spillovers from embodied technology 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 as we consider the US to be at the technology frontier. 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 this time period (see Berkes et al., 2022, for evidence). In this regard, the US best captures what we think of as the frontier economy. Second, setting the frontier economy to the US allows us to be consistent with data measurement. After adjusting for time and sectoral variation, US patents have a consistent interpretation in our regressions.[24] Further, the US has consistent data available to construct both the production and knowledge IO tables described in Section 4, which are necessary for our analysis.[25]

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 trade data for many countries and including later years would cause truncation issues for patents and forward citations, which are the main data used for our innovation outcome variables.

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 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 amongst 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 restrict our sample to countries that trade for reasons unrelated to production or consumption, such as countries that primarily act as intermediaries. Finally, we keep only those countries that are above the 25th percentile of total triadic patents across all years amongst the remaining countries, which corresponds to a cutoff of just under ten triadic patents over the sample. This restriction excludes countries where innovations are either infrequent or of a 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. Consistent with this expectation, we find substantially larger coefficient

[24] For example, patenting in other countries depends on the country's institutional structure. Consequently, we would be unable to aggregate our embodied knowledge variable across countries since the patent variables represent different stocks of knowledge.

[25] The strength of the shock depends on the relevance of the US-sector knowledge for innovation in downstream sectors in country i. We use the US production and citation IO structures to measure relevance. Measuring consistent production linkages for other countries at the level of sectoral aggregation available for the US is hard due to data limitations.

estimates when we restrict our sample to the top 40 countries in terms of overall patenting (see Appendix Table C.4).

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 knowledge spillovers and innovation outcomes. There, the focus is on the rate of 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 look at the average quality of patents—measured by the number of forward citations per patent (***FwdRate***)—to examine both the intensive and extensive margin effects of frontier knowledge 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

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

   We focus on forward citations received in the first five years of a patents life to mitigate truncation issues that would arise in later periods of the sample if citations received in any year were used instead.[26] We interpret forward citation-weighted patenting as a measure of quality-adjusted patenting. We do not take a stance on whether the expected coefficient for quality-adjusted patenting should be larger or smaller than that of the raw patent count. The coefficient may be larger for quality-adjusted patenting if

[26] 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, a patent filed in June 2000 will include forward citations up to December 31, 2005. We do not expect this choice to affect our results since our unit of measurement is a year.

higher-quality innovators receive larger spillovers and push low-quality innovators are out of the market. Conversely, the coefficient 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

$$FwdRate_{i,t}^{h} = \frac{FwdCites_{i,t}^{h}}{Patents_{i,t}^{h}}. \tag{6}$$

The forward citation rate is a measure of the average quality of patent applications. In this regard, the forward citation rate is a measure of the intensive margin response to a shock to frontier knowledge spillovers. 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 zeros. We exclude such observations from the estimation sample.

Diffusion Outcomes. Patents and citations data are also used to examine 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. Specifically, 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 overall flow of backward citations to US patents, the per-patent rate of backward citations to US patents, and the share of backward citations to US patents in the total backward citations to foreign patents. 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

$$USBackCites_{i,t}^{h} = \sum\nolimits_{s=0}^{t} \sum\nolimits_{l \neq h} Cites_{US,i,s,t}^{l,h}. \tag{7}$$

We exclude the own-sector backward citations from the outcome variable to be consistent with the focus on cross-sector imports of embodied technology described below.[27] That

[27]We look at the entire history of backward citations since backward citations do not suffer from the same truncation bias in later periods as with forward citations. However, we expect that the estimated coefficient will be biased downwards to the extent that more recently developed knowledge is embodied in traded goods.

is, since our main shock is to cross-sector knowledge, we do not want to measure own-sector knowledge flows.

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

$$USBackRate^h_{i,t} = \frac{USBackCites^h_{i,t}}{Patents^h_{i,t}}.$$

   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 amount of cross-sector knowledge that flows from the US to sector h in country i, the second measures the intensity with which the typical sector h patent uses that cross-sector imported embodied knowledge.

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

$$USBackShare^h_{i,t} = \frac{USBackCites^h_{i,t}}{\sum_{j \neq i} \sum_{s=0}^{t} \sum_{l \neq h} Cites^{l,h}_{i,j,s,t}}.$$

   Relative to the other two outcomes, the backward citation share informs us on whether knowledge inputs are substituted towards US knowledge 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 that flow as they learn from those patents as well as the US patents underlying our measurement of embodied technology. This would lead to estimates of the effects of imports of embodied technology on this variable to be small relative to the estimates of effects on the first diffusion outcome.

Summary Statistics. In our baseline specification described below, we measure outcome variables using the average of the variables in the three-year window between year t and $t + 2$. Table 1 presents summary statistics of the main outcome variables used in our baseline analysis. The counts of observations for *FwdRate*, *USBackRate*, and *USBackShare* are smaller than for the other outcomes because, for *FwdRate* and *USBackRate*, the denominators of these rates (*Patents*) are zero for some country-sector-year observations and because, for $USBackShare$, some observations have no cross-sector citations to foreign patents. The summary statistics show that the distribution of $Patents$ and $FwdCites$ are highly skewed with the median country-sector-year having values close to zero. The distribution of outcomes based on backward citations are similarly skewed.

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 $t+2$. All statistics are calculated on the log of one plus the variable except for the statistics for $USBackShare$.

5.1.3 Embodied Technology Imports and Other Controls

Our main variable of interest is the frontier knowledge spillovers ($S^h_{i,t}$ in the conceptual framework). We also describe the construction of knowledge stocks ($K^p_{i,t}$), an important input in the knowledge spillovers, and the domestic stock of knowledge ($Z^h_{i,t}$), which is used as a control in our main specification.

Knowledge Stocks ($\boldsymbol{K^p_{i,t}}$). Before turning to our main variables of interest, we discuss the construction of knowledge stocks $K^p_{i,t}$ since this is used as an input variable. We measure the technological content of a sector's goods using patent data. We follow Hall et al. (2001) in using citation-weights to adjust for the relative quality of patents in the construction of knowledge stocks. Specifically, we use forward citations in the first five years after a patents application as our preferred measure of patent quality since this avoids issues with citation truncation in later periods of the data and requires minimal structure in constructing a comparable measure of quality.[28] We construct the knowledge stock as

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

[28]Additionally, Hall et al. (2001) note differences in the propensity to cite across sectors and that patenting behavior has changed over time. In our baseline results we include industry and time fixed effects to address these issues.

where $FwdCites^p_{i,t}$ is the five-year forwards citations in country-sub-sector-year (i, p, t); and δ is the depreciation rate of knowledge, which we set to 5%, consistent with commonly used values. For each country and sector, we initialize the stock of knowledge $K^p_{i,t}$ in 1940 with value $K^p_{i,1940} = FwdCites^p_{i,t}/\delta$. The initial value has relatively little influence the knowledge stocks over the period of our analysis since it occurs over 50 years prior to the start of this period.

Domestic Technology Stock (*OwnTech*). The countries own stock of knowledge $(Z^h_{i,t})$ also enters into the expression for innovation in Equation (2) because it captures domestic spillovers.[29] We construct the measure of the domestic knowledge stock as

$$OwnTech^h_{i,t} = \prod_l \left(1 + \sum_{p \in \mathcal{P}^l} w^l(p) K^p_{i,t}\right)^{\alpha^{l,h}_{i,t}}, \tag{8}$$

where $w^l(p)$ is the concordance link discussed in Section 2. Higher values of $OwnTech$ capture that the country-sector-year (i, h, t) is more capable with the technology because of its previous stock of innovations. We use the domestic knowledge IO linkages $\alpha^{l,h}_{i,t}$ in the construction because this is the best measure of the relevance of sector l knowledge in country i for innovations in sector h.[30]

Embodied Technology (*EmbTech*). Our main variable of interest reflects the amount of useful technology embodied in imported goods. We do not impose structure on whether knowledge flows between sectors are better captured by the knowledge or production IO links. In this regard, we set the links γ in the conceptual framework to be a combination of knowledge α and production β IO linkages. An important takeaway from our results is the relative importance of knowledge and production linkages for innovation and diffusion outcomes.

Following the conceptual framework, we measure the frontier knowledge spillover in two steps. First, we construct the imported embodied technology flow as the product of the US knowledge stock $K^l_{US,t}$ and US imports $M^l_{US,i,t}$. Second, we weight the imported embodied technology flow by the upstream knowledge and production IO tables. Our measure of

[29] This variable can also be thought of as capturing past frontier knowledge spillovers to the extent that these become embodied within the domestic knowledge stock through innovations.

[30] We do not have consistent measures of production IO linkages for most country-years, which prevents us from constructing a similar measure with $\beta^{l,h}_{i,t}$.

knowledge-weighted embodied technology imports is given by

$$EmbTechK_{i,t}^{h} = \prod\nolimits_{l \neq h} \left(1 + (\sum\nolimits_{p \in \mathcal{P}^l} w^l(p) K_{US,t}^{p} M_{US,i,t}^{p})\right)^{\alpha_{US,t}^{l,h}}, \tag{9}$$

and production-weighted embodied technology imports is given by

$$EmbTechP_{i,t}^{h} = \prod\nolimits_{l \neq h} \left(1 + (\sum\nolimits_{p \in \mathcal{P}^l} w^l(p) K_{US,t}^{p} M_{US,i,t}^{p})\right)^{\beta_{US,t}^{l,h}}. \tag{10}$$

The amount of embodied technology depends on the flow of knowledge into country i from the United States in every sector l. This flow is increasing in the volume of imports and the stock of knowledge in sector l. Countries that spend more on sector l goods from the United States have a higher flow of knowledge 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 readily available and in higher number. The effect of a flow of knowledge from a given sector l is weighted by the tendency of that sector's knowledge to be used in sector h. Table 2 provides a summary of the two measures of embodied technology imports.

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

We construct the measures of embodied technology from the value of US imports, rather than trade scaled by output used in the conceptual framework, due to data limitations with output data being unavailable at the level of aggregation we examine. A potential issue with this construction is that higher imports could simply reflect that the destination is economically larger or more populous.[31] We include granular fixed effects as a best attempt to deal with this issue. An alternative would be to use US import shares, i.e., US imports to country i over all imports to country i. This construction leads to misleading conclusions because trends in trade—e.g., all countries tend to trade more over this period—lead to declining US import shares for most countries.

[31] It is also worth noting that this is not an issue with the US knowledge stock $K_{US,t}^{l}$ since it is country specific and captures the relative abundance of knowledge embodied within imports, meaning the level is important for the interpretation of our results.

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 reasons. One, within sector demand shocks can lead to countries importing more foreign products to satisfy demand, while at the same time invest more in innovation effort in the sector due to increased returns. A second concern is that own sector imports can also affect innovation outcomes in a country through import competition effects, wherein firms may invest more in innovation in order to escape foreign competition. Finally, productivity shocks with a country-sector can also lead to both imports and innovation within the sector to move together. We discuss this endogeneity concern further in Section 5.3.

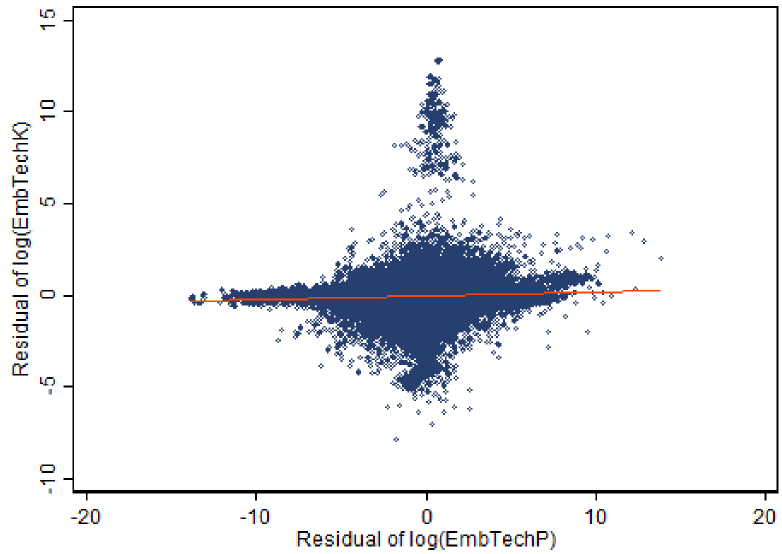
Identifying and estimating the separate effects of knowledge linkages and production linkages requires that there is sufficient variation across observations in our sample in these two measures. To assess this, we regress the logs of both $EmbTechK$ and $EmbTechP$ on the set of fixed effects included in our baseline specifications 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 structures explored in Section 4 show that own-sector inputs 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 (the diagonal of the IO table) embodied technology as

$$EmbTechDiag_{i,t}^{h} = 1 + \sum_{p \in \mathcal{P}^h} w^h(p) K_{US,t}^{p} M_{US,i,t}^{p}. \tag{11}$$

We do not scale the knowledge inputs by the IO weights $\alpha_{US,t}^{l,l}$ or $\beta_{US,t}^{l,l}$ since, as mentioned above, we expect that this variable captures factors not directly related to the effects of technology diffusion, such as import competition. 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 increased exposure

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 specifications discussed in Section Section 5.2.

to trade from other low-wage countries).[32] In contrast, they find that changes in import penetration of high-wage countries like the US had no effect on patenting. Nevertheless, we include this variable to mitigate concerns that the estimated effects of imported embodied knowledge pick up these import competition effects.

5.2 Estimation Equation

We now present the empirical counterpart of Equation (2) in terms of our constructed variables that we estimate:

$$\begin{aligned} \ln(Outcome^h_{i,t}) = \theta_1 \ln EmbTechK^h_{i,t-\tau} + \theta_2 \ln EmbTechP^h_{i,t-\tau} + \theta_3 \ln OwnTech^h_{i,t-\tau} \\ + \theta_4 \ln EmbTechDiag^h_{i,t-\tau} + V^h_{i,t}\beta + f_{i,t} + f^{n(h)}_t + f^h_i + \epsilon^h_{i,t}, \end{aligned} \tag{12}$$

where the outcomes were discussed previously; $EmbTechK^h_{i,h,t-\tau}$ is the US knowledge flows

[32] Autor et al. (2020) find instead 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.

to country-sector (i, h) in period $t - \tau$ using the knowledge IO linkages; $EmbTechP^h_{i,t-\tau}$ is the US knowledge flows to the country-sector (i, h) in period $t - \tau$ using the production IO linkages; $OwnTech^h_{i,t-\tau}$ captures the stock of domestic knowledge in sector h for country i in year $t - \tau$; $V^h_{i,t}$ is a vector of controls that includes own-sector imports from the world and exports to the world in the baseline results; and $f_{i,t}$, $f_t^{n(h)}$ and f_i^h are country-year, (summary) sector-year, and country-sector fixed effects. In the baseline regressions we average outcomes over a three-year window from t to $t + 2$ to reduce noise and to allow for a more gradual diffusion of knowledge. We also transform the outcome variable as $\ln(1 + Outcome^h_{i,t})$ to keep observations that have zero innovation/diffusion outcomes. We show that our results are robust to other data transformations in Appendix C. We also present the results for the model where outcomes are for period t only and the input variables are measured at lags $\tau \in \{1, 2, ..., 5\}$.

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).

We expect similar outcomes for the innovation and diffusion outcomes, with a few exceptions. The coefficient estimates of θ_1 and θ_2 should be positive since our hypothesis is that spillovers from embodied technology imports should improve innovation outcomes. We also expect that the estimates for θ_1 will in general be larger than for θ_2 since the knowledge weights reflect a more direct measure of the relevance of embodied technology imports for patenting. Similarly, the estimate of θ_3 should be positive for the rate of innovating ($Patents$ and $FwdCites$) but may be ambiguous for the quality of patents if higher knowledge stocks correspond to higher rates of low-quality innovations. For diffusion outcomes, the estimate of θ_3 should be positive for overall US citations and near zero for the other outcomes. Finally, the coefficient estimate of θ_4 is also ambiguous since the variable captures both embodied technology imports but also higher competition from US firms, which would tend to discourage domestic innovation.

5.3 Endogeneity Concerns

The fixed effects in Equation (12) 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 useful technology 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 $u_{i,t}^{h}$ 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 this concern, we use an instrumental variable strategy that focuses on variation in US imports that is a function of supply shocks to US exports. Specifically, we instrument each regressor that includes US imports with an instrument that constructs the variable using US exports to all countries outside of a country-specific cluster (discussed below). For our main outcomes, we construct the instrumental variables as

$$IVEmbTechK_{i,t}^{h} = \prod\nolimits_{l \neq h} (1 + (K_{US,t}^{l} \sum\nolimits_{j \notin \mathcal{G}_i} M_{US,i,t}^{l}))^{\alpha_{US,t}^{l,h}}, \tag{13}$$

and

$$IVEmbTechP_{i,t}^{h} = \prod\nolimits_{l \neq h} (1 + (K_{US,t}^{l} \sum\nolimits_{j \notin \mathcal{G}_i} M_{US,i,t}^{l}))^{\beta_{US,t}^{l,h}}. \tag{14}$$

where $\mathcal{G}_i$ is a cluster of countries with similar characteristics to country i. For each country i, we construct the cluster $\mathcal{G}_i$ as the countries that fall into both the same quintile of GDP-per-capita and total trade (imports plus exports) to GDP ratio as country i. For the cluster $\mathcal{G}_i$, we first construct two groupings of countries based on quintiles of GDP-per-capita and quintiles of total trade (imports plus exports) divided by GDP. We include the first variable to capture similarities in technological development across countries and the second variable to capture similarities in trade structures across countries. Our instrumental variable strategy isolates trade to the domestic country that stems from supply shocks to the US. Intuitively, the leave-one-out instrument excludes the domestic economy to avoid counting changes in trade that result from demand shocks. We extend this logic by not only excluding the domestic economy but also countries that share similar characteristics and, consequently, may have

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

correlated demand shocks.

6 Results

In this section, we discuss the results from estimating knowledge-weighted and production-weighted embodied technology imports measured on the innovation and diffusion outcomes. We begin by discussing estimates using the baseline specifications described in Equation (12). We show that our main conclusions hold at different lags. We also use the empirical model to provide a quantification of the magnitude of the results. We close this section by discussing the robustness of the results to other considerations and concerns.

6.1 Baseline Results

Table 3 presents the main innovation results for both the OLS and IV estimates, which we interpret as $X_{i,t}^{h}$ in the context of the conceptual framework. As previously mentioned, all regressions include country-sector, summary sector-year, and country-year fixed effects as described in the model and standard errors are clustered at the country-sector and sector-year levels.

The OLS results in columns (1) to (3) suggest that knowledge-weighted embodied technology imports $EmbTechK$ has a positive impact on the innovation rate as measured by $Patents$ and $FwdCites$. Despite a larger point estimate for $FwdCites$ and a positive point estimate for $FwdRate$, the OLS results do not point to a statistically significant increase in $FwdRate$. Qualitatively, production-weighted embodied technology $EmbTechP$ has a similar affect on the innovation rate as $EmbTechK$. That said, the elasticity is substantially lower, less than a quarter for $Patents$, suggesting that technology spillovers are primarily through knowledge linkages. We show later that the broad quantitative comparison holds after accounting for the relative variation in $EmbTechK$ and $EmbTechP$.

The IV estimates in column (4) to (6) are larger and, for the case of $FwdRates$ also statistically significant compared with the OLS results. For $Patents$ and $FwdCites$ the IV coefficient estimate is around twice as large as the OLS estimate for the knowledge-weighted embodied technology imports. The coefficient estimates is also larger for production-weighted embodied technology imports. The coefficient estimates remain larger for the knowledge-weighted embodied technology imports consistent with our expectations that patent citations more accurately reflect the relevant knowledge for innovating. That said, the estimates for the production-weighted embodied technology imports are positive and statistically significant

Table 3: Innovation Outcomes

	OLS			IV		
	(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)
$\ln(EmbTechDiag)$	0.000	-0.000	0.002	0.061***	0.078***	0.009
	(0.000)	(0.001)	(0.001)	(0.010)	(0.014)	(0.011)
$\ln(OwnTech)$	0.011***	0.031***	-0.011***	0.011***	0.030***	-0.011***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Observations	478,880	478,880	361,290	478,880	478,880	361,290
F-Stat1				4467	4467	10890
F-Stat2				15029	15029	20754
F-Stat3				396	396	483

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where *Outcome* is the variable specified on column titles. Other controls include Lags of log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

indicating that the production process captures some element of the relevance of knowledge linkages.

The remaining coefficient estimates are in line with expectations. The coefficients for $OwnTech$ are positive for the patenting rate and negative for the quality of patents. This suggests that, at the margin, a higher stock of knowledge corresponds to more low-quality innovators, perhaps due to selection effects. The coefficient for $EmbTechDiag$ is positive in the IV estimates suggesting that additional spillovers from same sector imports. We do not focus on the spillovers through this channel for the aforementioned difficulties with interpretation. That said, the positive and relatively large coefficient estimate as well as the overall important of own-sector linkages (Section 4) imply that the first two coefficient estimates should be taken as a lower bound of spillovers from embodied technology imports.

Table 4 presents the main diffusion results for both the OLS and IV estimates. Since the signs in the IV estimates are the same as the OLS and the bias is in the same direction as in Table 3, we discuss the two sets of results together. The coefficient estimates for $USBackCites$ are positive for both the knowledge-weighted and production-weighted embodied technology imports. To some extent, this follows from the previous result since more frequent patenting implies more backward citations to the US, all else equal.

Table 4: Diffusion Outcomes

	OLS			IV		
	(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)
$\ln EmbTechDiag$	0.000	0.003**	0.001**	0.127***	0.037***	0.008***
	(0.001)	(0.001)	(0.000)	(0.021)	(0.013)	(0.003)
$\ln OwnTech$	0.045***	-0.014***	-0.000	0.044***	-0.014***	-0.000
	(0.003)	(0.003)	(0.001)	(0.003)	(0.003)	(0.001)
Observations	478,880	361,290	356,457	478,880	361,290	356,457
F-Stat1				4467	10890	11432
F-Stat2				15029	20754	21440
F-Stat3				396	483	472

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where $Outcome$ is the variable specified on column titles except for $USBackShare$. Other controls include Lags of log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

For $USBackRate$ and $USBackShare$ the IV estimates for the coefficients are statistically insignificant. The results suggest that while knowledge-weighted and production-weighted embodied technology imports improve innovation outcomes, knowledge may diffuse through channels that do not result in citations. That said, the estimates of $EmbTechK$ for $USBackRate$ are more mixed than when looking beyond Table 4. In many of the robustness tables in Appendix C the coefficient estimate becomes statistically significant at standard levels. This is consistent with the relative point estimate for $USBackCites$, which is almost twice as large as the point estimate for $Patents$ in Table 3. However, there is less evidence in favor of the estimates for $USBackRate$ and the production-weighted embodied technology imports. We expect that for both $USBackRate$ and $USBackShare$ the estimated coefficient is biased downwards since our measure of backward citations is noisier than for $Patents$ and $FwdCites$, with $USBackShare$ being the nosiest measure. Specifically, our measure includes citations on any previous US patent, many of which may include innovations that were already well-understood in the domestic economy. Additionally, many countries may not require domestic patents to cite foreign patents making backward citations a noisy measure of diffusion.[34] Another possibility is that innovators learn more broadly from embodied technology imports than from US patents. For example, an innovator may learn about the French innovation that is built on by the US product and cite that patent instead. In this regard, we think of both $USBackRate$ and $USBackShare$ as being relatively strict measures of diffusion.

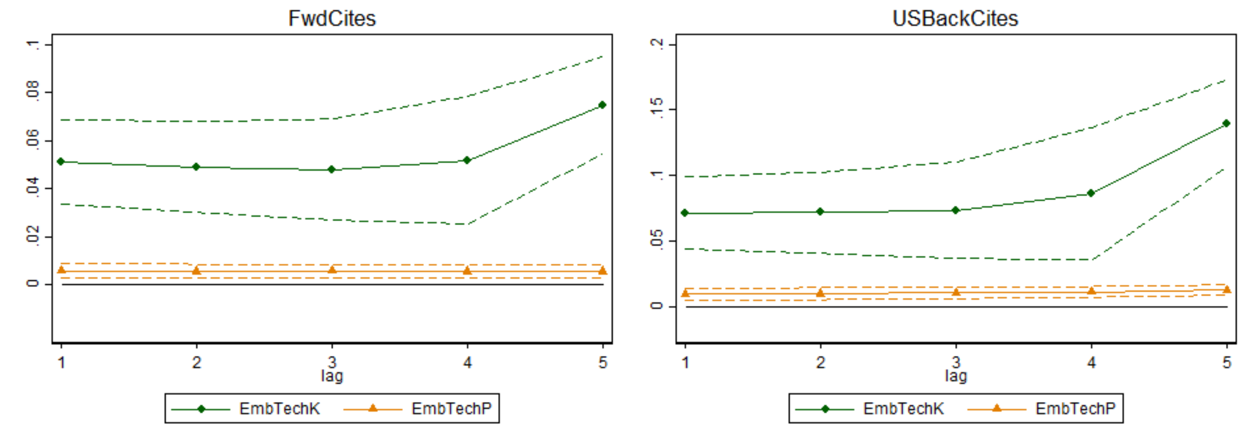
6.2 Lags

The results in Table 3 and Table 4 average the outcome variables over a three-year window, in part, to control for the gradual diffusion of technology. Diffusion is a gradual process that may take several years before knowledge spillovers from embodied technology imports are realized in patentable innovations. For this reason, we estimate Equation (12) with the outcome variable measured at period t (rather than a three-year window) and the regessors taken at lags $\tau \in \{1, 2, ..., 5\}$. Figure 3 summarizes the coefficient estimates for the two measures of embodied technology imports. In all regressions the set of controls are the same as in the baseline regressions. We only present the results for $FwdCites$ and $USBackCites$ in the main text for brevity and provide the remaining figures in Appendix C.

The results at different lags are similar in magnitude to the baseline results, with higher lags being stronger for $EmbTechK$ and relatively stable for $EmbTechP$. There is a similar pattern for coefficient estimates related to the other innovation and diffusion outcomes, where the statistical significance of the estimated coefficients are similar to the baseline results. We view this as supportive evidence for averaging the outcome variables in the baseline results over a three-year window since the results point to diffusion being relatively gradual. Consistent with this view, we find larger, albeit quantitatively similar, point estimates if instead the outcome is averaged over a five-year window (Appendix C).

[34] Our choice to allocate patents based on the location of the innovator would add noise to both measures. For example, a citation to a patent with 50% US innovators and 50% French innovators would not increase $USBackShare$ even if local innovators learned about the patent through interactions with the US innovators US employer.

Figure 3: Main Outcomes Estimated at Different Lags



(a) *FwdCites* (b) *USBackCites*

Notes: Coefficient estimates for *FwdCites* and *USBackCites* for five models, Equation (12), in which the lag is set to $\tau \in \{1, 2, ..., 5\}$. Outcome variables are calculated as there value in period t.

6.3 Quantitative Significance

To further understand the quantitative magnitude of the results, we examine the relative magnitude of the variation in the outcomes and the implied variation of outcomes attributable to the model variables. The results are summarized in Table 5. We focus on the residualized standard deviation (RSD) of the explanatory variables which is calculated as the standard deviation of the variable after removing the effect of all other regressors as well as fixed effects used in the baseline specification. The variation in the outcome variables is captured by the RSD of the outcome variable in question after removing the effect of the fixed effects. We do this to remove both variable trends as well as cross-country and cross-sector variation in the variables. These differences are important for both the outcomes (e.g., increases in patenting over time) and embodied technology imports (e.g., increase in trade over time). However, these trends are not important for understanding the economic significance of the coefficient estimates.[35] Using the residualized variables, we calculate the RSD implied by the model estimates and scaled by the total RSD of the outcome variable.

The table shows that, together, *EmbTechK* and *EmbTechP* explain 9% of the residualized variation in *Patents* and slightly less for *FwdCites* and *USBackCites*. Additionally, *EmbTechK* explains around 2% of the variation in *FwdRate*. Consistent with the earlier sum-

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

Table 5: Quantitative Significance

		Coefficient		RSD		Relative Implied RSD (%)	
Outcome	RSD	$EmbTechK$	$EmbTechP$	$EmbTechK$	$EmbTechP$	$EmbTechK$	$EmbTechP$
$Patents$	0.187	0.041	0.006	0.368	1.201	8.1	1.1
$FwdCites$	0.317	0.059	0.006	0.368	1.201	6.8	0.7
$FwdRate$	0.481	0.024	0	0.368	1.201	1.8	0
$USBackCites$	0.425	0.081	0.011	0.368	1.201	7.0	0.9
$USBackRate$	0.606	0.012	0.002	0.368	1.201	0.7	0.1
$USBackShare$	0.142	-0.001	0	0.368	1.201	-0.2	0.1

Notes: RSD refers to residualized standard deviation and is calculated as the standard deviation of the variable after controlling for the fixed effects used in the baseline specification, Equation (12), for the embodied technology import measures and the other regressors for the outcome variables. Coefficient estimates are from Table 3 and Table 4. Relative Implied RSD is calculated as the estimated coefficient multiplied by RSD of the *EmbTechK* and *EmbTechP* divided by the RSD of the outcome variable.

mary statistics, the table also shows that 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. The net impact of a one RSD shock to the production-weighted embodied technology imports is around one seventh the magnitude of the knowledge-weighted embodied technology imports, consistent with our earlier takeaway that the technology spillovers are mostly through knowledge linkages.

6.4 Robustness Checks

We conclude this section with a discussion of the robustness of the main results to alternative specifications. Overall, we find similar coefficient estimates. The tables are provided in Appendix C.

Alternative Instruments Our baseline instrument isolates US supply shocks by examining US exports to all countries outside of a country's cluster. We construct the cluster as the set of countries that fall in the same quintiles of GDP-per-capita and total trade (exports plus imports) to GDP. We also construct alternative instruments using both the traditional leave-one-out instrument, which can be viewed as a cluster with a single country, and an instrument using all other countries within the country's cluster. In both cases, the results hold with similar significance as our baseline results. In the latter case, the coefficient estimates for $USBackRate$ is also positive and statistically signficant.

Country Sample We also consider an alternative specification where we restrict the sample of countries to the top 40 countries based on the total number of patents. We find larger point estimates for the main outcomes, which we take as being suggestive that patenting does not fully capture innovative activity in many lower patenting countries. Additionally, in many of these countries, patenting as an institution may be prohibitively expensive, provide insufficient protection, or simply be underdeveloped.

Alternative Knowledge Transformations In our baseline specification, we take one plus embodied technology to avoid excluding zero-valued observations. We find similar results using both the log variable (excluding zeros) and the *asinh* transformation of embodied technology. In the former case we also find that the coefficient estimate for $USBackCites$ is positive and statistically significant.

Alternative Knowledge Stocks In the baseline variables, we construct knowledge stocks $K_{i,t}^{h}$ using the forward citations rates. Forward citations allow for a control on the relative quality of patents in the measure of knowledge stocks. We show that the results are robust to constructing knowledge stocks with raw patent counts.

Other Results and Controls We find similar results when we focus on alternative outcomes, such as including the own-sector outcomes in our diffusion measures and limiting our outcomes to triadic patents. In both cases we also find that $USBackRate$ is positive and statistically significant. We also find similar results for $Patents$, $FwdCites$ and $USBackCites$ when we restrict country-sector-year observations with positive 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 a measure of the stock of technology embodied within sectors' goods and data on sector-level trade flows between countries to construct measures of knowledge-weighted and production-weighted technology embodied in imports. We show that increases in both measures of embodied technology lead to higher rates of innovation in an importing country-sector pair.

Our results point to important directions for future research, including towards developing a better understanding of the mechanisms underlying the trade channel of technology diffusion. For example, since knowledge linkages are a more important source of diffusion than production linkages and the sources of knowledge linkages are distinct from the sources of production linkages, diffusion through trade of goods may not primarily occur within the firm-to-firm relationships that underpin our sector-level import data and instead there may be spillovers to other firms in importing countries. Future work using firm-level data can investigate the presence of these spillovers through knowledge IO linkages. The estimated elasticities in this paper could also 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), 357–74.

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.

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.[36]

Appendix A.1 Correlations of IO Linkages

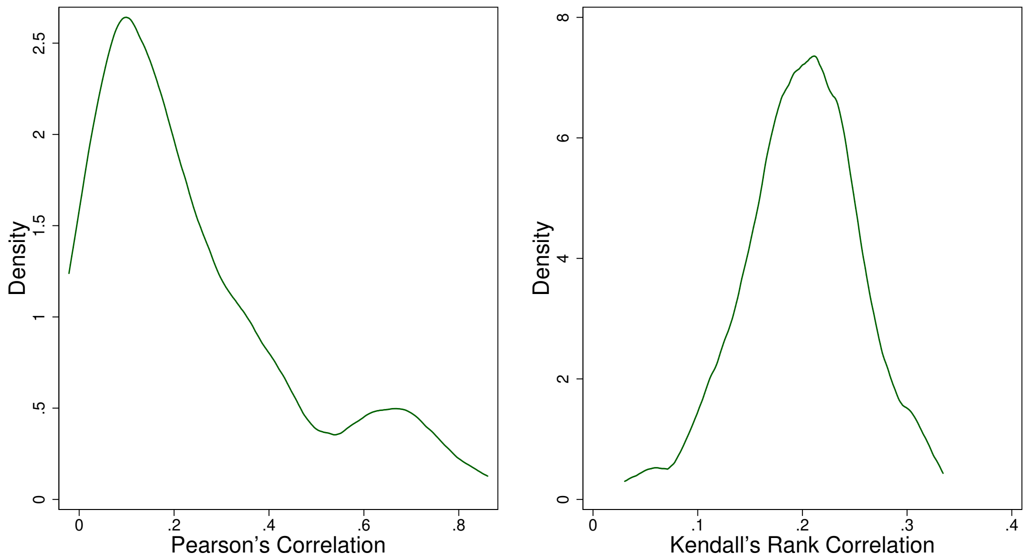
Our empirical analysis, which compares the effects of imports of technology embodied in knowledge and production inputs on patenting outcomes, depends to a large extent on there being distinct variation in the sources of those inputs for the average sector in order to draw the inferences that we do. 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 is not driving these low average correlations. We now state our first observation regarding the comparison of the knowledge and production IO tables.

[36] 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

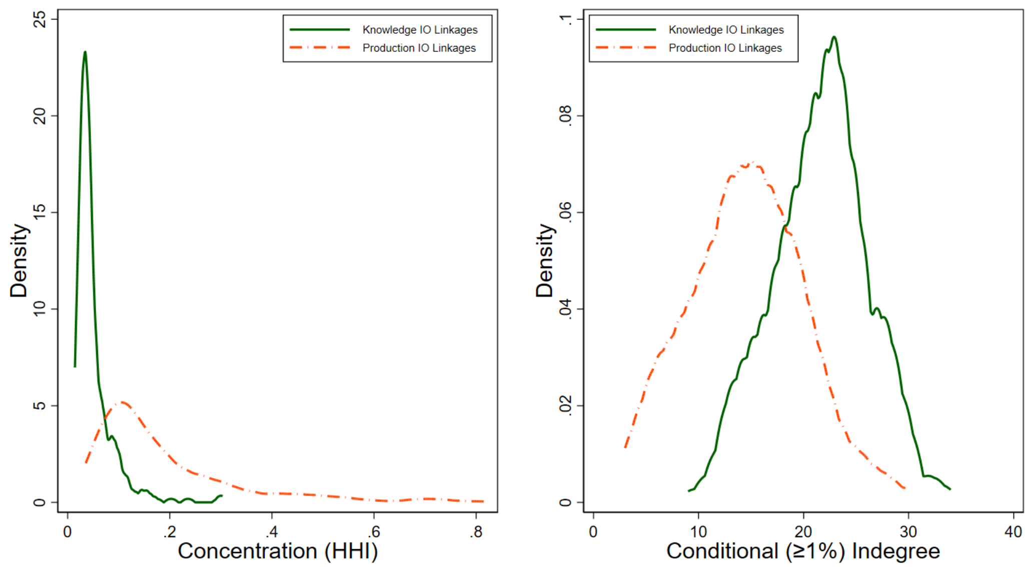
Notes: Table reports summary statistics of the distributions of correlation coefficients of IO linkages for the 291 output sectors plotted in Figure 1. Pearson is the linear correlation between knowledge and production IO linkages. Kendall is the rank correlation (adjusted for ties) of the knowledge 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.[37] 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

[37]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 $l \neq h$.[38] 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

[38]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.[39] In our context, these measures are defined by

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

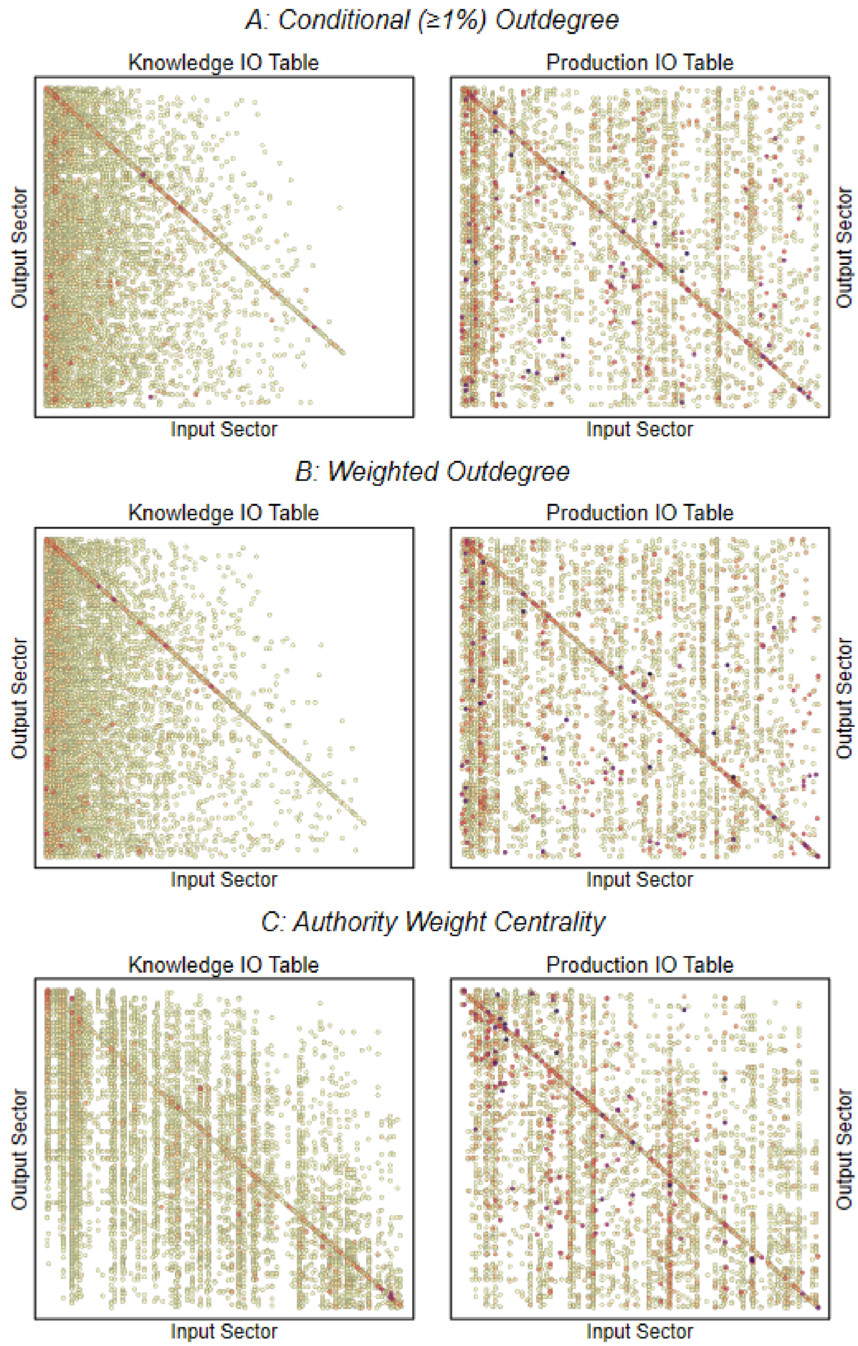
where λ_K (λ_P) and μ_K (μ_P) are the Euclidean norms of the vectors of $\{\text{AWC-K}^{l}_{US,2002}\}_{l\in\mathcal{H}}$ ($\{\text{AWC-P}^{l}_{US,2002}\}_{l\in\mathcal{H}}$) and $\{\text{HWC-K}^{l}_{US,2002}\}_{l\in\mathcal{H}}$ ($\{\text{HWC-P}^{l}_{US,2002}\}_{l\in\mathcal{H}}$), respectively.

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

[39] 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.[40] 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.[41] 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.[42]

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).[43] 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

[40]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.

[41]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.

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

[43]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. 71.8 million patent families have filing dates, while only 20.9 million have inventor country information.[44] 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.[45] 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.[46] 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.[47] In total,

[44]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.

[45]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.

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

[47]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.[48]

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.[49] 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.

[48]Only 17k focal patents cite patents that do not have IPC code data.

[49]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.[50] This second concordance uses equal weights for each 1992 6-digit HS code into which a given 2002 6-digit HS code maps.[51]

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.[52] The second concordance converts 1987 4-digit SIC codes into 2002 6-digit NAICS codes, again using an unweighted mapping between the classifications.[53] 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.[54] The third concordance applies the mapping of 2002 6-digit NAICS codes into the endpoint 2002 BEA sector codes.[55] The composite weights between 1992 6-digit HS codes and our endpoint classification implied by combining the three concordances of this second stage are precisely the weights mapping sub-sectors into sectors referred to in

[50]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`.

[51]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 labeling, this is the H2 to H0 concordance file).

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

[53]This file is available from the US Census Bureau at `https://www.census.gov/naics/?68967`.

[54]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`.

[55]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`.

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.[56] 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.[57] 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.[58] The second and third concordances map 2012 NAICS sectors into 2007 NAICS sectors and 2007 NAICS sectors into 2002 NAICS

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

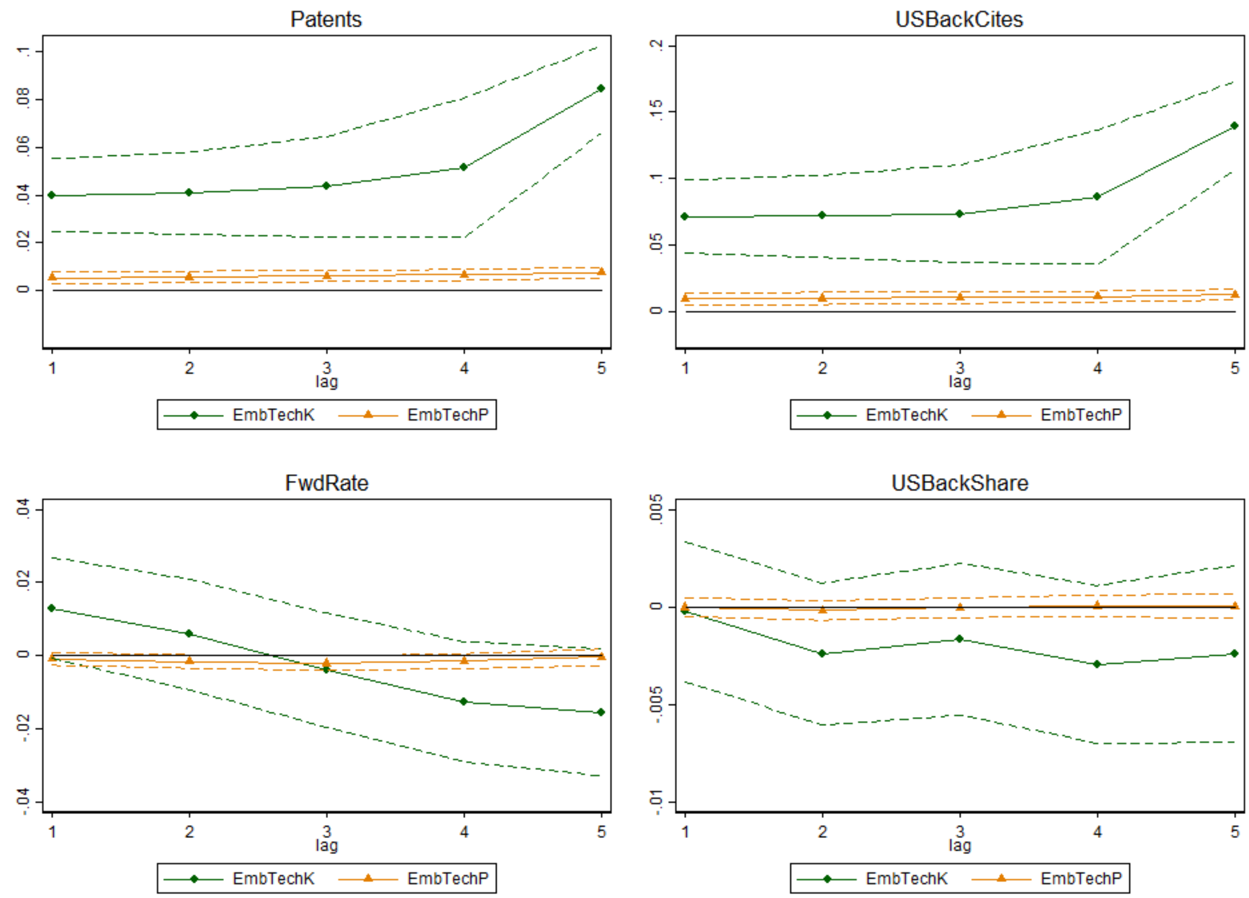
[57] The first of these concordances is available at `https://www.bea.gov/industry/benchmark-input-output-data` using the appendices of the 1997 Use table (after redefinitions) while the second concordance is available at `https://www.census.gov/naics/?68967`.

[58] This is available in the appendix of the 2007 Use table found at `https://www.bea.gov/industry/input-output-accounts-data`.

sectors, respectively.[59] 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.

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

Appendix C Additional Figures and Tables



Appendix Table C.1: Five-Year Average for Outcomes

5-year Forward Averages of Outcomes						
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	4467.582	4467.582	8225.481	4467.582	8225.481	8672.958
F-Stat2	15029.801	15029.801	19520.15	15029.801	19520.15	20247.786
F-Stat3	396.829	396.829	469.325	396.829	469.325	464.055

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where $Outcome$ is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

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

IV: Leave-One-Out from Total US Exports to World						
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	4477.005	4477.005	10924.322	4477.005	10924.322	11472.592
F-Stat2	15031.43	15031.43	20786.991	15031.43	20786.991	21475.515
F-Stat3	395.742	395.742	482.88	395.742	482.88	471.943

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where $Outcome$ is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

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

IV: Leave-One-Out from Total US Exports to Country Cluster

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	1195.037	1195.037	1106.028	1195.037	1106.028	1110.95
F-Stat2	1217.131	1217.131	2050.855	1217.131	2050.855	2254.552
F-Stat3	100.917	100.917	77.69	100.917	77.69	78.400

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where *Outcome* is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

Appendix Table C.4: Top 40 Countries by Total Patents

Top 40 Patenting Countries

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	5753.031	5753.031	14376.865	5753.031	14376.865	15861.01
F-Stat2	22155.752	22155.752	27403.875	22155.752	27403.875	28348.766
F-Stat3	366.13	366.13	418.4	366.13	418.4	415.653

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where *Outcome* is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

Appendix Table C.5: $log(Y)$ Transformation

	Data Transformation: log(Y)					
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	10890.478	11853.41	11853.41	11759.435	11759.435	11432.904
F-Stat2	20754.325	23813.57	23813.57	23653.542	23653.542	21440.735
F-Stat3	483.847	469.661	469.661	465.975	465.975	472.995

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where *Outcome* is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

Appendix Table C.6: $asinh(Y)$ Transformation

	Data Transformation: asinh(Y)					
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	4467.582	4467.582	10890.478	4467.582	10890.478	11432.904
F-Stat2	15029.801	15029.801	20754.325	15029.801	20754.325	21440.735
F-Stat3	396.829	396.829	483.847	396.829	483.847	472.995

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where *Outcome* is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

Appendix Table C.7: Other Diffusion Outcomes

	Other Diffusion Outcomes						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Inc. Domestic	Inc. Diagonal	Inc. Diagonal	Inc. Diagonal	Triadic	Triadic	Triadic
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$	
$\ln EmbTechK$	0.001	0.071***	0.018**	-0.002	0.041***	0.016*	0.001
	(0.002)	(0.013)	(0.008)	(0.002)	(0.008)	(0.010)	(0.002)
$\ln EmbTechP$	0.000	0.009***	0.000	0.000	0.006***	-0.000	0.000
	(0.000)	(0.002)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Observations	356,779	478,880	361,290	359,483	478,880	259,847	259,662
F-Stat1	11299.02	4467.582	10890.478	10864.852	4467.582	13555.713	13622.151
F-Stat2	21595.063	15029.801	20754.325	20935.564	15029.801	34681.241	34783.334
F-Stat3	473.867	396.829	483.847	477.964	396.829	398.009	396.711

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: ln(1+*Outcome*), where *Outcome* is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

Appendix Table C.8: Knowledge Stocks Construct Using Patent Count

	Knowledge Stock Constructed Using Patent Counts					
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	3576.401	3576.401	7894.512	3576.401	7894.512	8131.503
F-Stat2	11513.898	11513.898	16366.745	11513.898	16366.745	16735.513
F-Stat3	382.703	382.703	536.722	382.703	536.722	528.105

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: ln(1+*Outcome*), where *Outcome* is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year

Appendix Table C.9: Non-Zero Patent Sample

	Non-Zero Patents Sample					
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	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-Stat1	10890.478	10890.478	10890.478	10890.478	10890.478	11432.904
F-Stat2	20754.325	20754.325	20754.325	20754.325	20754.325	21440.735
F-Stat3	483.847	483.847	483.847	483.847	483.847	472.995

Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where *Outcome* is the variable specified on column titles. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, log total exports to world and log total imports to world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. All standard errors are clustered twoways: Country*Sector and Sector*Year