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

February 20, 2023

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 imports 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 discussants, Anna Ignatenko and Simone Lenzu, 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 traded. These product flows potentially convey information about the innovations embodied within them to the users of the products. Second, the foundational knowledge on which new innovations are based originates from many distinct sectors; these sources vary across sectors and need not be related to sectors' sources of production inputs. Since countries' patterns of trade depend in part on patterns of comparative advantage, their imports of technology embodied in trade flows affect innovation in different sectors in those countries in different ways. Incorporating variation in sectors' sources of knowledge and production inputs is necessary to assess the impacts of a given amount of technology embodied in a set of trade flows on different sectors. 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 technology imports on innovation and diffusion outcomes. The channel underlies many

[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 patents data includes Bloom et al. (2016), Bombardini et al. (2018), Autor et al. (2020) for import competition, and Coelli et al. (2020) and Aghion et al. (2021b) for market access.

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 patents 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 import data 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 empirical 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 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 US 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, that the sectors that are key economy-wide sources of inputs differ between the knowledge and production IO tables, and that the sources of knowledge inputs are more persistent than the sources of production inputs. We also find that knowledge IO linkages

[3]Though not the focus of our paper, we are among the first to provide a descriptive comparison of the knowledge and production IO tables of an economy. Concurrent work in Hötte (2021) and Liu and Ma (2022) construct similar knowledge and production IO tables and compare them.

are less persistent on average in sectors in non-US countries than in the US.

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 exclude the own-sector component in the construction of embodied technology as imports and innovation activity in a country-sector can be correlated due to domestic demand shocks or import competition effects. 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 patent 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 the US knowledge stock, each country's own knowledge stock—constructed using country-specific knowledge IO linkages—and own-sector imports. Additionally, our long panel of data, spanning from 1995 through 2015, allows us to control for high-dimensional fixed effects. We include country-sector fixed effects to account for differences in the propensity to patent and innovate across countries and sectors, country-year fixed effects to account for changes in patenting markets and development within countries over time, as well as aggregate sector-year fixed effects to control for structural changes in sectoral technologies over time.

A potential concern with estimating the effects of the trade channel of technology diffusion on domestic innovation is that domestic shocks to future potential profits or shocks to R&D productivity can lead to R&D activity in a country-sector and the demand for production and knowledge intermediate inputs being correlated. [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

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

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

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 patenting by around 0.67%. In comparison, a 1% increase in the production-weighted embodied technology imports increases patenting by 0.003%. To quantify the size of the of the estimated coefficients, we show that a one standard deviation increase of the residualized of knowledge-weighted and production-weighted embodied technology imports account for a 13.2% and 0.06% standard deviations of residualized patenting. The considerably larger estimate of the knowledge-weighted measure is consistent with our expectation that the knowledge IO table better 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 a similar proportion for the knowledge-weighted measure and 0.005% for the production-weighted measure. Knowledge-weighted and production-weighted embodied technology imports account for 8.8% and 0.04% 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.

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, consistent with the idea that patenting activity is better measured in these countries or more representative of innovation activity. Finally, we find similar results using alternative constructions of the main variables, alternative innovation and diffusion outcomes, or other controls.

[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. Our results are robust to using this instrument instead.

Related Literature. Our work contributes to the literature on the channels of international technology diffusion (most recently surveyed by Keller, 2021), particularly 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 concurrent work by 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.

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 is Cai et al. (2022) who examine inter-sectoral and cross-country technology diffusion. In their model, technology diffuses exogenously within and across borders based on parameters estimated using citation linkages from the US patent and trademark office. We show that this diffusion across countries depends on endogenous trade flows between countries.

The empirical approach we take to evaluate the effects of diffusion of technology across countries is complementary to recent work using 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).[7] To the best of our knowledge, ours is the first paper to include inter-sectoral knowledge IO measures based on these data to estimate the trade channel of technology diffusion. Fons-Rosen et al. (2019) use patents-based sector-pair measures of technological similarity adapted from Bloom et al. (2013), which are distinct from our citations-based IO measures, to investigate the foreign direct investment channel of technology diffusion. Berkes et al. (2022) show that there has been a large increase in international knowledge spillovers since the 1990s as measured by cross-country patent citations and that the innovations induced by this increase in diffusion lead to an increase in the growth rates of sectoral output per worker and total factor productivity. Closely related is the empirical exercise in Liu and Ma (2022) that documents that global spillovers from past patenting activity that depend on the network of patent citations across countries and sectors lead to increases in innovation.

[7]We also build on work examining the inter-sectoral patterns of knowledge flows in single-country settings, such as Acemoglu et al. (2016) and Cai and Li (2019).

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.

Outline. The remainder of this paper proceeds as follows. Section 2 describes the data used in our analysis. Section 3 presents the conceptual framework used to guide our empirical analysis. Section 4 describes the constructions of the knowledge and production IO tables. Section 5 describes our empirical strategy and baseline 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,

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

which are represented by International Patent Classification (IPC) codes; the set of inventors of the patent application and their countries of residence; and citations to other patents listed in the patent application.[9] Throughout our analysis, we focus on patent applications rather than patent grants as grant dates are unavailable in the Google Patents database for patents applied for at many national patent offices, whereas application dates are available.[10] Furthermore, as we examine technology diffusion and its effects, patent application events better reflect the timing of diffusion than do patent grant events.

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

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

Inter-sectoral input purchases. To measure production input-output relationships, we employ the Bureau of Economic Analysis (BEA) Supplementary Use Tables. These tables

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

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

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

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

[13] We also include patents applied for at the JPO, the USPTO, and 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).

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.

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

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

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

[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.

For the first stage, we convert the data on patent counts, forward citations, stocks of knowledge (the measurement of which we describe in Section 5.2), 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]

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

3 Conceptual Framework

Before turning to our empirical analysis, we describe a stylized conceptual framework to guide our analysis. Time is discrete and indexed by t. The economy is populated by a unit mass of identical 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

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

[16] Related papers that use the concordances introduced by Lybbert and Zolas (2014) and extended to other classifications in Goldschlag et al. (2020) include Kukharskyy (2020) and Hötte (2021), among others. Kukharskyy (2020) uses the concordances with citations data to construct 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 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.

structure and the empirical approach described in Section 5, we define three levels of sectoral aggregation, where we denote n as a summary sector (the highest aggregation), h as a sector (the focus of our analysis), and p as a subsector (or product). We also define the sets $\mathcal{P}^h$ as the set of subsectors p in sector h and $n(h)$ as the summary sector n that contains sector h.[18]

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.[19] 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 profits in $\bar{\pi}^{n(h)}_t$ to be consistent with our empirical specification. A firm (i, h, t) that invests $R^h_{i,t}$ into R&D produces innovations in the next period at rate

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

where $\psi^h_{i,t}$ governs the relative cost of R&D across country-sector-years, $Z^h_{i,t}$ is the domestic stock of relevant knowledge for sector h, and $S^h_{i,t}$ is a spillover from the frontier of knowledge (described below). 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. Alternatively, the variable $X^h_{i,t}$ could be interpreted as the average quality of innovations or the quality-adjusted rate of innovations.

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

$$Z^h_{i,t} = \prod_l \mathcal{G}_Z \left(\sum\nolimits_{p \in \mathcal{P}^l} K^p_{i,t}\right)^{\kappa^{l,h}_{i,t}}$$

where $\mathcal{G}_Z(\cdot)$ is a monotonic function that dictates the strength of spillovers from domestic knowledge in an input sector, which we set to $\mathcal{G}_Z(x) = (1+x)^{\eta_Z}$.[20] The other variables are the knowledge stock $K^p_{i,t}$ of country-subsector-year (i, p, t) and the relevance of knowledge

[18] $\mathcal{P}^h$ can be thought of as the set of products that are associated with a sector h. Our raw trade data is collected at the product level and we first convert our patent data to this level of aggregation. We introduce the summary sector $n(h)$ since we use fixed effects at this level of aggregation in our empirical application. These details are further explained in Section 5.

[19] 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 proportional to expected profits.

[20] This specification of $\mathcal{G}_Z(x)$ is consistent with our treatment of zeros in the empirical analysis.

from sector l for producing innovations in country-sector-year (i, h, t), denoted by $\kappa_{i,t}^{l,h}$.

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

$$S_{i,t}^{h} = \prod_{l} \mathcal{G}_S \left(\sum\nolimits_{p \in \mathcal{P}^l} (m_{F,i,t}^{p} \times K_{F,t}^{p}) \right)^{\gamma_{F,t}^{l,h}}$$

where $\mathcal{G}_S(\cdot)$ is a monotonic function that dictates the strength of spillovers from the embodied frontier technology, which we set to $\mathcal{G}_S(x) = (1+x)^{\eta_S}$. The value of $\gamma_{F,t}^{l,h}$ captures the usefulness of knowledge from sector l to sector h in the frontier economy in period t. We allow for this parameter to change over time to capture dynamics in knowledge linkages over time. 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 different types of goods result in knowledge spillovers to different countries.

The spillover of ideas from the frontier economy depend on two components: $K_{F,t}^{p}$ is the frontier stock of knowledge in subsector 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 subsector 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 $m_{F,i,t}^{p}$. Intuitively, embodied technology can be thought of as capturing the probability that a domestic innovator encounters a frontier good by chance multiplied by the probability that the innovator realizes a new innovation given the encounter (as in, for example, Bloom et al., 2013; Lucas Jr. and Moll, 2014; Perla and Tonetti, 2014; Buera and Oberfield, 2020). Whenever $\gamma_{F,t}^{l,h} > 0$ and $\eta_S > 0$, increased domestic abundance of frontier goods in subsectors $p \in \mathcal{P}^l$ (higher $m_{F,i,t}^{p}$) and increased technology embodied in those goods (higher $K_{F,t}^{p}$) both increase innovation in sector h.

The problem of a firm is to maximize net expected profits 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

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

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

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

where $f_{i,t}=(\ln\bar{\pi}_{i,t}-\ln\psi_{i,t})/(\zeta-1)$, $f_{t}^{n(h)}=(\ln\bar{\pi}_{t}^{n(h)}-\ln\psi_{t}^{n(h)})/(\zeta-1)$, $f_{i}^{h}=(\ln\bar{\pi}_{i}^{h}-\ln\psi_{i}^{h})/(\zeta-1)$, and $\epsilon_{i,t}^{h}=(u_{i,t}^{h}-v_{i,t}^{h})/(\zeta-1)$. Substituting our functional forms for the domestic and frontier spillovers, $\mathcal{G}_S$ and $\mathcal{G}_Z$, we get:

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

where

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

The expression in (2) provides the foundation for our empirical strategy. The parameters, η_S and η_Z, are the elasticities of the spillover functions $\mathcal{G}_S$ and $\mathcal{G}_Z$ respectively. The rest of the paper focuses on identifying these elasticity parameters which modulate the impact of frontier and domestic knowledge spillovers on innovation outcomes of a country. In the next sections, we construct variables that correspond to the values of *EmbTech* and *OwnTech*, which stand for "embodied technology imports" and "own-country technology" respectively.

The conceptual framework highlights the relationship between embodied technology imports and innovation outcomes. 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 Inter-Sectoral Technology Linkages

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

4.1 Knowledge and Production Input-Output Tables

Our analysis estimates the effects of embodied technology imports on patenting outcomes. We focus on two 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 tightly linked with our focus on innovation outcomes since patent citations represent a direct report of flows of technology. The second is *production input-output linkages*, which describe the relative flow of intermediate inputs across sectors. While less directly linked to innovation outcomes, the use of intermediate inputs captures another channel through which technology can diffuse within and between sectors. We collect these measures of technology relevance into separate knowledge and production IO tables and document patterns of inter-sectoral technology flows.

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

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

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 $\kappa_{i,t}^{l,h}$ denote the knowledge IO linkage between sectors l and h in country i in year t. 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

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

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.[22] The knowledge IO linkages capture the country-sector (i,h) citations made by patents filed over a ten-year window to all prior sector l patents from all countries as a share of total citations made by country-sector (i,h) patents filed over the ten-year window.

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

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

where $\text{Sales}_{i,t}^{l,h}$ is the total value of sector l goods sold to country-sector-year (i,h,t). Production IO linkages measure, for year t, the sales from sector l to sector h as a share of the total sales from all sectors to sector h.

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

[22] For example, Berkes et al. (2022) and Ayerst (2022) find that ICT sectors have become more important sources of innovations over the period of our analysis, highlighting the need for dynamic IO linkages. Baslandze (2018) finds an overall increase in the interconnectedness of sectors over this period.

in exposure variation is applied to both knowledge and production IO linkages.[23]

4.2 Description of Knowledge and Production IO Tables

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

Both knowledge and production IO linkages take on values between zero and one. Values closer to one indicate stronger relationships whereas values further 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 $\kappa^{l,h}_{US,2002}$ represented in the left panel and $\rho^{l,h}_{US,2002}$ 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 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 from Figure 1 is that there are clear differences in the patterns of knowledge and production IO linkages for many sectors. We formalize and build on this visual intuition through five descriptive observations that highlight the differences between the knowledge and production IO tables.[24]

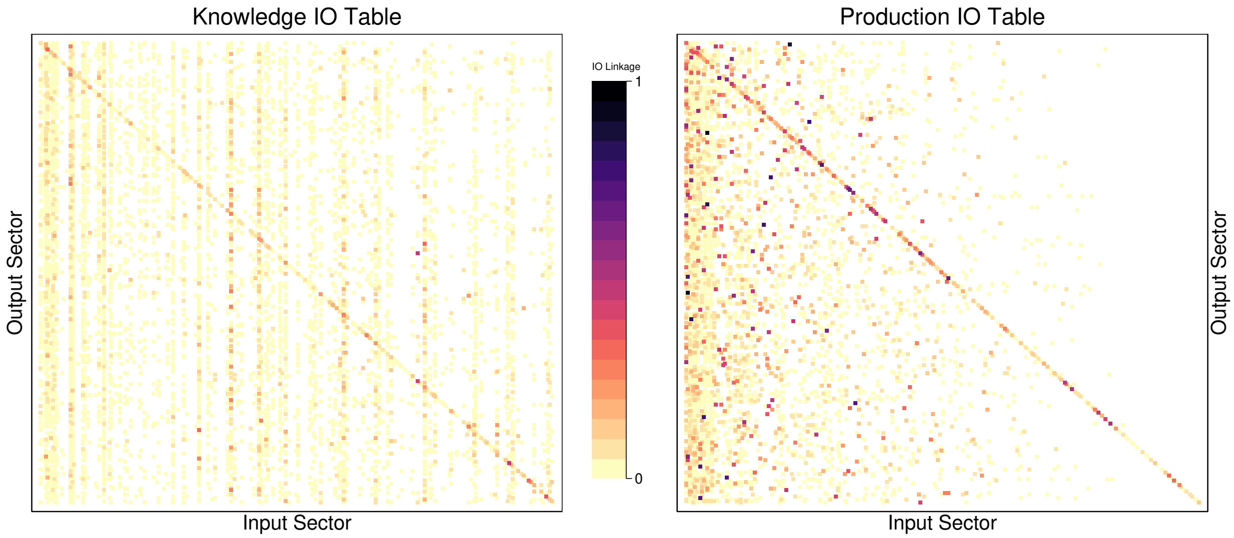
Observation 1*: The sources of knowledge and production inputs are not highly correlated for the average sector in the US.*

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

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

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

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 292 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.

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

Observation 4: *For the average sector in the US, the sources of knowledge inputs are more highly correlated across time than the sources of production inputs.*

Observation 5: *The sources of knowledge inputs are less persistent on average in non-US country-sectors than in US sectors. In part, this reflects a convergence on average between each non-US country-sector and the same sector in the US.*

We relegate the elaboration of these observations to Appendix A as a comparison of the IO tables is tangential to our main objectives. That said, a key implication of the observations is that the knowledge and production IO structures of the economy capture different relationships between sectors and, consequently, may capture different potential sources of technology spillovers. Furthermore, although US knowledge IO linkages tend to be highly persistent, there is a considerable degree of variation over time in US production IO linkages (as well as in non-US knowledge IO linkages). In our baseline analysis we explore the diffusion of knowledge through imports of embodied technology weighted using both dynamic knowledge IO linkages and production IO linkages.

The approach used to measure knowledge IO linkages in this paper shares many similarities with the method developed concurrently in Liu and Ma (2022). Like them, we construct knowledge IO tables for many countries across a long period of time using data from Google Patents and provide a descriptive analysis of the IO tables. Although both papers use different sources of data to construct production IO tables, both document a low correlation between the knowledge and production IO tables across all IO linkages in the tables. The analysis in this paper goes one step further to show that this low correlation prevails across most output sectors and at different points in time.[25] Although Liu and Ma (2022) describe their global knowledge IO table as being stable across time and across countries, our more disaggregated analysis reveals that there are many non-US country-sectors for which the sources of knowledge inputs change substantially across our sample period. Similarly, while Liu and Ma (2022) find that the knowledge IO tables in most top patenting countries are highly correlated with the US knowledge IO table when using patents from all years, our analysis suggests that there is considerable variation across country-sectors and across time in the correlation of US and non-US knowledge IO linkages.

5 Empirical Specification

We now describe the main empirical specification of our analysis and the construction of key variables. Following Our main regressions involve regressing innovation outcomes on measures of embodied technology imports. We start by specifying our outcome variables. We then use the knowledge and production IO tables to develop the 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.

The US as the technological frontier. Throughout the analysis, we focus on the effects of imports from the US, which we consider to be 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,

[25]Although the comparable regression-based analysis in both papers is undertaken at the level of output sectors and not countries, Liu and Ma (2022) do not document patterns of variation across output-sectors and across time in the knowledge IO tables other than for domestic citation shares.

for evidence), which reflects the characteristics of the frontier economy. Second, setting the frontier economy to the US allows us to be consistent with our data measurement, since the US has detailed data available for multiple years to construct both the production and knowledge IO tables described in Section 4, which are necessary for our analysis.

Sample Description. The unit of observation in our analysis is a country-sector-year. We limit our final panel of data to the years 1995 to 2015. Years prior to 1995 lack 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 their population, triadic patenting activity, and export-to-GDP and import-to-GDP ratios. We do this to avoid including countries where patenting outcomes may be too noisy, or countries that act as intermediaries in the global trade network. Appendix B provides the details of our sample selection procedure.

5.1 Outcome Variables

Table 1 summarizes, defines, and describes the main outcomes variables. The conceptual framework illustrates the relationship between frontier knowledge spillovers and innovation outcomes. 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 ($\boldsymbol{Patents}$) and the citation-weighted rate of patenting ($\boldsymbol{FwdCites}$). We also look at the average quality of patents, measured by the number of forward citations per patent ($\boldsymbol{FwdRate}$), to examine both the intensive and extensive margin effects of frontier knowledge spillovers.

We also examine 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 measuring total backward citations to US patents ($\boldsymbol{USBackCites}$), the per-patent rate of backward citations to US patents ($\boldsymbol{USBackRate}$), and the share of backward citations to US patents in the total backward citations to foreign patents ($\boldsymbol{USBackShare}$).

In our baseline specification, we measure outcome variables using the average of the variables in the three-year window between year t and $t + 2$. Panel A of Table 2 presents summary statistics of the main outcome variables. $FwdRate$ and $USBackRate$ are calculated only for observations with non-zero patent counts, while $USBackShare$ is calculated only for observations with non-zero cross-sector citations to foreign patents. The summary statistics

Table 1: Description of Outcome Variables

Variable	Description
$Patents^h_{i,t}$	The count of patent applications using the allocation rules described in Section 2.
$FwdCites^h_{i,t} = \sum_{l\in\mathcal{H}} \sum_{j\in\mathcal{I}} \sum_{s=t}^{t+5} Cites^{h,l}_{i,j,t,s}$	Measures total citation-weighted patenting activity. The measure only includes citations received in the five years following the calendar year of the patent's application to mitigate truncation issues that would arise in later periods of the sample if citations received in all years were used.
$FwdRate^h_{i,t} = \frac{FwdCites^h_{i,t}}{Patents^h_{i,t}}$	Measures the average quality of patent applications as the citations received by patents divided by total patents.
$USBackCites^h_{i,t} = \sum_{l\neq h} \sum_{s=0}^{t} Cites^{l,h}_{US,i,s,t}$	The number of cross-sector citations to US patents. This measure excludes own-sector backward citations to be consistent with the focus on cross-sector imports of embodied technology (discussed below).
$USBackRate^h_{i,t} = \frac{USBackCites^h_{i,t}}{Patents^h_{i,t}}$	Measures the rate at which patents cite cross-sector US patents. In this regard, the measure captures the intensive margin of technology diffusion.
$USBackShare^h_{i,t} = \frac{USBackCites^h_{i,t}}{\sum_{l\neq h} \sum_{j\neq i} \sum_{s=0}^{t} Cites^{l,h}_{i,j,s,t}}$	The share of cross-sector citations to US patents in total cross-sector citations to foreign patents. The measure is informative of the extent to which knowledge inputs are substituted towards US knowledge in response to larger embodied technology flows from the US.

Notes: Variables are defined for a country-sector-year (i, h, t). For *FwdRate*, *USBackRate*, and *USBackShare*, we exclude (i, h, t) observations with zero value in the denominator.

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.

5.2 Embodied Technology Imports and Other Controls

Our main variable of interest is the frontier knowledge spillovers through embodied technology imports ($EmbTech^h_{i,t}$ in the conceptual framework). Before turning to our main variables of interest, we discuss the construction of knowledge stocks $K^p_{i,t}$, which are used in the construction of the main variables.

We measure the technological content of a subsector's goods using patents data as

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

where $FwdCites^p_{i,t}$ is the five-year forward citations in country-subsector-year (i,p,t) and δ is the depreciation rate of knowledge that we set to 5% to be consistent with commonly used values. For each country and subsector, we initialize the stock of knowledge in 1940 with the value $K^p_{i,1940} = FwdCites^p_{i,1940}/\delta$. The initial value has little influence on the knowledge stocks used in the period of our analysis because of the long time period.

Embodied technology imports. We set the linkages γ in the conceptual framework to be a combination of knowledge κ and production ρ IO linkages, $\gamma^{l,h}_{F,t} = \alpha\kappa^{l,h}_{F,t} + (1-\alpha)\rho^{l,h}_{F,t}$. We do not impose structure on whether knowledge flows between sectors are better captured by the knowledge or production IO linkages. For each country-sector-year (i,h,t), our measure of knowledge-weighted embodied technology imports is given by

$$EmbTechK^h_{i,t} = \prod\nolimits_{l\neq h}\left(1+\sum\nolimits_{p\in\mathcal{P}^l}(w^l(p)K^p_{US,t}M^p_{US,i,t})\right)^{\kappa^{l,h}_{US,t}}, \tag{5}$$

and production-weighted embodied technology imports is given by

$$EmbTechP^h_{i,t} = \prod\nolimits_{l\neq h}\left(1+\sum\nolimits_{p\in\mathcal{P}^l}(w^l(p)K^p_{US,t}M^p_{US,i,t})\right)^{\rho^{l,h}_{US,t}}. \tag{6}$$

The amount of imported embodied technology depends on the flow of knowledge into country i from every sector l in the United States. 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 abundant and readily available. 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 in the US.

An important distinction from the conceptual framework is that we construct the measures of embodied technology from the levels of US imports ($M^p_{US,i,t}$), rather than trade scaled by output ($m^p_{US,i,t} = M^p_{US,i,t}/Y^p_{i,t}$), since output data is unavailable at the level of aggregation we examine. A potential issue with this construction is that higher imports could simply reflect that the importing country has a larger population or economy.[26] We include granular fixed

[26]It is also worth noting that this is not an issue with the US technology stocks $K^p_{US,t}$ since they are common to all importers and capture the abundance of technology embodied within imports, meaning their levels are important for the interpretation of our results.

effects as a best attempt to deal with this issue.[27]

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. Within-sector demand shocks can lead to countries importing more foreign products to satiate demand while at the same time investing more in innovation activities in the sector due to increased returns. Moreover, own-sector imports can also affect innovation outcomes in a country through import competition effects, since firms may invest more in innovation in order to escape foreign competition. Finally, R&D productivity shocks and profitability shocks to a country-sector can lead to imports and innovations by the country-sector to be correlated. Given these concerns, we include a control for the own-sector component of embodied technology spillovers constructed as

$$EmbTechDiag^h_{i,t} = 1 + \sum\nolimits_{p \in \mathcal{P}^h} w^h(p) K^p_{US,t} M^p_{US,i,t}. \tag{7}$$

We do not scale the knowledge inputs by the IO weights, $\kappa^{l,l}_{US,t}$ or $\rho^{l,l}_{US,t}$, since we expect that this variable also captures factors not directly related to the effects of technology diffusion, such as import competition.[28]

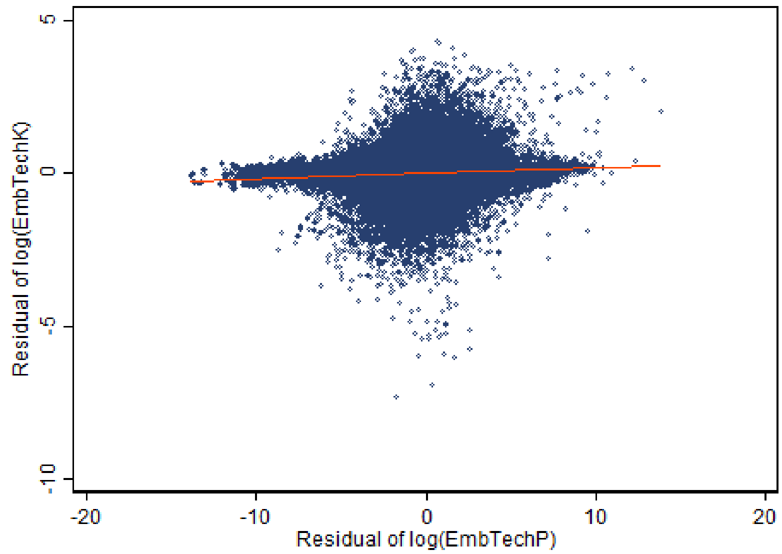
Estimating the separate effects of knowledge- and production-weighted imports of embodied technology requires that there is sufficient variation across observations in our sample in these two measures. To assess this, we regress the logs of both $EmbTechK$ and $EmbTechP$ on the set of fixed effects included in our baseline specifications. Figure 2 plots the fitted residuals from these regressions and a line of best fit from a regression of one set of residuals on the other. The figure highlights that there is considerable orthogonal variation in the two residualized variables.

Other variables. We also include measures of the country-sector domestic and US stocks of knowledge to control for alternative channels of diffusion. We construct the measure of the

[27] 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., trade increases for most country sectors over this period—lead to declining US import shares for most countries.

[28] The evidence on the effect of import competition on innovation activity is mixed. For example, Bloom et al. (2016) find that increased trade with China between 2000 and 2005 led to an increase in patenting activity in European firms that were more exposed to that competition (which was also the case for increased exposure to trade from other low-wage countries). In contrast, Autor et al. (2020) find that import competition due to increased trade with China decreased patenting activity in publicly listed US firms and technology classes more exposed to that competition. 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.

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 5.3.

domestic knowledge stock as

$$OwnTech_{i,t}^{h} = \prod_{l} \left(1 + \sum_{p \in \mathcal{P}^{l}} w^{l}(p) K_{i,t}^{p}\right)^{\kappa_{i,t}^{l,h}}, \tag{8}$$

where $w^l(p)$ is the concordance weight discussed in Section 2. Higher values of $OwnTech$ reflect higher stocks in country-sector-year (i, h, t) knowledge that tend to be cited by sector h. We use the domestic knowledge IO linkages $\kappa_{i,t}^{l,h}$ in the construction because this is the best measure of the relevance of sector l knowledge for producing innovations in country-sector-year (i, h, t).[29]

Analogous to $EmbTechK$ and $EmbTechDiag$, we also control for US knowledge stocks

[29]We do not have highly detailed and dynamic measures of production IO linkages for most countries, which prevents us from constructing a similar measure with $\rho_{i,t}^{l,h}$.

relevant to a sector constructed as:

$$USTech_t^h = \prod_{l \neq h} (1 + \sum_{p \in \mathcal{P}^l} w^l(p) K_{US,t}^p)^{\kappa_{US,t}^{l,h}}, \tag{9}$$

$$USTechDiag_t^h = 1 + \sum_{p \in \mathcal{P}^h} w^h(p) K_{US,t}^p \; . \tag{10}$$

These controls serve two purposes. First, they control for the effect of the growth in frontier knowledge on each sector's innovation activity that is common to all countries irrespective of each country's trade integration with the frontier economy. Since our choice of instrumental variables (described in 5.4) preclude us from including the stricter sector-year level fixed effects, these controls help in capturing part of sector-year trends relevant for knowledge production that are common to all countries. Second, we expect that knowledge diffuses across countries through mechanisms other than trade, and controlling for $USTech$ captures some of these aspects of knowledge diffusion. Diffusion can occur through direct technology licensing, FDI, immigration, and other channels (Keller, 2004, 2010, 2021). Including $USTech$ allows us to separately identify the effect of traded embodied technology from broader technology diffusion across countries. Panel B of Table 2 provides summary statistics for all the technology related variables discussed in this section.

Table 2: Summary Statistics

	A: Outcome Variables					B: Technology Variables			
	N	Median	Mean	SD		N	Median	Mean	SD
$Patents_{i,t}^h$	478880	0.083	0.786	1.384	$EmbTechK_{i,t}^h$	478880	16.090	15.788	3.119
$FwdCites_{i,t}^h$	478880	0.171	1.203	1.858	$EmbTechP_{i,t}^h$	478880	13.129	12.604	4.088
$FwdRate_{i,t}^h$	368950	1.394	1.405	0.745	$EmbTechDiag_{i,t}^h$	478880	14.174	13.425	5.397
$USBackCites_{i,t}^h$	478880	0.403	1.585	2.151	$OwnTech_{i,t}^h$	478880	2.471	3.160	3.003
$USBackRate_{i,t}^h$	368950	2.198	2.129	0.988	$USTech_t^h$	478880	10.988	10.827	1.200
$USBackShare_{i,t}^h$	364724	0.498	0.486	0.194	$USTechDiag_t^h$	478880	9.435	8.998	2.859

Notes: All outcome variables are averaged over the three-year window t to $t+2$. All statistics for outcome variables are calculated on the log of one plus the variable except for the statistics for $USBackShare$. All statistics for the technology variables are calculated on the log of the variable.

5.3 Estimation Equation

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

$$
\begin{aligned}
\ln(1 + Outcome^h_{i,t}) = {} & \theta_1 \ln EmbTechK^h_{i,t-1} + \theta_2 \ln EmbTechP^h_{i,t-1} + \theta_3 \ln OwnTech^h_{i,t-1} \\
& + \theta_4 \ln EmbTechDiag^h_{i,t-1} + \theta_5 \ln USTech^h_{t-1} + \theta_6 \ln USTechDiag^h_{t-1} \\
& + V^h_{i,t-1}\beta + f_{i,t} + f^{n(h)}_t + f^h_i + \epsilon^h_{i,t},
\end{aligned}
\tag{11}
$$

where $V^h_{i,t}$ is a vector of controls that includes own-sector imports from the world and exports to the world and $f_{i,t}$, $f^{n(h)}_t$ and f^h_i are country-year, (summary) sector-year, and country-sector fixed effects. In the baseline regressions we lag all our regressors as well as 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. With the exception of $USBackShare$, we also transform the outcome variable as $\ln(1 + Outcome^h_{i,t})$ to keep observations that have zero innovation/diffusion outcomes.

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

The coefficients of interest are related to model parameters as $\theta_1 = \eta_S\alpha, \theta_2 = \eta_S(1-\alpha), \theta_3 = \eta_Z$, where α is the contribution of knowledge linkages towards cross-sectoral spillovers of technology. The 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.

5.4 Endogeneity Concerns

The fixed effects in Equation (11) control for time-invariant characteristics of country-sector pairs, factors that vary at the country level over time, and sector-year shocks that are common to sectors within a summary sector. Despite the inclusion of these fixed effects, there remain potential endogeneity concerns with our regressors of interest.

One possibility is that variation across country-sector-years in the amount of relevant technology inputs embodied in a country's imports in prior years could reflect demand shocks for those inputs that also directly affect patenting outcomes. For example, shocks to expected profits, captured by $u_{i,t}^h$ in the conceptual framework,[30] would both increase R&D investment but also the imports of intermediate inputs used in the production of goods in (i, h, t).[31] Similarly, shocks to domestic R&D productivity, captured by $v_{i,t}^h$ in the conceptual framework, can give rise to endogeneity concerns. 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 bias on the OLS estimates. If these shocks were serially correlated, there would be a spurious correlation between innovation output and imports of embodied technology in past years arising from the shocks. Since there is no data available on R&D expenditures at the level of sectoral disaggregation used in our analysis, we cannot control for these innovation inputs which may cause an omitted variable bias to affect our estimates.

To address these concerns, we use an instrumental variable strategy that focuses on variation in US imports that is a function of supply shocks to US exports. 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

$$IVEmbTechK_{i,t}^h = \prod\nolimits_{l \neq h} \left(1 + \sum\nolimits_{p \in \mathcal{P}^l} (w^l(p) K_{US,t}^p X_{US,i,t}^p)\right)^{\kappa_{US,t}^{l,h}}, \tag{12}$$

and

$$IVEmbTechP_{i,t}^h = \prod\nolimits_{l \neq h} \left(1 + \sum\nolimits_{p \in \mathcal{P}^l} (w^l(p) K_{US,t}^p X_{US,i,t}^p)\right)^{\rho_{US,t}^{l,h}}, \tag{13}$$

where $X_{US,i,t}^p = \sum_{j \notin \mathcal{C}_i} M_{US,j,t}^p$ and $\mathcal{C}_i$ is a cluster of countries with similar characteristics to

[30]For example, demand shocks or policy driven shocks in a sector can change expected future profits in that sector.

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

country i. We construct the cluster $\mathcal{C}_i$ as the countries that fall into both the same quintile of GDP-per-capita (in 1995) and total trade (imports plus exports) to GDP ratio as country i.[32] We include the first variable to capture similarities in technological development across countries and the second variable to capture similarities in trade patterns across countries.

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

Although the instrumental variables correct for demand shocks to imports that vary at the country-sector-year level, we would be remiss not to mention the principal endogeneity concern that may still be present. If there are demand shocks to imports at the sector-year level for sectors within a BEA summary sector that are common across all countries, our instruments may reflect these demand shocks rather than US export supply shocks. Since the instruments is constructed at the sector-year level, we are unable to include sector-year fixed effects, at the same level of disaggregation, and so cannot preclude that these common demand shocks affect our estimates.[33]

6 Results

We discuss the results from estimating the effects of knowledge-weighted and production-weighted embodied technology imports on the innovation and diffusion outcomes. We report the estimates using the baseline specifications described in Equation (11) and use the empirical model to quantify the magnitudes of the results. We close this section by discussing the robustness of the results to alternative specifications.

6.1 Baseline Results

Table 3 reports the results for the innovation and diffusion outcomes. We interpret the innovation results as corresponding to $X^h_{i,t}$ in the context of the conceptual framework. All

[32]We use the entire sample of countries that we have data, not just those included in the final sample.

[33]While this may be the case, because our sample includes many countries at a wide range of stages of development and with different patterns of trade with the US, it is unlikely that there are sector-year demand shocks common to all the countries in our sample.

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 IV estimates are larger than the OLS estimates (reported in Appendix D) for knowledge-weighted embodied technology imports $EmbTechK$ and are similar for production-weighted embodied technology imports $EmbTechP$.

Table 3: Main Results

	(1) *Patents*	(2) *FwdCites*	(3) *FwdRate*	(4) *USBackCites*	(5) *USBackRate*	(6) *USBackShare*
ln $EmbTechK$	0.666***	0.545***	0.052	1.023***	0.076	0.002
	(0.048)	(0.054)	(0.048)	(0.082)	(0.055)	(0.011)
ln $EmbTechP$	0.003**	0.002*	-0.001	0.005*	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)
ln $EmbTechDiag$	0.019	0.001	-0.013	0.023	0.012	0.004
	(0.012)	(0.012)	(0.012)	(0.021)	(0.015)	(0.003)
ln $OwnTech$	0.005**	0.006**	-0.047***	0.019***	-0.051***	-0.005***
	(0.002)	(0.002)	(0.004)	(0.004)	(0.005)	(0.001)
ln $USTech$	-0.761***	-0.596***	-0.086	-1.130***	-0.128	-0.007
	(0.064)	(0.072)	(0.070)	(0.110)	(0.079)	(0.017)
ln $USTechDiag$	0.097***	0.165***	-0.003	0.261***	0.109***	0.009
	(0.015)	(0.016)	(0.022)	(0.026)	(0.026)	(0.006)
Observations	478,880	478,880	368,950	478,880	368,950	364,724
F-Stat *EmbTechK*	412.6	412.6	609.6	412.6	609.6	686.7
F-Stat *EmbTechP*	12806.4	12806.4	17535.9	12806.4	17535.9	18557.1
F-Stat *EmbTechDiag*	322.1	322.1	349.5	322.1	349.5	350.0

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 from 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

The innovation outcomes, reported in columns (1) to (3), point to a positive impact from knowledge-weighted and production-weighted embodied technology imports $EmbTechK$ and $EmbTechP$ on the innovation rate, as measured by $Patents$ and $FwdCites$. The effect appears to be primarily through the extensive margin (i.e., more patenting) rather than the intensive margin (i.e., more citations). The elasticity of innovation outcomes with respect to shocks to $EmbTechP$ is substantially lower than for $EmbTechK$, by multiple orders of magnitude, for both $Patents$ and $FwdCites$, suggesting that technology spillovers are primarily through knowledge linkages. We show later that the quantitative comparison holds after accounting for the relative variation in $EmbTechK$ and $EmbTechP$.

The diffusion outcomes, reported in columns (4) to (6), point to positive, albeit more mixed, impacts of knowledge-weighted and production-weighted technology imports $EmbTechK$ and

$EmbTechP$. The coefficient estimates for $USBackCites$ are positive for both $EmbTechK$ and $EmbTechP$. To some extent, this follows from the previous result since more frequent patenting implies more backward citations to the US, all else equal.

For $USBackRate$ and $USBackShare$ the point estimates are positive, in line with our expectations, but the estimates 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 at alternative specifications (see Appendix D). This is consistent with the point estimate of $USBackCites$ relative to that of $Patents$, with the former being almost twice as large as the latter in Table 3. 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 $FwdCites$. Specifically, our measure includes citations on any previous US patent, many of which may include innovations that are already well-understood in the domestic economy. Additionally, many countries may not require or enforce that domestic patents cite foreign patents making backward citations a noisy measure of diffusion.[34] Another possibility is that innovators learn both about US patents but also other foreign innovations through US patents. For example, an innovator may learn about a French innovation that is built on by a US product and cite that patent instead. In this regard, we think of both $USBackRate$ and $USBackShare$ as being relatively strict measures of diffusion.

The coefficient for $EmbTechDiag$ is near zero reflecting that this variable captures several channels with potentially opposing impacts on innovation, such as, for example, competition from imports and knowledge spillovers. We do not focus on the spillovers through this channel due to the aforementioned difficulties with their interpretation. The coefficients for $OwnTech$ are positive for $Patents$ and $FwdCites$ suggesting that the stock of domestic knowledge relevant to a country-sector contributes positively to its innovation rate.

The coefficient on $USTech$, which measures relevant upstream knowledge produced by the US, is negative in our baseline results. While this measure captures broader technology diffusion from the US to other countries, it is also correlated with our measure of $EmbTechK$. A change in $USTech$ will also result in a change in $EmbTechK$ since both measures have the same underlying IO linkages and US technology levels by construction. This means that

[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.

the negative estimate cannot be interpreted as the effect of $USTech$, holding all else fixed. In order to understand the effect of $USTech$, we first residualize $EmbTechK$ by regressing $USTech$ on it and taking the residual. We then run our baseline regressions by replacing $EmbTechK$ with its residualized version and keeping all other regressors the same. By the Frisch-Waugh-Lowell theorem, the estimates on all regressors except $USTech$ will remain the same. The new estimate on $USTech$ is positive, and can now be interpreted as the effect of $USTech$ that is orthogonal to the trade-weighted US technology measure $EmbTech$. The details of this exercise are presented in Appendix C.1.

6.2 Quantitative Significance

Our conceptual framework implies the estimation Equation (2), where the innovation rate $X^h_{i,t+1}$ depends on embodied technology imports $EmbTech$ through an elasticity η_S and the IO linkages $\gamma^{l,h}_{F,t}$, which equals the knowledge $\kappa^{l,h}_{US,t}$ and production $\rho^{l,h}_{US,t}$ weights for the US scaled by α and $1-\alpha$, respectively. Taking the estimates for $Patents$ and $FwdCites$ implies that $\eta_S\alpha \in [0.54, 0.67]$ and $\eta_S(1-\alpha) = [0.002, 0.003]$.

To further understand the quantitative magnitude of the results, Table 4 compares the relative magnitude of the variation in the outcomes and the implied variation of outcomes attributable to the model variables. We focus on the residualized standard deviation of the explanatory variables, RSDE, which is calculated as the standard deviation of the variable after removing the estimated effect of all other regressors and fixed effects. The residual variation in the outcome variables, RSDO, is the standard deviation of the outcome variable in question after removing the estimated effect of the fixed effects. We do this to remove both variable trends as well as cross-country and cross-summary-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., increases in trade over time). However, these trends are not important for understanding the economic significance of the coefficient estimates.[35] We estimate the effect of the RSDE of the explanatory variables implied by the model and scale them by the RSDO of the outcome variables.

The table shows that $EmbTechK$ explains around 13% of the residualized variation in $Patents$, 6.5% of the variation in $FwdCites$, and around 8.8% for $USBackCites$. Consistent with the earlier summary statistics, the table also shows that there is more residualized variation in $EmbTechP$, which increases its relative quantitative importance, but this gap is

[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 4: Quantitative Significance

		Coefficient Estimate		Relative Implied RSD (%)	
Outcome	RSDO	*EmbTechK*	*EmbTechP*	*EmbTechK* (RSDE = 0.037)	*EmbTechP* (RSDE = 1.197)
Patents	0.185	0.666	0.003	13.2	0.06
FwdCites	0.308	0.545	0.002	6.5	0.03
FwdRate	0.478	0.059	-0.001	0.5	-0.01
USBackCites	0.427	1.023	0.005	8.8	0.04
USBackRate	0.616	0.084	0.000	0.5	0.00
USBackShare	0.139	0.002	0.000	0.0	0.01

Notes: RSDO is calculated as the standard deviation of the outcome variable after controlling for the fixed effects used in the baseline specification, Equation (11). RSDE is calculated as the standard deviation of the explanatory variable after controlling for the other regressors and fixed effects used in that specification. For each of *EmbTechK* and *EmbTechP*, relative implied RSD refers to the product of the coefficient estimate and the ratio of the RSDE to the RSDO. Coefficient estimates are taken from Table 3.

not large enough to offset the differences in coefficient estimates found in Table 3. That is, the overall impact of production-weighted embodied technology imports remains marginal compared with knowledge-weighted embodied technology imports.

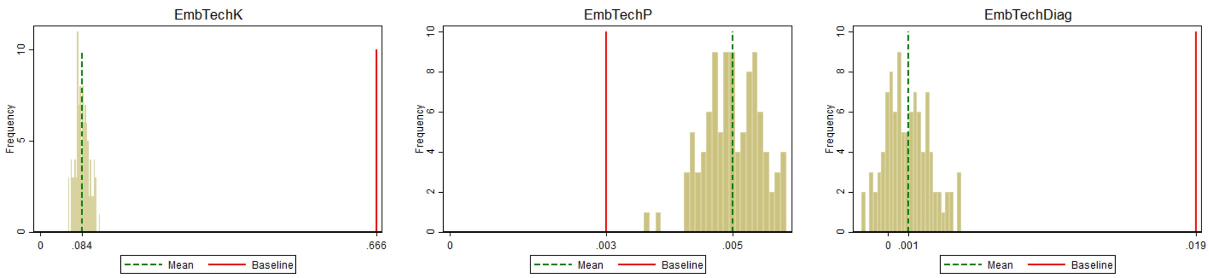
6.3 The Role of Trade Flows in Technology Diffusion

As discussed, we do not expect technology to diffuse across borders only through trade linkages. We now consider a falsification exercise to help understand the role trade plays in this diffusion process. A first step towards identifying the role of trade, *separate* from broader diffusion channels, is by controlling for sector-relevant upstream US technology, $USTech$, as in our baseline specification. Our baseline results suggest that trade plays an important role in the form of diffusing embodied technology even after controlling for the same upstream technology that is not augmented by trade flows.

To further address these concerns, we conduct a falsification test by randomizing the trade flows data used in the construction of our embodied technology variables. Specifically, we take the true imports from the US data for all countries that the data is available for, and randomize their allocation across years, countries, and sub sectors.[36] We construct the ran-

[36]See Appendix C.2 for details of the randomization procedure.

Figure 3: Falsification Test: Randomizing Trade flows



Notes: Figure displays the histograms of the coefficients θ_1^r, θ_2^r, and θ_4^r from the regression specified in Equation 14 for 100 draws of trade flows randomization. The dependent variable in all exercises is *Patents*. The green dashed line represents the mean of the corresponding distributions, while the red solid line represents the respective coefficient of interest from the baseline regression specified in Equation 11.

domized counterparts of the embodied technology variables, which we label as $EmbTechK(r)$, $EmbTeckP(r)$, and $EmbTechDiag(r)$ where r stands for the r^{th} randomization of trade flows. We then re-estimate our baseline specification using the randomized variables for 100 draws of randomized variables.

$$\begin{aligned}\ln(1 + Outcome_{i,t}^h) = \theta_1^r \ln EmbTechK_{i,t-1}^h(r) + \theta_2^r \ln EmbTechP_{i,t-1}^h(r) + \theta_3^r \ln OwnTech_{i,t-1}^h \\ + \theta_4^r \ln EmbTechDiag_{i,t-1}^h(r) + \theta_5^r \ln USTech_{t-1}^h + \theta_6^r \ln USTechDiag_{t-1}^h \\ + V_{i,t-1}^h \beta^r + f_{i,t}^r + f_t^{rn(h)} + f_i^{rh} + \epsilon_{i,t}^h.\end{aligned} \tag{14}$$

Figure 3 shows the distribution of estimates θ_1^r, θ_2^r, and θ_4^r from the regression specification in Equation (14) for *Patents* as the outcome variable. The mean of the randomized model coefficient estimates is plotted as a dashed green line and the baseline estimates are plotted as a solid red line. The falsification test confirms that the trade weights used in the $EmbTechK$ measure are important, with the baseline estimate (0.666) being eight times higher than the mean of the coefficients from the randomized data (0.084). However the baseline estimate on $EmbTechP$ and its counterpart from the randomized data are relatively smaller and close to each other (0.003 and 0.005). This could indicate that diffusion of embodied technology through production linkages might not be a robust feature of the data. Finally, the coefficients on the own-sector embodied technology, $EmbTechDiag$, from the randomized data (mean of 0.001) are also an order of magnitude smaller compared to the larger, but statistically insignificant, baseline estimate (0.019). We conclude that the randomized data is unable to replicate the coefficient estimates for knowledge-weighted embodied technology imports, highlighting the importance of trade.[37]

[37]We present the results of the falsification exercises for other outcome variables in Appendix C.2.

6.4 Robustness Checks and Additional Results

In Appendix D we consider the robustness of the main results to alternative specifications and controls as well as present additional related results. The results are quantitatively similar when alternative lag structures are used instead of the baseline specification. Our baseline results construct $EmbTechK$ and $EmbTechP$ using upstream linkages. We show that the results hold when the corresponding variables constructed with downstream linkages are included. The results become stronger when the country sample is restricted to the top 40 countries based on patenting, consistent with our view that less patenting countries are noisier. The results hold using alternative IV strategies, including the traditional leave-one-out IV. The results hold using alternative transformation of the outcome variables that avoid the log of one plus transformation. Finally, the results hold for alternative outcomes, based on triadic patents, and constructions of knowledge stocks.

7 Conclusion

Innovation activities are highly concentrated in a small number of countries, but new technology eventually diffuses to other countries. A 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, we construct knowledge and production IO tables using data on inter-sectoral patent citations and sales. We combine these measures with data on sector-level trade flows between countries to construct measures of the knowledge-weighted and production-weighted technology 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 better understanding 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 production linkages, diffusion through the trade of goods may not primarily occur between firm relationships that underpin our sector-level import data. Instead, spillovers may primarily be 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 to evaluate the aggregate growth and welfare implications of knowledge diffusion through trade and the 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.

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 descriptive comparisons of the knowledge and production IO tables of the US economy and of the US and non-US knowledge IO tables. We highlight five observations that emerge from these exercises. The first three concern the cross-sectional relationship between the US knowledge and production IO tables. The fourth observation assesses the differences in the time-series characteristics of the US knowledge and production IO tables. The final observation compares the knowledge IO tables of non-US countries with those of the US across different years.

A.1 Cross-Sectional Correlations of US IO Linkages

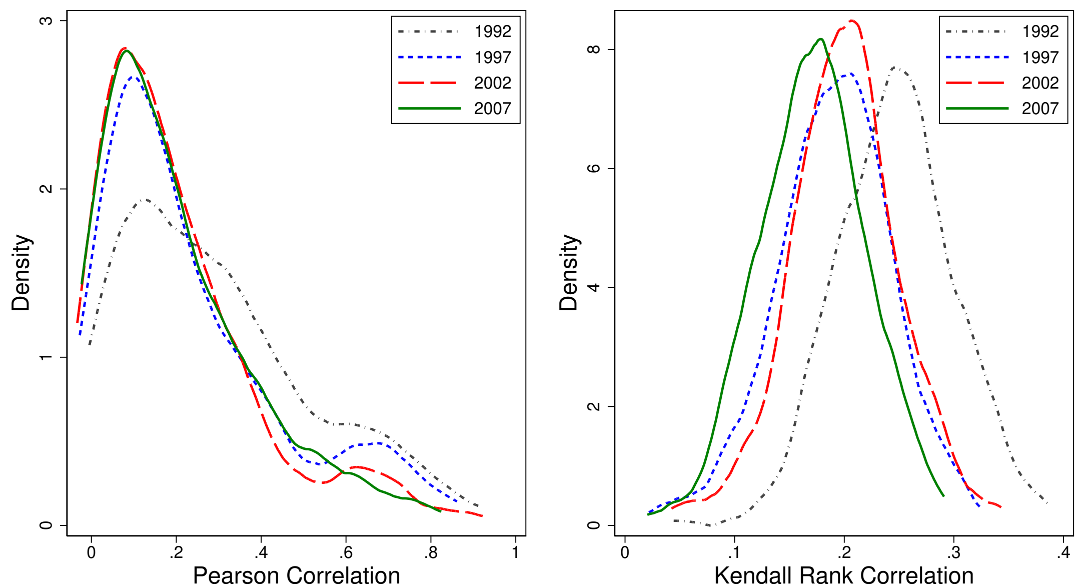
Our empirical analysis, which compares the effects of imports of technology embodied in knowledge and production inputs on patenting outcomes, depends to a large extent on there being distinct variation in the sources of those inputs for the average sector in order to draw the inferences that we do. 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 Pearson correlation of $\kappa^{l,h}_{US,2002}$ and $\rho^{l,h}_{US,2002}$ across all 85,264 sector-pair IO linkages for the 292 sectors depicted in the figure is 0.179, while for the off-diagonal IO linkages it is 0.130.

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 $\kappa^{l,h}_{US,t}$ and $\rho^{l,h}_{US,t}$ across all input sectors l for each output sector h for each of the four years t for which we use the IO tables to construct the explanatory variables in our empirical analysis (i.e., 1992, 1997, 2002, and 2007).[38] For each IO table year, 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 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 in each of the four IO table years.

More formally, we display summary statistics of these distributions in Table A.1. Looking

[38]For the exercises in all sections of this appendix except Appendix A.3, we drop industries that do not make use of inputs from any of the 292 input sectors used in our baseline analysis in at least one of the US knowledge or production IO tables, which leaves 287 sectors. We then renormalize the IO linkages by output sectors' total input usage of the remaining input sectors in each year.

Figure A.1: Distributions of Cross-Sectional Correlation Coefficients of US IO Linkages



Notes: Figure plots the distributions of correlation coefficients of US IO Linkages for different IO table years. Coefficients are computed as the correlation of US knowledge and production IO linkages across all input sectors for each output sector in each year. The left panel displays the distributions of the Pearson correlation coefficients while the right panel displays the distributions of the Kendall rank correlation coefficients (adjusted for ties).

across IO table years, there appears to be a decreasing trend in the average correlation of US knowledge and production IO linkages across our sample period. We also include statistics for the distributions of correlation coefficients computed using only off-diagonal IO linkages to show that differences in the intensity of use of own-sector knowledge and production inputs are not driving the low average correlations. We now state our first observation regarding the comparison of the knowledge and production IO tables.

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

A.2 Concentration and Sparsity of IO Linkages

Next, we investigate another major cross-sectional difference between the US 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.

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 from each of

Table A.1: Summary Statistics of Cross-Sectional US IO Linkage Correlation Coefficients

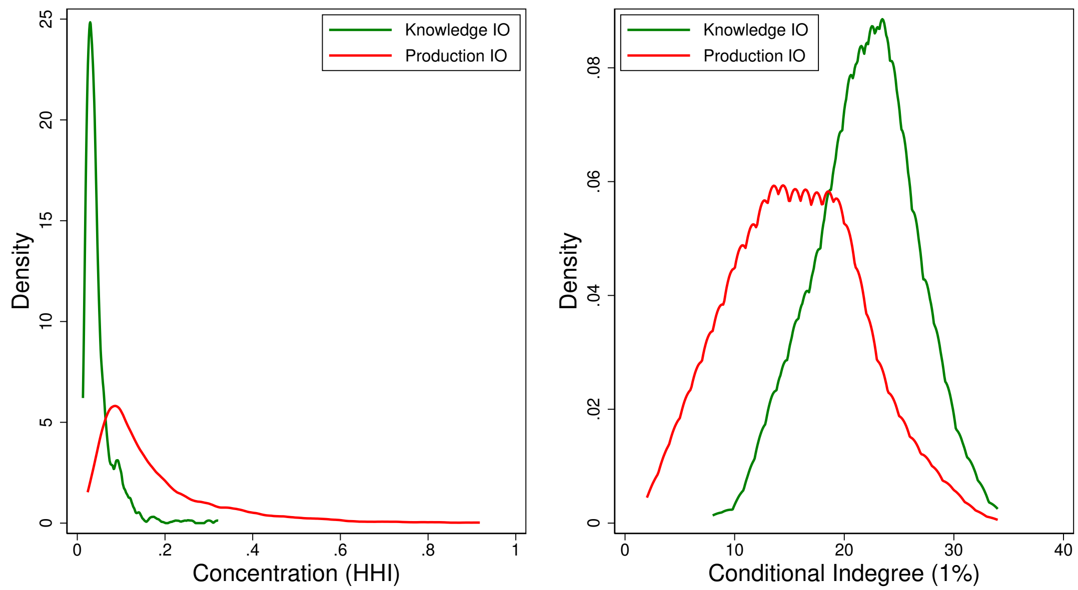
	N	Min	Max	Median	Mean	Std. Dev.
Pearson Correlation Across All Inputs						
1992	287	-0.004	0.921	0.249	0.295	0.224
1997	287	-0.027	0.861	0.170	0.237	0.213
2002	287	-0.033	0.922	0.156	0.204	0.191
2007	287	-0.023	0.825	0.155	0.202	0.181
All Years	1,148	-0.033	0.922	0.174	0.235	0.206
Kendall Rank Correlation Across All Inputs						
1992	287	0.044	0.386	0.247	0.247	0.053
1997	287	0.022	0.326	0.192	0.190	0.053
2002	287	0.042	0.343	0.201	0.200	0.051
2007	287	0.020	0.291	0.173	0.171	0.049
All Years	1,148	0.020	0.386	0.202	0.202	0.058
Pearson Correlation Across Off-Diagonal Inputs						
1992	287	-0.034	0.915	0.205	0.248	0.206
1997	287	-0.026	0.862	0.132	0.188	0.196
2002	287	-0.037	0.923	0.100	0.150	0.167
2007	287	-0.037	0.825	0.110	0.155	0.161
All Years	1,148	-0.037	0.923	0.132	0.185	0.187
Kendall Rank Correlation Across Off-Diagonal Inputs						
1992	287	0.032	0.381	0.243	0.242	0.053
1997	287	0.012	0.320	0.186	0.184	0.054
2002	287	0.044	0.336	0.194	0.193	0.051
2007	287	0.022	0.284	0.166	0.164	0.049
All Years	1,148	0.012	0.381	0.194	0.196	0.059

Notes: Table reports summary statistics of the distributions of cross-sectional correlation coefficients of US 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.

the US IO tables. 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 HHI-$\mathrm{K}^h_{US,t} = \sum_{l\in\mathcal{H}}(\kappa^{l,h}_{US,t})^2$ for knowledge IO linkages and HHI-$\mathrm{P}^h_{US,t} = \sum_{l\in\mathcal{H}}(\rho^{l,h}_{US,t})^2$ for production IO linkages. Second, we construct conditional indegrees (CID) for each IO table that measure the number of input sectors that have an IO linkage with an output sector that is larger than some threshold level c.[39] For output sector h, the conditional indegree for knowledge IO linkages is CID-$\mathrm{K}^h_{US,t}(c) = \sum_{i\in\mathcal{H}} \mathbb{1}(\kappa^{l,h}_{US,t} \geq c)$ and for production IO linkages is CID-$\mathrm{P}^h_{US,t}(c) = \sum_{i\in\mathcal{H}} \mathbb{1}(\rho^{l,h}_{US,t} \geq c)$, where $\mathbb{1}(\cdot)$ is the indicator function.

[39] 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).

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



Notes: Figure plots the distributions of the concentration and conditional indegree measures of US knowledge and production IO linkages across output sectors and the four IO table years. 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.

In Figure A.2, we depict the distributions of the HHI and CID measures for both knowledge and production IO linkages.[40] 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 the concentration of production IO linkages while the reverse is true for the distributions of conditional indegree measures.

Table A.2 lists summary statistics of these distributions as well as the distributions of the HHI and CID measures computed using only off-diagonal input sectors. For this latter group of distributions, we modify the definitions of the knowledge and production IO linkages such that, for output sector h, the denominators of Equation (3) and Equation (4) only sum over input sectors $l \neq h$.[41] Knowledge IO linkages are less concentrated than production IO linkages, in part because for the average output sector there are fewer significant knowledge input sectors than production input sectors (where significant means larger than 1% here).

[40]Since the distributions are quite similar across IO table years for both knowledge and production IO linkages, we group the concentration and sparsity measures across all four years into a single distribution for each IO table for visual clarity. The distributions for each IO table year are displayed in Figure D.2 and their summary statistics are shown in Table D.12.

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

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

	Min	Max	Median	Mean	Std. Dev.
All Inputs					
HHI-K$^h_{US,t}$	0.013	0.321	0.037	0.047	0.035
HHI-P$^h_{US,t}$	0.024	0.918	0.128	0.173	0.138
CID-K$^h_{USt}$(1%)	8	34	22	21.866	4.624
CID-P$^h_{US,t}$(1%)	2	34	15	15.582	6.062
Off-Diagonal Inputs					
HHI-K$^h_{US,t}$	0.013	0.319	0.033	0.041	0.028
HHI-P$^h_{US,t}$	0.024	0.970	0.123	0.180	0.159
CID-K$^h_{US,t}$(1%)	8	34	23	22.820	4.664
CID-P$^h_{US,t}$(1%)	1	34	16	15.824	6.158

Notes: Table reports summary statistics of the cross-sectional distributions of the Herfindahl-Hirschman Index (HHI) and conditional indegree (CID) measures of US knowledge and production IO linkages across the 1992, 1997, 2002, and 2007 IO tables. For measures computed using off-diagonal sectors, own-sector IO linkages are omitted from the calculation of the IO linkages. 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.

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 in the US.*

A.3 Key Input Sectors

The last major cross-sectional 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 the US 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 US knowledge and production IO tables.[42]

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 COD-K$^l_{US,t}(c) = \sum_{h \in \mathcal{H}} \mathbb{1}(\kappa^{l,h}_{US,t} \geq c)$ for knowledge IO linkages and

[42]We focus here on the 2002 US IO tables to illustrate our findings and include all 292 sectors as in Figure 1.

COD-P$^l_{US,t}(c) = \sum_{h \in \mathcal{H}} \mathbb{1}(\rho^{l,h}_{US,t} \geq c)$ for production IO linkages. Second, we use the (unconditional) weighted outdegree (WOD) of input sectors with WOD-K$^l_{US,t} = \sum_{h \in \mathcal{H}} \kappa^{l,h}_{US,t}$ for knowledge IO linkages and WOD-P$^l_{US,t} = \sum_{h \in \mathcal{H}} \rho^{l,h}_{US,t}$ 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.[43] In our context, these measures are defined by

$$\text{AWC-K}^l_{US,t} = \lambda_K \sum_{h \in \mathcal{H}} \kappa^{l,h}_{US,t} \text{HWC-K}^h_{US,t}$$
$$\text{HWC-K}^l_{US,t} = \mu_K \sum_{h \in \mathcal{H}} \kappa^{h,l}_{US,t} \text{AWC-K}^h_{US,t}$$

for knowledge IO linkages and

$$\text{AWC-P}^l_{US,t} = \lambda_P \sum_{h \in \mathcal{H}} \rho^{l,h}_{US,t} \text{HWC-P}^h_{US,t}$$
$$\text{HWC-P}^l_{US,t} = \mu_P \sum_{h \in \mathcal{H}} \rho^{h,l}_{US,t} \text{AWC-P}^h_{US,t}$$

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

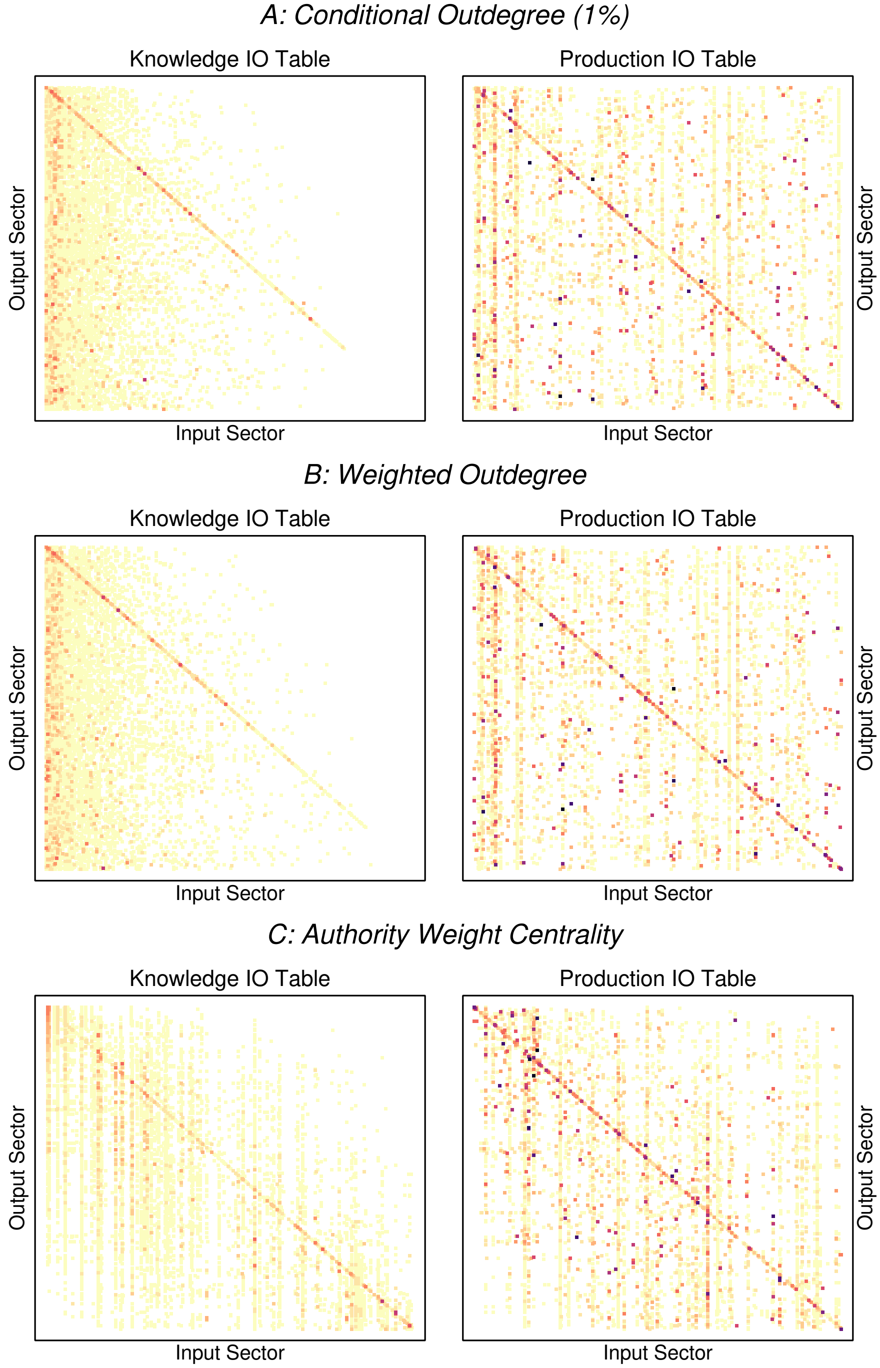
To illustrate that the key input sectors are different across the IO tables, we reproduce versions of Figure 1 in which we reorder sectors according to the ranking of sectors by these three centrality measures in the 2002 US IO tables. In Figure A.3, we order sectors in each panel by the rank of sectors 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 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 of inputs in the production IO table.[44] We close this section by stating our third observation from comparing the US knowledge and production IO tables.

[43] 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.

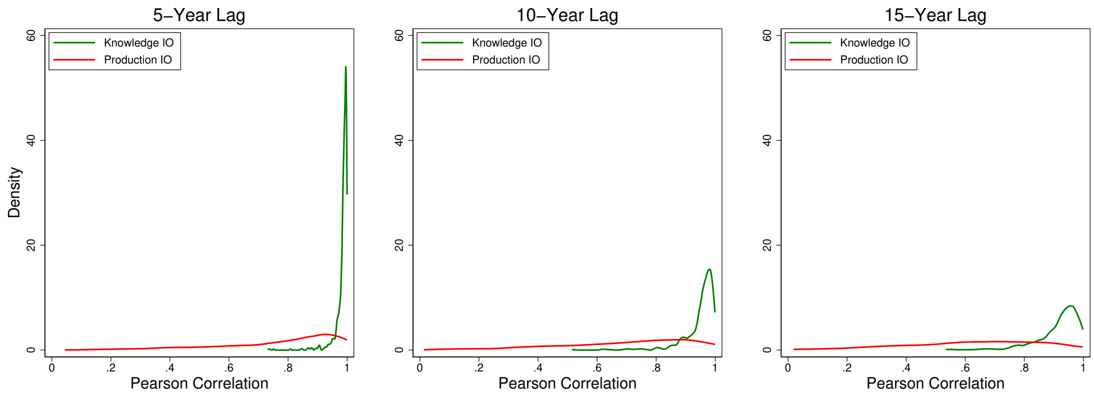
[44] When sectors are instead ordered by the rankings of the centrality measures constructed using production IO linkages, the reverse implication is visually apparent (see Figure 1).

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



Notes: Figure plots the US knowledge and production IO tables in 2002 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 defined in text. IO linkages are defined in Section 4.1. All plots only display IO linkages that account for at least 1% of the inputs used by an output sector.

Figure A.4: Distributions of Autocorrelations of US IO Linkages



Notes: Figure plots the distributions of autocorrelation coefficients of US IO Linkages for different lags using the 1992, 1997, 2002, and 2007 IO tables. Coefficients are computed as the Pearson autocorrelation of US knowledge and production IO linkages across all input sectors for each output sector. IO linkages are defined in Section 4.1.

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

A.4 Persistence of US IO Linkages

The measures of technology embodied in knowledge and production inputs imported from the US are constructed using IO linkages that vary across years. We allow for this temporal variation because the relevance of different sources of knowledge and production inputs for generating new innovations need not be fixed over time. We now describe the extent to which US knowledge and production IO linkages vary across time at different time horizons and compare the persistence of input sources on average across output sectors in the knowledge and production IO tables.

To do so, we make use of the IO tables for the four years used to construct the explanatory variables in our empirical analysis (i.e., 1992, 1997, 2002, and 2007). For both the knowledge and production IO tables, we compute the autocorrelations of the IO linkages of each output sector across all input sectors between pairs of IO tables for different years and group these autocorrelations by the number of years between the tables. This provides us with distributions of autocorrelations (measured using the Pearson correlation coefficient) for both knowledge and production IO linkages at lag lengths of five, ten, and fifteen years. We plot these distributions in Figure A.4, separating out the distributions for different lag lengths into their own charts for visual clarity.

Unsurprisingly, the distribution of autocorrelations shifts to the left as the lag length

Table A.3: Summary Statistics of US IO Linkage Autocorrelations

	N	Min	Max	Median	Mean	Std. Dev.
5-Year Lag						
Knowledge IO	861	0.730	1.000	0.990	0.984	0.025
Production IO	861	0.044	1.000	0.839	0.777	0.205
10-Year Lag						
Knowledge IO	574	0.514	0.999	0.967	0.950	0.056
Production IO	574	0.014	1.000	0.751	0.699	0.227
15-Year Lag						
Knowledge IO	287	0.534	0.999	0.938	0.918	0.074
Production IO	287	0.019	0.998	0.650	0.628	0.230

Notes: Table reports summary statistics of the distributions of Pearson autocorrelation coefficients of US knowledge and production IO linkages for the 287 output sectors that have non-zero input usage for at least one input sector in each of the 1992, 1997, 2002, and 2007 US knowledge and production IO tables. Std. Dev. is the standard deviation. IO linkages are defined in Section 4.1.

between the IO tables increases; the sources of inputs for the average output sector in each year are more highly correlated with the sources used in years that are closer in time to that year than with those used in more distant years. More interestingly, the charts make clear that at all three time horizons, the average autocorrelation of knowledge IO linkages is substantially higher than it is for production IO linkages. These visual insights are confirmed by the summary statistics of the distributions of autocorrelations presented in Table A.3.[45] This evidence supports our fourth observation on the differences between sources of knowledge and production inputs.

Observation 4: *For the average sector in the US, the sources of knowledge inputs are more highly correlated across time than the sources of production inputs.*

We also examine the inter-temporal cross-table correlation of US knowledge and Production IO linkages. For each sector, we compute the Pearson correlations of knowledge IO linkages and production IO linkages across all input sectors for each pair of IO table years. We report the means of the distribution of these cross-table correlations for each pair of IO table years in Table A.4. Across all pairs of IO table years, the average correlation is similar in the cross-section and across time. Despite these similarities, there average cross-table correlations tend to be slightly larger for earlier years than for later years for both the knowledge and production IO linkages.

[45]These patterns remain the same when using the Kendall rank correlation coefficient rather than the Pearson correlation coefficient to compute the autocorrelations and when using only off-diagonal IO linkages to compute the correlations. These additional results are available on request.

Table A.4: Cross-Table Correlations of US IO Linkages Across Time

		Production IO Linkages			
		1992	1997	2002	2007
Knowledge IO Linkages	1992	0.295	0.249	0.236	0.250
	1997	0.282	0.237	0.224	0.237
	2002	0.259	0.218	0.204	0.216
	2007	0.245	0.205	0.191	0.202

Notes: Table reports means of the distributions of cross-table correlation coefficients of US Knowledge and Production IO linkages across different pairs of IO table years. Coefficients are computed as the Pearson correlation of knowledge and production IO linkages across all input sectors for each output sector in each pair of years.

A.5 Evolution of Knowledge IO Linkages Across Countries

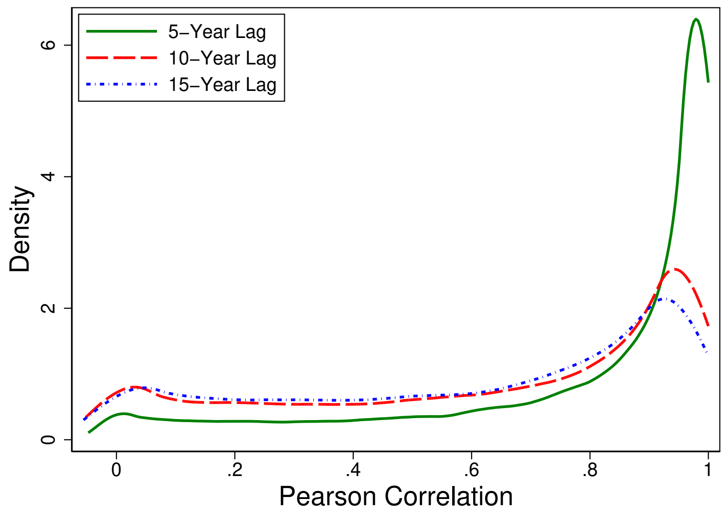
To control for the effects of past domestic innovation activity on new patenting and citation activity in sectors in importing countries, we weight upstream domestic knowledge stocks by the relevance of input sectors for generating new innovations in the importing country. As with the IO linkages used to weight US technology embodied in imported inputs, we allow these importer-sector-specific weights to vary over time. We now document the degree to which knowledge IO linkages in sectors in non-US importing countries change over time and the relationship between those IO linkages and the US knowledge IO linkages across the IO tables from the four years used to construct our explanatory variables.[46]

We begin by computing the autocorrelations of IO linkages in each country-sector at different time horizons using the same approach as in Appendix A.4. The distributions of these autocorrelations are plotted in Figure A.5, and the summary statistics of these measures are reported in the panel A of Table A.5. The knowledge IO linkages in sectors in importing countries are on average substantially less persistent than those of US sectors at each of the three time horizons we consider. In other words, the knowledge IO linkages in many country-sectors are not stable across time.

Given our focus on imports of embodied technology from the US, we examine the relationship between knowledge IO linkages in importing countries and the US. We compute the cross-sectional correlations of each output sector's IO linkages between each importer and the US for each of the four IO table years. We plot the distributions of these correlations in the left panel of Figure A.6 and display their summary statistics in panel B of Table A.5. Although the knowledge IO linkages in many country-sectors are highly correlated with the

[46]We use the same 287 sectors for the 1992, 1997, 2002, and 2007 IO tables as in Appendix A.4. The IO linkages are renormalized within each year such that they sum to one across input sectors for each output sector in each country in each IO table.

Figure A.5: Distributions of Autocorrelations of non-US Knowledge IO Linkages



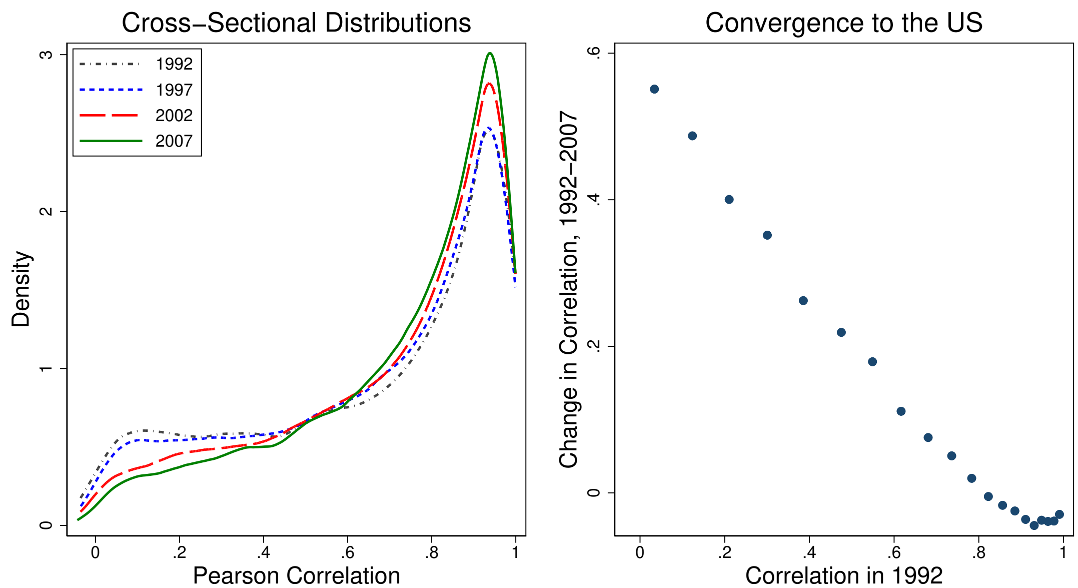
Notes: Figure plots the distributions of autocorrelation coefficients of non-US IO Linkages for different lags using the 1992, 1997, 2002, and 2007 IO tables. Coefficients are computed as the Pearson autocorrelation of knowledge IO linkages across all input sectors for each output sector in each country. IO linkages are defined in Section 4.1.

Table A.5: Summary Statistics of non-US knowledge IO Linkage Correlations

	N	Min	Max	Median	Mean	Std. Dev.
A: Autocorrelations of non-US Knowledge IO Linkages						
5-Year Lag	55,528	-0.048	1.000	0.926	0.788	0.281
10-Year Lag	34,196	-0.052	1.000	0.737	0.628	0.329
15-Year Lag	15,946	-0.056	0.999	0.698	0.603	0.324
B: Correlations with US Knowledge IO Linkages						
1992	16,444	-0.035	0.999	0.754	0.653	0.302
1997	19,419	-0.035	0.999	0.755	0.662	0.291
2002	21,223	-0.036	0.999	0.789	0.694	0.272
2007	23,414	-0.043	0.999	0.802	0.714	0.257

Notes: Panel A reports summary statistics of the distributions of Pearson autocorrelation coefficients of non-US knowledge IO linkages for the 287 output sectors that have non-zero input usage for at least one input sector in each of the 1992, 1997, 2002, and 2007 US knowledge and production IO tables. Panel B reports summary statistics of the distributions of Pearson correlation coefficients between US and non-US knowledge IO linkages across country-sectors for each IO table. Std. Dev. is the standard deviation. IO linkages are defined in Section 4.1.

Figure A.6: Correlations of Importer and US Knowledge IO Linkages



Notes: The left panel plots the distributions of cross-sectional correlations of knowledge IO linkages between the US and each other country for each of the 1992, 1997, 2002, and 2007 IO table years. Coefficients are computed as the Pearson correlation of knowledge IO linkages across all input sectors for each output sector in each country. IO linkages are defined in Section 4.1. The right panel is a binscatter plot of the difference between the correlations in 2007 and 1992 against the level of the correlation in 1992.

linkages in the same sector in the US, there are also many that have only low or moderate correlations.

Figure A.6 and Table A.5 also imply that across our sample period, the sources of knowledge inputs in the average sector in non-US countries have become more similar to the sources used in that sector in the US. However, there is heterogeneity in the extent to which knowledge IO linkages converge to those used in the US. In the right panel of Figure A.6, we display a binscatter plot of the change in correlation between non-US and US knowledge IO linkages between 1992 and 2007 (measured using the difference between the coefficients) against the level of the correlation in 1992. The plot displays a clear pattern: country-sectors that had knowledge IO linkages that were initially relatively less correlated with those in the same US sector underwent a relatively large increase in the correlation on average across the 15 years considered. We summarize our findings from this section in the following observation.

Observation 5: *The sources of knowledge inputs are less persistent on average in non-US country-sectors than in US sectors. In part, this reflects a convergence on average between each non-US country-sector and the same sector in the US.*

Appendix B Data Appendix

Google Patents Data. Our knowledge IO linkages, stocks of knowledge, 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.[47] 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.[48]

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

Out of a total of 74.8 million patent families in the Google Patents database, 67.9 million of them have at least one 4-character IPC code, which is a minimum requirement in order for them to be included in the data underlying the knowledge IO tables we construct. Meanwhile, 71.8 million patent families have filing dates, while only 20.9 million have inventor country

[47]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.

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

[49]92% of patent publications are primary series documents, and 98% of patent families have at least one primary series document filed.

information.[50] 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.[51] 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.[52] 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.[53] In total, 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.[54]

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

[50] 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.

[51] 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.

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

[53] 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.

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

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.[55] 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.

We first crosswalk the Google Patents data on knowledge stocks, defined in Section 5.2, 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.[56] This second concordance uses equal weights for each

[55]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).

[56]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/`

1992 6-digit HS code into which a given 2002 6-digit HS code maps.[57]

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.[58] The second concordance converts 1987 4-digit SIC codes into 2002 6-digit NAICS codes, again using an unweighted mapping between the classifications.[59] 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.[60] The third concordance applies the mapping of 2002 6-digit NAICS codes into the endpoint 2002 BEA sector codes.[61] 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 subsectors into sectors referred to in Section 5.2.

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.

PatentCrosswalk.

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

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

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

[60]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.

[61]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.

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.[62] 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.[63] 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.[64] The second and third concordances map 2012 NAICS sectors into 2007 NAICS sectors and 2007 NAICS sectors into 2002 NAICS sectors, respectively.[65] 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.

Sample Selection We limit the set of countries in our final sample based on the following criteria. We drop countries with no triadic patents in any sector or year of the panel and

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

[63] 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`.

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

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

countries with a population of less than one million in 1995 to avoid inclusion of countries where patenting outcomes may be too noisy. We also drop those countries that have exports to GDP or imports to GDP ratios in 2015 above the 98th percentile or below the 2nd percentile of those statistics among the remaining set of countries or countries that have imports to GDP or exports to GDP ratios in 2015 that are larger than one. This restricts countries that trade for reasons unrelated to production or consumption, such as countries that primarily act as intermediaries. Finally, we keep the top 25% of the remaining countries based on total triadic patents over the sample. This restriction excludes countries where innovations are either infrequent or 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. 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.

Our final sample contains the following 83 countries: Albania, United Arab Emirates, Argentina, Armenia, Australia, Austria, Belgium, Bangladesh, Bulgaria, Belarus, Bolivia, Brazil, Canada, Switzerland, Chile, China, Colombia, Costa Rica, Czech Republic, Germany, Denmark, Dominican Republic, Algeria, Ecuador, Egypt, Spain, Estonia, Finland, France, United Kingdom, Georgia, Ghana, Greece, Guatemala, Croatia, Hungary, Indonesia, India, Ireland, Iran, Israel, Italy, Jordan, Japan, Kazakhstan, Kenya, South Korea, Kuwait, Lebanon, Sri Lanka, Lithuania, Latvia, Morocco, Moldova, Mexico, Mauritius, Malaysia, Netherlands, Norway, New Zealand, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Paraguay, Russian Federation, Saudi Arabia, Singapore, Slovakia, Slovenia, Sweden, Thailand, Trinidad and Tobago, Tunisia, Turkey, Ukraine, Uruguay, Uzbekistan, Viet Nam, South Africa, and Zimbabwe.

Appendix C Trade Flows and Technology Diffusion

In this appendix, we show additional results to supplement the discussion in the main text on how to separately identify the effect of $USTech$ from that of $USEmbTech$, as well as additional results on the falsification exercises described in Section 6.3.

C.1 Effect of $USTech$ on Innovation Outcomes

As described in the main text, in our baseline results the estimate on $USTech$ is negative and significant. However, we cannot interpret this estimate as the only effect of $USTech$ since our measures of $USTech$ and $EmbTechK$ are correlated by construction. This means that a thought exercise in which only $USTech$ changes, keeping other regressors in the model fixed, is not possible as the underlying elements that make up $USTech$ are also the underlying elements that make up $EmbTechK$. In order to understand the effect of $USTech$ alone that is orthogonal to $EmbTechK$, we first residualize $EmbTechK$ with respect to $USTech$ as follows:

$$\ln EmbTechK_{i,t}^{h} = \gamma \ln USTech_{t}^{h} + \eta_{i,t} + \eta_{t}^{n(h)} + \eta_{i}^{h} + \nu_{i,t}^{h}$$
$$\ln Resid\ EmbTechK_{i,t}^{h} = \ln EmbTechK_{i,t}^{h} - \hat{\gamma} \ln USTech_{t}^{h}.$$

By the Frisch-Waugh-Lowell theroem, replacing $\ln EmbTechK_{i,t}^{h}$ with its orthogonalized form $\ln Resid\ EmbTechK_{i,t}^{h}$ will not change its estimate. However, a benefit of this orthogonalization process is that the variable $USTech$ will no longer move hand in hand with $EmbTech$. Table C.1 shows the results from the original baseline regression as well as the regression with the residualized $EmbTechK$ with Patents as the outcome variable. The estimate on $USTech$ in the residualized regression is 0.111, in contrast to the -0.761 in the baseline. We can hence say that the broader effect of US technology on patenting activity that is orthogonal to embodied technology is 0.111. We also run a OLS regression without any of the endogenous embodied technology variables considered in our main specification as a sanity check. We find that the effect of $USTech$ in a model without the correlated $EmbTech$ varibales is very close to the estimate in the residualized regression. We find similar results for the other dependent variables considered in our baseline.[66]

[66]We skip presenting results for all variables for brevity. Results are available upon request.

Table C.1: Effect of $USTech$ on Innovation Outcomes

	(1) Baseline: IV *Patents*	(2) Residualized: IV *Patents*	(3) OLS *Patents*
$\ln EmbTechK$	0.666***		
	(0.048)		
$\ln Resid\ EmbTechK$		0.666***	
		(0.048)	
$\ln EmbTechP$	0.003**	0.003**	
	(0.001)	(0.001)	
$\ln EmbTechDiag$	0.019	0.019	
	(0.012)	(0.012)	
$\ln OwnTech$	0.005**	0.005**	0.001
	(0.002)	(0.002)	(0.002)
$\ln USTech$	-0.761***	0.111***	0.113***
	(0.064)	(0.014)	(0.011)
$\ln USTechDiag$	0.097***	0.097***	0.106***
	(0.015)	(0.015)	(0.010)
Observations	478,880	478,880	478,880

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 from 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

C.2 Role of Trade Flows on Technology Diffusion

Randomization Procedure We randomize the allocation of all the available flows of US imports data across years, countries, and sub sectors. The randomization is done in three layers. First, for each year we randomly draw, without replacement, another year's trade data. Second, within each year, for each country we randomly draw, without replacement, another country's trade data. Finally within each year and for all countries within that year, each sub sector's trade data is drawn, without replacement, from the set of trade flows within that year and country. The final sub sector allocation is kept consistent across all countries to ensure that the instruments used in the first stage are still valid. The following is an example of one such random allocation of trade flows: trade data for 1998 is drawn from 2011 trade data. Data for 1998 China is drawn from 2011 France data, while data for 1998 Australia is drawn from 2011 Canada data. Finally, Chinese imports of shoes from the US in 1998 is given by French imports of chairs from the US in 2011, and Australian imports of shoes in 1998 is given by Canadian imports of chairs in 2011.

Additional Results Table C.2 reports the results from the randomization exercise outlined in Section 6.3 for three outcome variables - $Patents$, $FwdCites$, and $USBackCites$. For each outcome variable we report the baseline estimates as well as the mean and standard deviations of the coefficients from 100 draws of the randomization procedure described above.

Table C.2: Falsification Exercise

	(1)	(2)	(3)	(4)	(5)	(6)
	$Patents$		$FwdCites$		$USBackCites$	
	BD	RD	BD	RD	BD	RD
$\ln EmbTechK$	0.666	0.084	0.545	0.091	1.023	0.141
	(0.048)	(0.014)	(0.054)	(0.017)	(0.082)	(0.026
$\ln EmbTechP$	0.003	0.005	0.002	0.004	0.005	0.008
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
$\ln EmbTechDiag$	0.019	0.001	0.001	0.001	0.023	0.002
	(0.012)	(0.001)	(0.001)	(0.002)	(0.021)	(0.003)

Notes: Table compares estimates from the baseline regression described in Equation (11) to the estimates from the regression described in Equation (14) that uses variables constructed with randomized trade data. Columns (1), (3), (5), titled "BD", uses baseline data and reports the appropriate coefficient and clustered standard errors in parenthesis. Columns (2), (4), (6), titled "RD", uses randomized data and reports the mean and standard deviation (in parenthesis) of the coefficients from 100 randomizations of the underlying trade trade. 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 $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from world. The following fixed effects are included in each column: Country*Sector, Country*Year, and Summary-Sector*Year. Baseline standard errors are clustered twoways: Country*Sector and Sector*Year

Appendix D Additional Figures and Tables

OLS results. Table D.1 reports the OLS results for the baseline specification.

Table D.1: Main Results OLS

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.010***	0.003	-0.003	0.004	0.000	-0.004**
	(0.002)	(0.003)	(0.006)	(0.004)	(0.008)	(0.002)
$\ln EmbTechP$	0.003***	0.003***	-0.001	0.005***	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)
Observations	478,880	478,880	368,950	478,880	368,950	364,724
R-squared	0.983	0.974	0.594	0.962	0.615	0.501

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 $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

Lagged effects of knowledge diffusion. We estimate Equation (11) with the outcome variable measured at period t (rather than a three-year window) and the regessors taken at one- to five-year lags. The results show similar patterns as the baseline results and are reported in Figure D.1. In all regressions the set of controls are the same as in the baseline regressions. The results at different lags are similar in magnitude to the baseline results. In most cases, the coefficient estimates on longer lags are smaller for $EmbTechK$, albeit quantitatively similar. This is consistent with a theory in which diffusion is strongest early on, when the technology is most likely to be embodied in traded goods, and gradually becomes weaker as the technology frontier progresses beyond the technology. On the other hand, the coefficients on $EmbTechP$ get bigger over time, but remain relative very small compared to the coefficients on $EmbTechK$. For $USBackRate$, the coefficient estimate becomes statistically significant at a three-year lag for $EmbTechK$ but remains statistically insignificant at other lags.

Table D.2: Five-Year Forward Average Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.609***	0.414***	0.039	0.880***	0.068	-0.006
	(0.044)	(0.049)	(0.052)	(0.075)	(0.063)	(0.013)
$\ln EmbTechP$	0.002*	0.001	-0.000	0.003	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)
Observations	478,880	478,880	391,742	478,880	391,742	388,237
F-Stat $EmbTechK$	412.6	412.6	558.1	412.6	558.1	627.0
F-Stat $EmbTechP$	12806.4	12806.4	16632.3	12806.4	16632.3	17451.9
F-Stat $EmbTechDiag$	322.1	322.1	340.7	322.1	340.7	343.5

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 $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

Downstream embodied technology. In our baseline results, we use upstream linkages from the IO tables—i.e., how important a sector is as a supplier of knowledge and products—to construct our measure of embodied technology imports. A concern with this implementation is that the upstream linkages are not important. Instead, the empirical results are capturing technology spillovers to closely related sectors. To test this possibility, we construct alternative variables of knowledge-weighted and production-weighted embodied technology imports using downstream linkages (the importance of sectors as users of knowledge and products). The alternative variables capture important economic relationships but are based on the reverse flow of knowledge and products. We find quantitatively similar and statistically significant coefficient estimates on the upstream variables ($EmbTechK$ and $EmbTechP$) when controlling for the downstream variables.

Country sample. We consider an alternative specification where we restrict the sample of countries to the top 40 countries based on the total number of patents (Table D.4). 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

Table D.3: Downstream Linkages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IV	IV	IV	IV	IV	IV	IV
	Patents	*Patents*	*FwdCites*	*FwdRate*	*USBackCites*	*USBackRate*	*USBackShare*
ln *EmbTechK*		0.562***	0.436***	0.018	0.828***	0.009	-0.015
		(0.042)	(0.048)	(0.048)	(0.072)	(0.055)	(0.011)
ln *EmbTechP*		0.003**	0.003**	-0.001	0.007***	0.001	0.000
		(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)
ln *EmbTechDownK*	0.190***	0.084***	0.093***	0.023	0.173***	0.061**	0.015***
	(0.018)	(0.019)	(0.023)	(0.020)	(0.035)	(0.025)	(0.005)
ln *EmbTechDownP*	-0.001*	-0.001*	-0.002***	0.001	-0.005***	-0.002**	-0.000**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Observations	478,880	478,880	478,880	368,950	478,880	368,950	364,724
F-Stat *EmbTechK*		328.1	328.1	377.8	328.1	377.8	379.0
F-Stat *EmbTechP*		4279.8	4279.8	13738.3	4279.8	13738.3	15599.5
F-Stat *EmbTechDiag*		33208.5	33208.5	21813.0	33208.5	21813.0	22159.3
F-Stat *EmbTechDownK*	1768.7	647.7	647.7	1039.8	647.7	1039.8	1172.9
F-Stat *EmbTechDownP*	30561.9	16029.3	16029.3	21181.8	16029.3	21181.8	22241.7

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*, ln *USTech*, ln *USTechDiag*, log total exports to world and log total imports from 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

many of these countries, patenting as an institution may be prohibitively expensive, provide insufficient protection, or simply be underdeveloped. We also find that the coefficient for *EmbTechK* for *USBackRate* becomes statistically significant, consistent with our view that this variable may be noisier in some countries.

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 construct alternative instruments using both the traditional leave-one-out instrument, which can be viewed as a cluster with a single country (Table D.5), and an instrument using all other countries within the country's cluster (Table D.6). The coefficient estimates are a similar magnitude for the traditional leave-one-out instrument.

Table D.4: Top 40 Countries by Total Patents

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	Patents	*FwdCites*	*FwdRate*	*USBackCites*	*USBackRate*	*USBackShare*
ln *EmbTechK*	0.793***	0.631***	0.106***	0.985***	0.139***	0.014
	(0.052)	(0.065)	(0.041)	(0.082)	(0.046)	(0.009)
ln *EmbTechP*	0.002	0.001	-0.003**	0.003	-0.002*	-0.000
	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)	(0.000)
Observations	233,600	233,600	225,109	233,600	225,109	224,308
F-Stat *EmbTechK*	900.4	900.4	1000.4	900.4	1000.4	1133.2
F-Stat *EmbTechP*	18115.4	18115.4	22968.8	18115.4	22968.8	23722.8
F-Stat *EmbTechDiag*	274.4	274.4	299.6	274.4	299.6	301.7

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*, ln *USTech*, ln *USTechDiag*, log total exports to world and log total imports from 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 the within-cluster instrument, the coefficient estimates for *Patents* and *USBackCites* become smaller, but remain statistically significant. This is consistent with our view that excluding countries within the country cluster, as with our baseline instrument, should remove correlated demand shocks.

Alternative transformations of outcomes. In our baseline specification, we take the log of one plus outcome variable to avoid excluding zero-valued observations. We find similar results using the *asinh* transformation (Table D.7) and when we remove zero observations (Table D.8).[67]

Other results. We find similar results when we focus on alternative measures of the outcomes. Table D.9 reports the results using triadic patents—patents that are applied for in

[67] We find different coefficient estimates using the log(Y) transformation (as opposed to $1 + Y$). However, these results are driven by observations with small values of patenting, where we note that observations may have fewer than one patent because of our probabilistic mapping. If we additionally include a cutoff (e.g., only observations with greater than one patent) then the results are similar to the baseline results. Results are available upon request.

Table D.5: Leave-One-Out Instrument

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.665***	0.545***	0.052	1.022***	0.075	0.002
	(0.048)	(0.054)	(0.048)	(0.082)	(0.055)	(0.011)
$\ln EmbTechP$	0.003**	0.002*	-0.001	0.005*	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)
Observations	478,880	478,880	368,950	478,880	368,950	364,724
F-Stat $EmbTechK$	412.0	412.0	609.5	412.0	609.4	687.2
F-Stat $EmbTechP$	12804.0	12804.0	17580.2	12804.0	17580.1	18605.6
F-Stat $EmbTechDiag$	321.1	321.1	348.4	321.1	348.4	349.0

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 $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

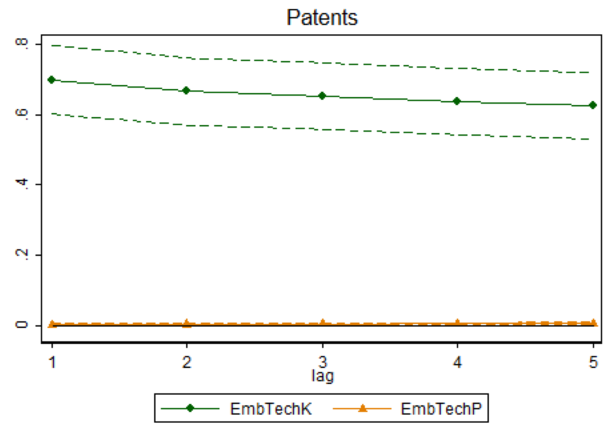
the US, European, and Japanese patent offices—and only foreign citations. Table D.10 reports the results for the diffusion outcomes when domestic citations are included in $USBackShare$, when the own-sector citations are included in the diffusion outcomes, and for the subset of triadic patents. We also find similar results, albeit quantitatively smaller, when we construct knowledge stocks using patent counts, rather than citation-weighted patenting (Table D.11). We suspect that citation-weighting patents helps to remove low-quality patents (e.g., frivolous patents) that do not capture much knowledge content. Consequently, we expect that this measure is noisier and, as a result, downward bias compared to the true estimate of knowledge spillovers.

Table D.6: Leave-One-Out Within Cluster Instrument

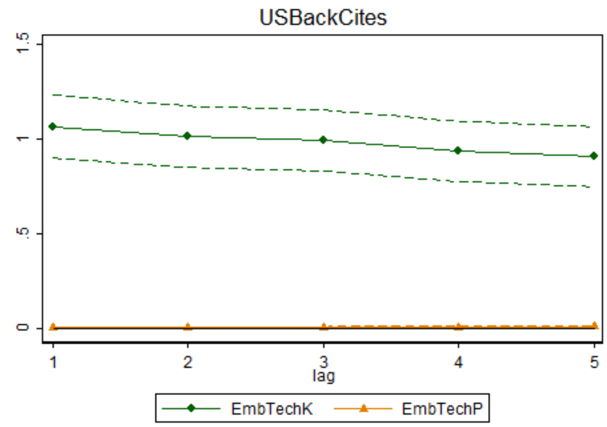
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.106***	0.013	-0.021	0.166***	0.024	-0.001
	(0.022)	(0.030)	(0.048)	(0.040)	(0.059)	(0.013)
$\ln EmbTechP$	0.004***	0.003***	-0.001	0.006***	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)
Observations	478,880	478,880	368,950	478,880	368,950	364,724
F-Stat $EmbTechK$	379.1	379.1	252.3	379.1	252.3	251.6
F-Stat $EmbTechP$	945.9	945.9	1482.1	945.9	1482.1	1587.5
F-Stat $EmbTechDiag$	88.1	88.1	62.9	88.1	62.9	63.3

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 $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

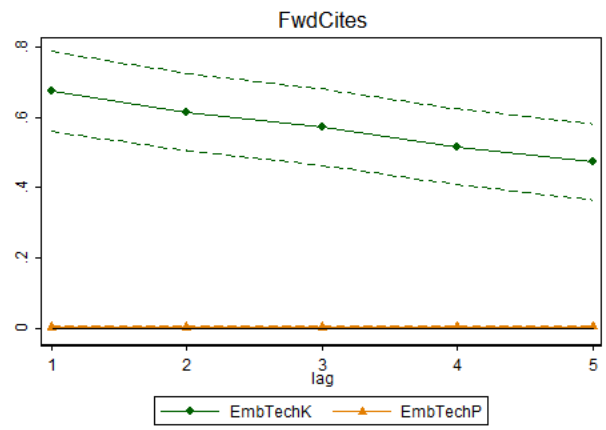
Figure D.1: Main Outcomes Estimated at Different Lags



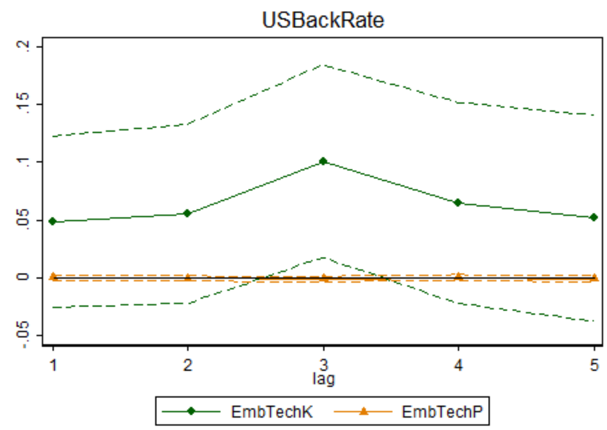
(a) *Patents*



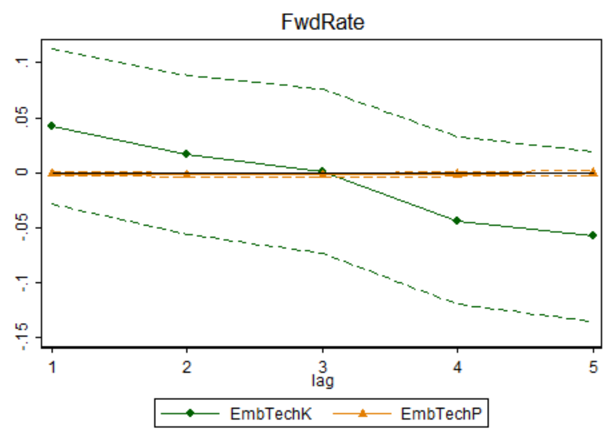
(b) *USBackCites*



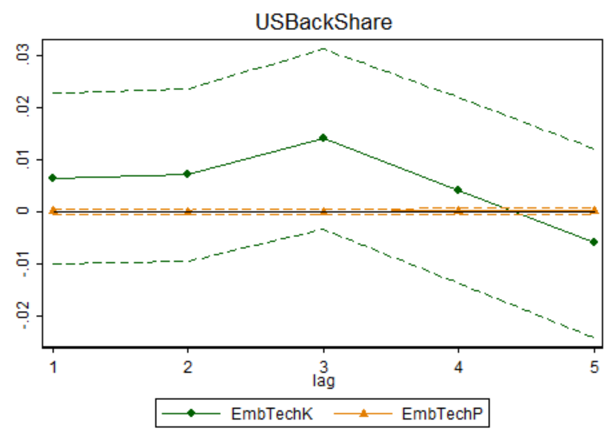
(c) *FwdCites*



(d) *USBackRate*



(e) *FwdRate*



(f) *USBackShare*

Notes: Coefficient estimates for five models based on Equation (11), in which the lag of the regressors is set from one to five years. Outcome variables are calculated as the value in period t.

Table D.7: $asinh(Y)$ Transformation

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.769***	0.600***	0.041	1.082***	0.095	0.002
	(0.055)	(0.061)	(0.058)	(0.090)	(0.063)	(0.011)
$\ln EmbTechP$	0.003**	0.002	-0.001	0.005*	0.000	0.000
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.000)
Observations	478,880	478,880	368,950	478,880	368,950	364,724
F-Stat $EmbTechK$	412.6	412.6	609.6	412.6	609.6	686.7
F-Stat $EmbTechP$	12806.4	12806.4	17535.9	12806.4	17535.9	18557.1
F-Stat $EmbTechDiag$	322.1	322.1	349.5	322.1	349.5	350.0

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 $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

Table D.8: Non-Zero Patent Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.796***	0.574***	0.052	1.099***	0.076	0.002
	(0.052)	(0.062)	(0.048)	(0.083)	(0.055)	(0.011)
$\ln EmbTechP$	0.004**	0.002	-0.001	0.005**	0.000	0.000
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.000)
Observations	368,950	368,950	368,950	368,950	368,950	364,724
F-Stat $EmbTechK$	609.6	609.6	609.6	609.6	609.6	686.7
F-Stat $EmbTechP$	17535.9	17535.9	17535.9	17535.9	17535.9	18557.1
F-Stat $EmbTechDiag$	349.5	349.5	349.5	349.5	349.5	350.0

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 $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

Table D.9: Other Innovation Outcomes

	(1)	(2)	(3)	(4)	(5)
	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$FwdCites$	$FwdRate$
$\ln EmbTechK$	0.311***	0.308***	0.126***	0.506***	0.065
	(0.027)	(0.040)	(0.047)	(0.053)	(0.049)
$\ln EmbTechP$	0.002**	0.001	-0.003**	0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Outcome	Triadic	Triadic	Triadic	Foreign	Foreign
Observations	478,880	478,880	264,815	478,880	368,950
F-Stat $EmbTechK$	412.6	412.6	1073.3	412.6	609.6
F-Stat $EmbTechP$	12806.4	12806.4	30214.5	12806.4	17535.9
F-Stat $EmbTechDiag$	322.1	322.1	298.5	322.1	349.5

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. Columns (1)-(3) are counts, forward citations, and forward citation rate of triadic patetns in country-sector-year. Columns (4)-(5) are forward citations received by all patents in a country-sector-year from foreign patents, and the corresponding citation rate. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

Table D.10: Other Diffusion Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IV	IV	IV	IV	IV	IV	IV
	$USBackShare$	$USBackCites$	$USBackRate$	$USBackShare$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.008	0.895***	0.050	-0.012	0.590***	0.225***	0.032***
	(0.011)	(0.071)	(0.043)	(0.011)	(0.050)	(0.045)	(0.010)
$\ln EmbTechP$	0.000	0.004**	-0.001	0.000	0.003**	-0.002	0.000
	(0.000)	(0.002)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Outcome	Inc. Domestic	Inc. Diagonal	Inc. Diagonal	Inc. Diagonal	Triadic	Triadic	Triadic
Observations	365,014	478,880	368,950	367,376	478,880	264,815	264,657
F-Stat $EmbTechK$	670.1	412.6	609.6	618.162	412.6	1073.3	1075.9
F-Stat $EmbTechP$	18634.2	12806.4	17535.9	18043.5	12806.4	30214.5	30302.5
F-Stat $EmbTechDiag$	350.2	322.1	349.5	348.6	322.1	298.5	298.0

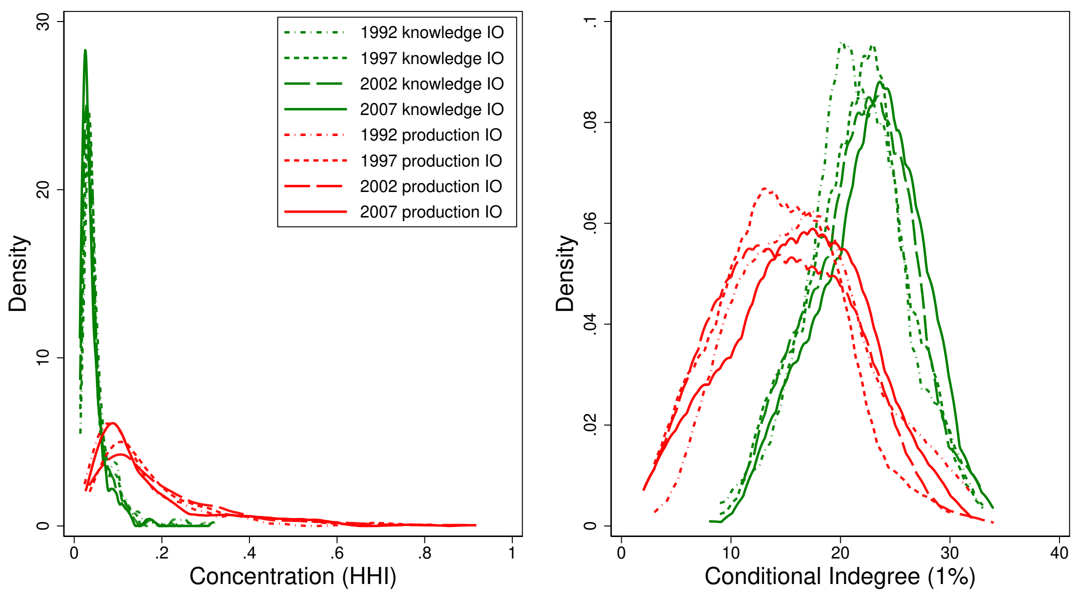
Notes: All dependent variables are first averaged over the three-year window t to $t+2$, and transformed as follows: $\ln(1+Outcome)$, where $Outcome$ is the variable specified on column titles. Column (1) is share of backward citations to US patents out of all citations including to domestic patents. Columns (2)-(4) include citations to own-sector patents. Columns (5)-(7) are citations made by triadic patents in a country-sector-year. Other controls include Lags of $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

Table D.11: Knowledge Stocks Constructed Using Patent Count

	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
	$Patents$	$FwdCites$	$FwdRate$	$USBackCites$	$USBackRate$	$USBackShare$
$\ln EmbTechK$	0.200***	0.194***	0.014	0.350***	0.014	-0.003
	(0.014)	(0.017)	(0.012)	(0.027)	(0.015)	(0.003)
$\ln EmbTechP$	0.005***	0.004**	-0.001	0.008***	0.000	0.000
	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)	(0.000)
Observations	478,880	478,880	368,950	478,880	368,950	364,724
F-Stat $EmbTechK$	5203.2	5203.2	4925.9	5203.2	4925.9	5038.8
F-Stat $EmbTechP$	12038.4	12038.4	16441.8	12038.4	16441.8	17044.8
F-Stat $EmbTechDiag$	380.2	380.2	496.052	380.2	496.1	497.9

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 $\ln EmbTechDiag$, $\ln OwnTechK$, $\ln USTech$, $\ln USTechDiag$, log total exports to world and log total imports from 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

Figure D.2: Distributions of Concentration and Sparsity of US IO Linkages



Notes: Figure plots the distributions of the concentration and conditional indegree measures of US knowledge and production IO linkages across output sectors for different IO table years. 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.

Table D.12: Summary Statistics of IO Linkage Concentration Measures

	N	Min	Max	Median	Mean	Std. Dev.
HHI-$K^h_{US,t}$ Across All Inputs						
1992	287	0.014	0.321	0.041	0.053	0.041
1997	287	0.014	0.300	0.038	0.049	0.036
2002	287	0.014	0.231	0.035	0.045	0.031
2007	287	0.013	0.319	0.032	0.041	0.031
All Years	1,148	0.013	0.321	0.037	0.047	0.035
HHI-$P^h_{US,t}$ Across All Inputs						
1992	287	0.024	0.611	0.112	0.139	0.097
1997	287	0.035	0.823	0.142	0.190	0.146
2002	287	0.034	0.900	0.149	0.192	0.149
2007	287	0.026	0.918	0.114	0.171	0.145
All Years	1,148	0.024	0.918	0.128	0.173	0.138
CID-$K^h_{USt}(1\%)$ Across All Inputs						
1992	287	9	33	21	21.387	4.492
1997	287	9	33	22	21.509	4.576
2002	287	11	33	22	21.906	4.618
2007	287	8	34	23	22.662	4.725
All Years	1,148	8	34	22	21.866	4.624
CID-$P^h_{USt}(1\%)$ Across All Inputs						
1992	287	3	32	17	16.808	6.055
1997	287	3	30	14	14.387	5.524
2002	287	2	34	15	14.948	6.135
2007	287	2	32	16	16.185	6.233
All Years	1,148	2	34	15	15.582	6.062
HHI-$K^h_{US,t}$ Across Off-Diagonal Inputs						
1992	287	0.013	0.260	0.036	0.044	0.030
1997	287	0.013	0.259	0.034	0.042	0.029
2002	287	0.014	0.231	0.032	0.040	0.027
2007	287	0.013	0.319	0.030	0.038	0.028
All Years	1,148	0.013	0.319	0.033	0.041	0.028
HHI-$P^h_{US,t}$ Across Off-Diagonal Inputs						
1992	287	0.024	0.822	0.110	0.145	0.113
1997	287	0.036	0.888	0.144	0.202	0.172
2002	287	0.035	0.929	0.138	0.200	0.175
2007	287	0.027	0.970	0.115	0.174	0.161
All Years	1,148	0.024	0.970	0.123	0.180	0.159
CID-$K^h_{USt}(1\%)$ Across Off-Diagonal Inputs						
1992	287	11	33	23	22.596	4.379
1997	287	11	33	23	22.676	4.644
2002	287	12	34	23	22.753	4.698
2007	287	8	34	24	23.254	4.916
All Years	1,148	8	34	23	22.820	4.664
CID-$P^h_{USt}(1\%)$ Across Off-Diagonal Inputs						
1992	287	2	34	17	16.969	6.238
1997	287	2	29	15	14.505	5.599
2002	287	1	34	15	15.268	6.220
2007	287	1	31	17	16.554	6.264
All Years	1,148	1	34	16	15.824	6.158

Notes: Table reports summary statistics of the cross-sectional distributions of the Herfindahl-Hirschman Index (HHI) and conditional indegree (CID) measures of US knowledge and production IO linkages. For measures computed using off-diagonal sectors, own-sector IO linkages are omitted from the calculation of the IO linkages. 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.