Discussion Paper Series

RIEB

Kobe University

DP2026-32

Identifying FDI Knowledge Spillovers Using Innovation Networks: Evidence from China

Li JIANG
Canlei LIU
Yue LV
Jing ZHANG
Laixun ZHAO

November 16, 2026

Research Institute for Economics and Business Administration

Kobe University

2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

Identifying FDI Knowledge Spillovers Using Innovation Networks: Evidence from China

Li Jiang Canlei Liu Yue Lv Jing Zhang Laixun Zhao *

November 15, 2026

Abstract

We examine knowledge spillovers along the innovation network, using data on foreign direct investment (FDI) firms in China and global patent records. We measure FDI-embedded technology with a *knowledge input–output (KIO) table* based on patent citations and compare it with the conventional production input–output (PIO) framework. KIO-based FDI-embedded technology raises domestic patenting, whereas the PIO analogue becomes insignificant when both are considered jointly. Regarding mechanisms, FDI-embedded technology increases cross-border backward citations, with collaboration as an important observable channel. Spillovers are stronger where technological gaps are larger and technologies are more central to the innovation network.

Keywords: *FDI; knowledge spillover; innovation network*
JEL Classification: *F23; O31; O33*

*Contact authors: Li Jiang, School of Economics, Renmin University of China, Beijing 100872, China, email: jiangli0029@163.com; Canlei Liu, School of Economics, University of International Business and Economics, Beijing 100029, China, email: canlei_liu@163.com; Yue Lv, School of International Trade and Economics, University of International Business and Economics, Beijing 100029, China, email: lvyue@uibe.edu.cn; Jing Zhang, School of Economics and Statistics, Guangzhou University, Guangzhou 510405, China, email: zh_jing19@163.com; Laixun Zhao, Research Institute for Economics & Business, Kobe University, Kobe 657-8501, Japan, email: zhao@rieb.kobe-u.ac.jp. We thank seminar and conference participants at China's Economic Transition Workshop, UIBE, Zhejiang University, and CUFE for helpful comments and suggestions. Canlei Liu thanks research support from National Natural Science Foundation of China (Grant No. 72573035). Yue Lv thanks research support from National Natural Science Foundation of China (Grant No. 72073025) and National Social Science Foundation of China (Grant No. 21&ZD098). Laixun Zhao is grateful to JSPS No. 24H00014. All errors are our own.

1 Introduction

Foreign direct investment (FDI) brings capital and know-how to host economies, which is particularly important for developing countries that lag behind the technological frontier. Whether and how FDI promotes technological progress in these countries has long been a central question to scholars and policy makers alike (Blomström and Kokko, 1998; Harrison and Rodríguez-Clare, 2010; Keller, 2021). Yet empirical evidence on FDI spillovers remains mixed (Havranek and Irsova, 2011). One reason, we argue, lies in how the spillovers are measured. Most existing studies trace them along production input–output (**PIO**) linkages, either across industries or, more recently, through supplier–customer relationships.[1] Knowledge flow, however, needs not dance to the same tune as goods flow. Innovation in one field can build on ideas originating in technologically related fields that are not its input suppliers, giving rise to knowledge input–output (**KIO**) linkages that need not coincide with production linkages.[2] In this paper, we ask whether FDI knowledge spillovers are better captured by an *innovation network* than by the production linkages commonly used in the literature.

We examine this distinction by combining near-complete registration records of foreign-invested enterprises (FIEs) in China with global patent data and detailed Chinese patent records. Following work that uses patent citations to map inter-field knowledge flows (Acemoglu et al., 2016; Liu and Ma, 2021; Cai et al., 2022; Ayerst et al., 2023), we construct a KIO table from backward citations on U.S. patents. An element ω_{ij} of the KIO table measures the share of field i's backward citations directed to field j, which is used to define the FDI-embedded intermediate technology in field i. Specifically, we aggregate the foreign knowledge that FDI brings into each upstream field j, mea-

[1] Industry-level input–output linkages are used by Aitken and Harrison (1999), Javorcik (2004), and Lu et al. (2017), among others; firm-level supplier–customer relationships are used by Alfaro-Ureña et al. (2022) and Amiti et al. (2024).

[2] Appendix Table B1 illustrates this for battery technology (H01M) by listing its top 20 upstream fields in the innovation and production networks: the former reflects the knowledge on which battery inventors build, the latter what battery producers buy. Through the innovation network, H01M draws on electric power storage (H02J), inorganic chemistry (C01B), catalytic processes (B01J), and capacitors (H01G); and through the production network, on metal alloys (C22C), semiconductor devices (H01L), building elements (E06B), and plastics shaping (B29C). FDI in energy storage or catalysis can thus expose battery technologies to foreign knowledge even when it does not directly enter battery manufacturing.

sured by source economies' knowledge stocks in j which are scaled by their investment in China, using ω_{ij} as weights. For comparison, we similarly construct a production analogue, PIO, using input shares from the U.S. benchmark use table, mapped to the same technology fields. The two measures therefore use the same FDI, foreign knowledge stocks, and technology resolution; their key difference lies in the networks through which upstream foreign technology is aggregated.

To address endogenous allocation of FDI, we exploit the 2002 revision of China's *Catalogue for the Guidance of Foreign Investment Industries* (hereafter the *Catalogue*), which substantially eased restrictions on foreign investment in connection with China's WTO accession. We first estimate how the reform changed FDI across technology fields and use the resulting policy-induced investment component to construct an instrument for FDI-embedded technology. Specifically, we propagate this investment response through the pre-reform knowledge network while holding source-economy knowledge stocks at their 2001 levels.[3] Also, we construct the PIO instrument symmetrically, using the same policy-induced investment and foreign knowledge stocks but replacing citation weights with production weights. This symmetric design allows us to ask whether, given the same policy-induced variation in FDI, KIO exposure captures the foreign technology shock relevant for domestic innovation where its PIO analogue does not.

Our central finding is that FDI knowledge spillovers propagate along the innovation network rather than the production network. Foreign technology arriving through knowledge linkages raises domestic patenting. Nevertheless, considered on its own, production-network exposure also predicts higher patenting, which is the pattern the conventional approach relies on. However, once both exposures are instrumented jointly, the knowledge-network effect is virtually unchanged, whereas the production-network effect loses statistical significance. Additional evidence suggests

[3] Identification rests on the assumption that, conditional on the controls and fixed effects, reform-induced variation transmitted through predetermined knowledge linkages affects innovation only through FDI-embedded technology. It requires that reform-induced exposure be uncorrelated with other determinants of innovation in receiving fields and that it affect innovation only through FDI-embedded technology. We address the first condition by allowing pre-reform field characteristics to generate differential trends, documenting similar pre-reform investment paths in liberalized and unchanged fields, and controlling for concurrent WTO-related policies. We address the second with a placebo instrument built from source economies without pre-reform investment in China, which does not predict patenting, and by showing that the effect persists among fields whose own access status did not change and when own-field liberalization is controlled for (Sections 4.2 and 5.3).

that this pattern is not driven by measurement error in production linkages or by simple collinearity between the two exposures. Our finding does not imply that production linkages cannot transmit knowledge. Rather, a production-network measure considered in isolation may attribute to production linkages an innovation response that operates through underlying technological relationships, thereby mischaracterizing the channel through which FDI-driven technology diffusion occurs.

Mechanism evidence further supports the knowledge-spillover interpretation. We find that greater FDI-embedded technology leads domestic inventors to make more cross-border backward citations, and this response is not explained by foreign-affiliated patenting in China or by a general acceleration in access to global knowledge. The evidence also points to interaction-based exchanges (e.g., joint patents) rather than market-based transfers (e.g., patent assignments, licensing) as the primary channel. The collaboration response is inconsistent with a simple *quid pro quo* (technology for market access) explanation (Bai et al., 2020). Finally, spillovers are larger where the technological gap is greater and where the receiving field is more central in the innovation network. This is consistent with greater returns where there is more to learn and where knowledge can propagate more broadly.

Our paper contributes to three strands of literature. First, in the extensive literature on FDI spillovers,[4] a longstanding debate is whether estimated "FDI spillovers" reflect knowledge transmission or instead confound other channels such as scale, competition, or compositional change. Earlier work documents positive effects of foreign entry on domestic productivity and innovation, especially through vertical linkages using PIO tables (Javorcik, 2004), while more recent contributions emphasize technological proximity as a key condition for learning from foreign firms (Fons-Rosen et al., 2017). We contribute to this literature by measuring FDI exposure through an explicit knowledge network and comparing it directly with the conventional production-network approach under the same identifying variations. In this respect, our work is also related to the dis-

[4]Early studies emphasize horizontal channels, including demonstration and competition effects (Aitken and Harrison, 1999; Branstetter, 2001), while Javorcik (2004) shifts attention toward vertical linkages and establishes industry input–output relationships as a benchmark for studying FDI spillovers. Subsequent work examines learning and knowledge-transfer mechanisms, including those operating through joint ventures (Bai et al., 2020; Jiang et al., 2024; Ma and Zhang, 2025).

tinction drawn by Bloom et al. (2013) between technology spillovers and product-market rivalry, but transposes the distinction to an international setting and operationalizes the technology channel through inter-field knowledge flows.

Second, our study speaks to the literature on the drivers and micro-level mechanisms of international knowledge diffusion. While major technologies historically emerged in advanced countries and gradually spread to less developed ones, cross-border knowledge flows remain highly localized due to language barriers, institutional mismatches, and divergent patent systems (Jaffe et al., 1993; Jaffe and Trajtenberg, 1999; Hegde et al., 2023). Openness through trade and investment can facilitate access to the global knowledge frontier and contribute to long-run growth (Grossman and Helpman, 1991; Coe and Helpman, 1995; Buera and Oberfield, 2020; Cai et al., 2022), with multinational firms serving as important vehicles for transmitting knowledge across borders (Keller, 2021). We contribute to this literature by leveraging comprehensive firm registration and patent data to trace both the use of foreign knowledge and observable channels through which it is transmitted across technology fields.

Third, we contribute to the growing literature on innovation networks. Recent research shows that innovation networks can differ substantially from production networks, with relatively weak correspondence between technological knowledge linkages and input–output relationships (Liu and Ma, 2021; Ayerst et al., 2023). This distinction implies that production-based measures may fail to capture the technology relationships through which knowledge propagates across fields. Building on this insight, we apply patent-based KIO tables to the study of FDI, capturing the heterogeneous interconnections across technology fields through which foreign knowledge propagates and interacts with the host-country innovation system.

The remainder of the paper is organized as follows. Section 2 describes China's foreign-investment policy and the 2002 *Catalogue* reform. Section 3 describes the data and construction of the knowledge and production networks. Section 4 presents the empirical strategy. Section 5 reports the main results and compares the two networks. Section 6 examines mechanisms and heterogeneity. Section 7 concludes.

2 Institutional Background

China's Reform and Opening-up in 1978 marked a major milestone in the country's integration into the global economy. The *Law on Chinese-Foreign Equity Joint Ventures* (1979), the *Law on Wholly Foreign-Owned Enterprises* (1986), and the *Law on Chinese-Foreign Contractual Joint Ventures* (1988), collectively known as the "Three FDI Laws," established the legal framework for foreign-invested enterprises (FIEs). FDI inflows grew rapidly thereafter, especially after Deng Xiaoping's southern tour in 1992, rising from $4.37 billion in 1991 to over $27 billion in 1993.

China initially pursued FDI to address capital shortages. As domestic savings and foreign exchange reserves improved, the policy focus shifted toward attracting foreign technology. In 1995, the government introduced the *Catalogue for the Guidance of Foreign Investment Industries*, which classified industries into four categories: encouraged, permitted, restricted, and prohibited. Inward FDI in high-tech industries subsequently increased sixteenfold between 1998 and 2018, reaching 23.5% of total FDI.[5]

We construct our instrument using the 2002 revision of the *Catalogue*. To meet its WTO accession commitments, China overhauled the *Catalogue* in 2002, significantly loosening FDI restrictions in more than one-quarter of manufacturing industries (Liu et al., 2024).[6] Over time, mainland China gradually emerged as one of the world's top FDI destinations. After a brief adjustment during the 2008 global financial crisis, China's actual FDI inflows stabilized above $100 billion annually.

Anecdotal evidence illustrates how such investment can generate knowledge spillovers beyond the investor's own production chain. ZEISS, a German supplier of optical and precision-metrology technologies with manufacturing operations in China for over two decades, does not directly serve battery manufacturing (H01M). Its capabilities in inspection, precision imaging, and materials analysis (G01N), however, are applied throughout battery research and production, as leading pro-

[5]Source: Special lecture series of the Standing Committee of the 13th National People's Congress, http://www.npc.gov.cn/zgrdw/npc////xinwen/2019-02/20/content_2071908.htm.

[6]This was the first systematic revision of the *Catalogue* after China's WTO entry. Although subsequent revisions followed in 2004, 2007, 2011, 2015, and 2017, the 2002 revision was a pivotal step in liberalization, sharply increasing FDI openness and triggering sustained growth in foreign investment inflows (Lu et al., 2017).

ducers such as CATL have intensified quality control.[7]

3 Data and Measurement

3.1 Data

We draw on three main data sources: firm-level registration records of FIEs in China, global patent data, and Chinese patent records on assignments, licenses, and joint applications involving FIEs. The estimating sample is a balanced panel of 576 IPC4 technology fields observed annually over 1996–2016. Details on data processing are in Appendix A.

Firm-level data come from the Comprehensive Management Platform for Foreign Investment of the Ministry of Commerce (MOFCOM). Because registration or approval was mandatory, the platform covers every FIE registered before 2019, recording firm characteristics such as establishment date, industry, and ownership structure, together with each investor's origin and investment amount. We identify 824,565 FIEs established before 2019, involving investors from 205 economies across 1,364 four-digit China Industry Classification (CIC4) industries.[8]

Because FDI is recorded by industry and the analysis is conducted by technology field, we link the two using the *Reference Table between International Patent Classification and the Classification of National Economic Activities (2018)*, compiled by the China National Intellectual Property Administration (CNIPA).[9] Where an industry maps to several fields, we divide its FDI equally

[7] According to an industry report by NE Times (https://d1ev.com/kol/209799, September 2023), ZEISS's focused ion beam and field-emission scanning electron microscopes are used to analyze the morphology, composition, and three-dimensional structure of battery materials; its X-ray microscopes provide non-destructive 3D scans of electrodes, yielding parameters such as porosity and tortuosity that help manufacturers optimize process parameters; and its industrial CT systems are used for defect identification and failure analysis of finished cells, modules, and packs. The report situates these tools within the "extreme manufacturing" strategy of leading producers such as CATL, whose production process includes more than 6,800 quality-control points.

[8] Appendix Figure A1 shows that annual counts and registered capital closely track published national aggregates. Appendix Tables A1 to A3 report the distribution of entrants by source economy, industry, and technology field, and Appendix Table A4 reports summary statistics. Hong Kong is the largest source, accounting for 41.27% of new FIEs, followed by Taiwan (10.52%), South Korea (9.79%), the United States (6.91%), and Japan (5.71%). Offshore financial centers such as the British Virgin Islands and Samoa also account for a sizable share, reflecting tax-motivated routing; Appendix Table D2 controls for their influence.

[9] Appendix A documents the concordance and Appendix Table A7 reports its coverage and density.

among them to avoid double counting.[10]

Global patent data come from the Autumn 2023 release of PATSTAT Global. We group 92.3 million applications into 60.2 million DOCDB families so that an invention filed in multiple jurisdictions is counted once.[11] Restricting attention to families filed between 1996 and 2016 and granted within the observation window yields a citation network with 6.49 million citing and 5.92 million cited families. Innovation output is the number of granted patent families originating in China in field i, dated by application year t; unlike the grant year, the application year reflects the decision to seek protection rather than examination delays.[12]

Records of assignments and licenses registered with CNIPA, and of joint applications involving FIEs in China, come from the IncoPat Global Patent Database for 1985–2022. Matching FIE names from the MOFCOM registry to CNIPA applicants, we identify 10,668 assignments from FIEs to non-FIEs, 1,138 licenses, and 44,885 applications filed jointly by an FIE and at least one domestic partner.[13] Appendix Figure A2 shows all three series rising over the sample period, with joint applications accelerating after 2005 and becoming the most prevalent.

3.2 FDI-embedded Intermediate Technology

Our measure builds on the knowledge production function approach to technology spillovers (Griliches, 1979; Jaffe, 1986). In this framework, innovation in a unit depends not only on its own research inputs but also on a pool of external knowledge, constructed as a weighted sum of other units' knowledge stocks, with weights reflecting how useful each source is to the recipient. Because the spillover pool enters the knowledge production function multiplicatively, it appears in logs in the patent equation, and its coefficient captures the elasticity of innovation with respect to accessible

[10] Appendix Table D3 instead assigns an industry's full FDI to each mapped field. Appendix Table D5 reports sensitivity checks using alternative concordances and aggregation levels.

[11] Patent rights are territorial, so applicants seeking protection in multiple jurisdictions file separately, using the priority system to preserve the original filing date. A DOCDB family groups filings that share a common priority.

[12] Appendix Table D4 reports alternative outcome measures, including forward citations and patents granted by offices other than CNIPA.

[13] Joint applications are filed by separate legal entities and therefore exclude innovation carried out entirely within a joint venture.

external knowledge.[14]

We adapt this structure to foreign knowledge that FDI brings into upstream technology fields. For each technology field i and year t, we define the knowledge input–output spillover as

$$KIOSPILL_{it} = \sum_j \omega_{ij} \sum_c K_{cjt} FDI_{cjt}. \tag{1}$$

The inner sum, $\sum_c K_{cjt} FDI_{cjt}$, plays the role of the source unit's knowledge stock (G_j in Bloom et al., 2013): it is the pool of foreign knowledge that multinational firms bring into China's technology field j. FDI_{cjt} is investment from source economy c into field j in year t, and K_{cjt} is economy c's knowledge stock in field j, so that a dollar of investment from a technologically stronger source carries more knowledge. The outer weight, ω_{ij}, plays the role of the proximity weight ($TECH_{ij}$): it is the share of field i's backward citations directed to field j, measuring how much field i builds on knowledge from field j. Unlike the symmetric proximity measure in Bloom et al. (2013), ω_{ij} is directed, consistent with evidence that knowledge flows between technology fields are asymmetric (Acemoglu et al., 2016). Field i is thus more exposed to foreign technology when the fields it builds on attract more FDI from economies with stronger knowledge bases in those fields. If FDI is an effective channel of technology diffusion, greater exposure should raise innovation in field i.

Knowledge stocks accumulate with depreciation,

$$K_{cjt} = (1-\delta)K_{cjt-1} + FwdCite_{cjt},$$

where $FwdCite_{cjt}$ counts citations received within five years by patents that economy c filed in field j in year t, and $\delta = 0.05$ following Ayerst et al. (2023). The fixed citation window equalizes citation opportunities across patent cohorts. The initial stock is set to $K_{cj1940} = FwdCite_{cj1940}/\delta$, a choice with negligible influence given the long accumulation period.[15]

[14] Bloom et al. (2013) formalize this approach and provide its microfoundation. In their model, a firm's knowledge output depends on its own R&D and on a technology spillover pool, $SPILLTECH_i = \sum_{j \neq i} TECH_{ij} G_j$, where G_j is firm j's R&D stock and $TECH_{ij}$ measures the technological proximity between firms i and j.

[15] Appendix Table D3 reports estimates using a 15% depreciation rate, the conventional choice for R&D stocks (Bloom et al., 2013), a ten-year citation window, knowledge stocks constructed from patent counts rather than citation-

3.3 The Innovation Network

Following Acemoglu et al. (2016), we treat citations from patents in one field to patents in another as knowledge inputs and define

$$\omega_{ij} = \frac{cite_{ij}}{\sum_k cite_{ik}}, \tag{2}$$

where $cite_{ij}$ is the number of citations from field i to field j. The resulting matrix Ω is the KIO table: ω_{ij} is the share of field i's knowledge inputs supplied by field j, just as a PIO coefficient is the share of an industry's intermediate purchases supplied by another industry. Baseline weights are computed from citations on U.S. patents over 1996–2016; the instrument uses weights from the pre-reform window (1992–2001).

Because each row is normalized by the citing field's total backward citations, only the composition of what a field cites enters, not how much it cites. Building the network from patents of a single origin subjects all fields to common citation practices, and using U.S.-origin rather than Chinese-origin patents keeps the network separate from the records that supply the dependent variable. The network is also highly stable: Appendix Table B2 reports correlations above 0.9 between the 1996–2016 network and those computed over three sub-periods and the pre-reform window, and Appendix Table B3 reports correlations above 0.8 among the networks of the 10 largest patenting economies, with China's network correlated 0.95 with the global network.

3.4 The Production Network

The existing FDI spillover literature typically assumes that foreign knowledge travels along supply chains, as multinationals transfer technology to local suppliers or sell more advanced inputs to domestic customers (Javorcik, 2004). Empirical work captures these vertical channels by weighting foreign presence in upstream or downstream industries by input–output coefficients (Javorcik, 2004; Lu et al., 2017).

We take production weights from the 1997 benchmark use table of the U.S. Bureau of Eco-

weighted patents, and an alternative rule for allocating FDI across fields.

nomic Analysis, which predates the 2002 reform. For each using industry, we compute input shares over total intermediate inputs, excluding value added. We map input–output commodities and industries to four-digit NAICS 1997 codes and then to IPC4 fields using the probabilistic concordance of Lybbert and Zolas (2014), applying the concordance probabilities to supplying commodities and using industries separately and summing the resulting shares over field pairs. Denoting the resulting weights by v_{ij}, we define production-network spillovers as

$$PIOSPILL_{it} = \sum_{j} v_{ij} \sum_{c} K_{cjt} FDI_{cjt}. \tag{3}$$

$PIOSPILL_{it}$ thus differs from $KIOSPILL_{it}$ only in the network that aggregates foreign knowledge.

3.5 Innovation vs Production Networks

Figure 1 maps the two networks at the IPC3 level. Ranking fields by their off-diagonal column sums, which measure their importance as upstream sources for all other fields, reveals a sharp contrast. Apart from basic electric elements (H01), which ranks highly in both, the innovation network is dominated by general-purpose technologies such as medical science (A61), computing (G06), measurement (G01), and physical or chemical processes (B01), whereas the production network is dominated by fields supplying physical inputs, such as metal working (B21), weapons (F41), metallurgy (C22), and building (E04).[16] The same contrast holds for individual fields, as the battery example in Appendix Table B1 illustrates.

[Figure 1 here]

Across the technology fields for which both networks are defined, the Pearson correlation between ω_{ij} and v_{ij} over the 363,609 ordered field pairs is 0.17; at the IPC3 level, it is 0.32. This weak correspondence, consistent with earlier evidence that knowledge linkages differ substantially

[16] Appendix Tables B4 and B5 report the complete rankings at the IPC3 and IPC4 levels, respectively, and Appendix Figure B1 compares the two networks at the IPC4 level used in the estimation.

from input–output relationships (Liu and Ma, 2021; Ayerst et al., 2023), provides the independent variation needed to distinguish knowledge from production linkages.

4 Empirical Strategy

Of the three components of $KIOSPILL_{it}$ in equation (1), two are largely outside China's control. The knowledge weights ω_{ij} are computed from U.S. patent citations and reflect the technological structure of invention rather than Chinese conditions, and the knowledge stocks K_{cjt} reflect foreign economies' own research histories. Endogeneity is therefore concentrated in FDI_{cjt}, because multinational firms choose where to invest. Fields with faster growth in Chinese innovation may attract more technologically sophisticated entrants, creating reverse causality, and unobserved factors such as industrial policy, local institutions, or demand shifts may simultaneously attract foreign investment and stimulate domestic patenting.

4.1 Instrumental Variables

Our strategy replaces realized foreign investment with the component induced by a national policy change. The 2002 revision of the *Catalogue*, described in Section 2, moved roughly a quarter of Chinese industries along a four-step ladder of foreign access in connection with China's WTO accession. We use this reclassification to estimate the proportional investment response to liberalization and propagate it through technology fields using weights fixed before the reform.[17]

Comparing the 1997 and 2002 editions of the *Catalogue*, which assign each four-digit industry to one of four access categories, 339 of 1,346 industries moved to a more liberal category, 984 were unchanged, and 23 were tightened (Appendix Table C1). We map industries into technology fields using the concordance in Section 3.1, compute the share of each field's constituent industries that were liberalized, and set $Treat_j = 1$ when this share exceeds the median.[18] $Post_t$ equals one

[17] Estimating the investment equation first, rather than propagating the liberalization indicator directly, documents that the reform raised foreign investment conditional on pre-reform characteristics and concurrent policies, and expresses the instrument in units of policy-induced proportional changes in investment.

[18] Results are not sensitive to this threshold. Table 1 reports the investment equation with $Treat_j = 1$ for any positive

from 2003 onward and 3/4 in 2002, reflecting the April implementation of the revised *Catalogue*.

We estimate the policy-induced component of investment using a difference-in-differences specification at the source-economy × field × year level:

$$\ln(FDI_{cjt}+1) = \alpha + \pi\,(Treat_j \times Post_t) + \Gamma' Z_{jt} + \mu_{cj} + \delta_t + \eta_{cjt}, \tag{4}$$

where μ_{cj} absorbs time-invariant differences in source economies' propensity to invest in each field and δ_t absorbs aggregate investment shocks. Z_{jt} interacts year indicators with pre-reform field characteristics and the 1995 restricted category, allowing fields with different initial conditions to follow different paths, and includes the contemporaneous policy controls described below. Standard errors are two-way clustered by source economy and technology field. The coefficient π captures the differential change in foreign investment into liberalized fields after 2002.

Let $\widehat{FDI}^{policy}_{cjt} = \hat{\pi}(Treat_j \times Post_t)$ denote the policy-induced component of investment. We aggregate it in the same way as the endogenous regressor, fixing the other two components at pre-reform values:

$$IV^{KIO}_{it} = \sum_j \omega^0_{ij} \sum_c K_{cj,2001}\, \widehat{FDI}^{policy}_{cjt}, \tag{5}$$

where ω^0_{ij} is computed from citations over 1992–2001 and knowledge stocks are fixed at their 2001 values. Post-reform variation in the instrument is thus generated solely by the policy shock, while the network structure and foreign technological capabilities are held at pre-reform levels; cross-field variation arises from differences in fields' pre-reform dependence on knowledge from subsequently liberalized fields.

The production-network instrument replaces only the citation weights with the production input shares ν_{ij} from equation (3):

$$IV^{PIO}_{it} = \sum_j \nu_{ij} \sum_c K_{cj,2001}\, \widehat{FDI}^{policy}_{cjt}. \tag{6}$$

liberalized share, Appendix Table D1 the corresponding second stage, and Appendix C a continuous measure of movement along the access ladder.

The two instruments thus propagate the same policy-induced shock through different networks, allowing a horse race in Section 5.2 to assess which network better captures the foreign technology exposure relevant for domestic innovation.

4.2 Validity of the Instrument

Instrument validity requires relevance and an exclusion restriction.[19]

Relevance. Relevance rests on two links: the 2002 reform must raise foreign investment in liberalized fields, and this investment must translate into differential exposure across receiving fields through their predetermined knowledge dependence on those fields. Table 1 establishes the first link. Liberalization increases foreign investment by 0.141 log points under the baseline treatment definition and by 0.189 when any positive liberalized share counts as treatment, both significant at the 1% level. Section 5.1 shows that the resulting instrument strongly predicts realized FDI-embedded technology.

[Table 1 here]

Exclusion. The exclusion restriction requires that, conditional on the controls and fixed effects, pre-reform knowledge exposure to fields affected by the 2002 liberalization influences subsequent patenting only through the resulting FDI-embedded technology. We consider four threats. The first two concern whether this exposure is correlated with other determinants of innovation in receiving fields; the last two concern whether the reform affects innovation through channels other than foreign investment. Section 5.3 reports the corresponding evidence.

The first threat is endogenous selection into liberalization. China's accession commitments emerged from bilateral and multilateral negotiations and may reflect the interests of foreign firms

[19] Although the instrument has a shift-share-like structure, propagating a common policy change through predetermined network weights, identification does not rely on many independently realized shocks: the underlying shock is the single 2002 reform, and cross-field variation arises from predetermined exposure to it. The final instrument also applies a log transformation to the aggregate and so does not take the standard linear shift-share form. We therefore conduct inference at the technology-field level rather than applying many-shock asymptotics.

or the priorities of Chinese policy makers. Such selection threatens identification only if the determinants of liberalization are correlated with the counterfactual innovation paths of receiving fields exposed to liberalized fields through knowledge linkages. Because the policy is assigned to industries while identifying exposure is generated by technological linkages, which Section 3.5 shows are only weakly related to production linkages, industry-level determinants of liberalization do not map mechanically into the variation we use. More directly, following Lu et al. (2017), we interact pre-reform field characteristics with year indicators in both the investment and patent equations to allow unrestricted differential trends,[20] and the investment equation additionally interacts the 1995 restricted category with year indicators. Panel (a) of Figure 2 reports an event-study version of equation (4). Foreign investment in liberalized and unchanged fields shows no clear differential movement before the reform, followed by a sustained relative increase after 2002.[21]

[Figure 2 here]

The second threat is that the *Catalogue* revision coincided with other accession-related changes that could generate innovation trends correlated with exposure. Both the investment equation (4) and the patent equation (7) control for (1) trade liberalization, using 2001 input-, output-, and partner-weighted tariffs (Brandt et al., 2017); (2) SOE reform, using the 2001 SOE share (Hsieh and Song, 2015); (3) U.S. PNTR, using the 1999 tariff gap (Pierce and Schott, 2016); (4) textile and clothing quota removal, using product-level quota fill rates (Brambilla et al., 2010); and (5) TRIPS commitments, using a classification of intellectual-property-intensive fields. The tariff and SOE measures are interacted with year indicators, and the PNTR and TRIPS exposures with an indicator for years from 2001 onward. Quota removal enters as a time-varying fill rate that falls to zero as quotas are eliminated, from 2002 for the first three phases and from 2005 for the final phase (Appendix A.5). Domestic innovation policies introduced later may themselves respond to

[20] These characteristics are measured in the 1998 Annual Survey of Industrial Enterprises (ASIE) at the CIC4 level and mapped to IPC4 fields using the concordance in Section 3.1.

[21] Appendix Figure C2 examines patenting directly. Liberalized and unchanged fields differ in levels but show no clear sustained divergence before the reform; given some year-to-year variation in the coefficients, we regard this evidence as suggestive.

innovation outcomes, so we exclude them from the baseline but add them as robustness checks.[22]

The third threat is that liberalization raised access to foreign knowledge in liberalized fields through channels other than FDI, which would then reach receiving fields through the same knowledge linkages. Because the instrument weights liberalized fields by the knowledge stocks of all source economies, it would capture these channels as well as foreign investment. We test this threat using a placebo component of the instrument built from the 41 of 192 source economies that had no FDI in China in any year of 1999–2001. Their knowledge stocks enter the instrument through the same network weights, but without a pre-reform presence in China they have little scope to respond to liberalization through investment. Panel (b) of Figure 2 confirms this: their investment in liberalized fields shows no sustained response to the reform, whereas that of economies with baseline FDI rises after 2002. If the instrument affected patenting through non-FDI channels, this component should also predict patenting. Because it shares the treatment assignment and network structure of the baseline instrument, the test also provides a check on selection.

The fourth threat is that liberalization affects patenting in the receiving field directly through channels other than foreign investment, for example by relaxing joint-venture requirements that govern technology sharing or by changing access to imported inputs and equipment. Foreign entry into the receiving field itself does not pose such a threat, because same-field investment is part of FDI-embedded technology. We address the remaining own-field channels in two ways: by restricting the sample to fields whose own access status did not change, so that no own-field policy channel is active, and by controlling directly for the receiving field's own liberalization interacted with the post-reform indicator, which absorbs any direct effect common to liberalized fields. Baseline controls for field-level imports and 2001 input tariffs interacted with year indicators further account for changes in access to imported inputs.

What the estimates identify. The 2SLS estimate is a local average treatment effect (Imbens and Angrist, 1994), identified from source-economy and field pairs whose investment responded to the 2002 liberalization, not from fields the reform left untouched or investors who would have

[22] These include R&D tax relief, high-technology enterprise certification, development zones, and the expansion of higher education, most of which took effect from 2006 or 2008.

entered regardless. Policy-induced entry was concentrated among source economies with stronger technology in the field, whereas the expansion of existing investment shows no such tilt (Appendix Table E3).

Two qualifications bear on magnitude. First, because the instrument comes from a single, permanent reform, post-reform observations combine responses at different horizons, and the same variation cannot separately instrument exposure at each horizon. We therefore interpret the estimate as summarizing the response over the post-reform sample rather than as an annual elasticity; Appendix Figure D1 describes the OLS association by horizon. Second, liberalization may have changed the composition as well as the volume of foreign investment, and residual effects of the broader accession process cannot be ruled out. We therefore interpret the estimate as the effect of *Catalogue*-induced foreign investment transmitted through the knowledge network, rather than as a causal parameter on embedded foreign knowledge alone.

4.3 Specification

We estimate

$$\ln patnum_{it} = \alpha + \beta \ln KIOSPILL_{it-1} + \gamma' X_{it-1} + \mu_i + \delta_t + \varepsilon_{it}, \tag{7}$$

where i indexes IPC4 technology fields, t indexes years, and $\ln KIOSPILL_{it-1}$ is instrumented by $\ln IV^{KIO}_{it-1}$. Lagging the right-hand-side variables by one year attenuates reverse causality, since a field on an upward trajectory may attract investment both before and after its innovation accelerates. The dependent variable is the log of one plus the number of granted Chinese patent families in field i, dated by application year t (Section 3.1). Field fixed effects μ_i absorb time-invariant differences such as field size, citation practices, and network position; year fixed effects δ_t absorb aggregate shocks such as the secular growth of Chinese patenting. Standard errors are clustered by technology field.[23]

[23] Patenting is right-skewed, but zeros account for only 9.85% of field-year observations, and neither exposure measure contains zeros (Appendix Figure A3 and Table A6). We also estimate PPML specifications in levels (Chen and Roth, 2024). Mechanism outcomes contain more zeros, from 26% for backward citations to 96% for licenses, so for these we also report Poisson and extensive-margin specifications. Appendix Tables D3 and D4 examine alternative

The control vector X_{it-1} contains three groups of variables. The first captures two non-FDI sources of foreign technology exposure: global technological progress, measured by log patent applications outside China in the field, and learning through trade, measured by the log value of Chinese imports in the field (Bloom et al., 2016; Liu and Qiu, 2016). The second interacts year indicators with pre-reform field characteristics, namely new-product intensity, export intensity, the number of firms, and average firm age in 1998. The third captures the WTO-related policy shocks described in Section 4.2.

5 Results

5.1 Baseline Results

Table 2 establishes the effect of FDI-embedded technology on innovation in China. Columns 1 and 2 report OLS estimates of equation (7). Column 1 includes technology-field and year fixed effects, while column 2 adds the full set of controls. The estimated elasticity is positive and significant at the 1% level in both specifications and is virtually unchanged by the controls, moving from 0.173 to 0.172. Column 6 estimates the corresponding relationship by PPML using patent counts in levels. The coefficient remains positive and significant, although smaller at 0.066. We view this result as evidence that the positive association is not an artifact of the log transformation or the treatment of zero patent counts.[24]

[Table 2 here]

Columns 3 and 5 report the first-stage and reduced-form relationships underlying the IV estimate. The first stage in column 3 confirms that the policy-induced investment shock translates into

constructions of the explanatory variable and outcome.

[24] Appendix Figure G1 decomposes ln *KIOSPILL* into counterfactual components that alternately hold 2001 knowledge stocks and FDI fixed. Growth in foreign knowledge stocks accounts for much of the raw time-series variation but is largely absorbed by the fixed effects; the residual within-field variation closely tracks the FDI-driven component targeted by the instrument.

field-level exposure: the coefficient is 0.054 and significant at the 1% level, with a Kleibergen–Paap rank F statistic of 12.46, rising to 15.12 when production-network spillovers are included as a control. Column 5 reports the reduced form, linking patenting directly to the instrument; the coefficient is positive and significant at the 1% level. Because the reduced form does not scale the effect on patenting by the estimated first stage, it provides direct evidence that fields with greater pre-reform knowledge exposure to fields affected by the liberalization subsequently experienced greater patenting.

Column 4 reports the 2SLS estimate of 2.000, significant at the 1% level. Taken at face value, the estimate implies that a 10% increase in FDI-embedded intermediate technology is associated with approximately a 20% increase in patenting. The estimate is more than ten times the corresponding OLS coefficient, and this difference warrants caution in interpreting its magnitude. Reverse causality, multinationals investing disproportionately in fields where innovation is already strong, would tend to bias OLS upward and therefore cannot by itself account for the larger IV estimate. Three considerations are particularly relevant.[25]

First, the 2SLS coefficient scales the reduced form by a moderately strong first stage, so uncertainty in the first stage can produce a large and imprecise magnitude. Second, the estimate is local to the margin shifted by the reform: policy-induced entry was concentrated among technologically stronger sources (Appendix Table E3), and the return to exposure is larger where China's technology gap is wider (Table 8). Third, because the instrument derives from a permanent reform, the coefficient reflects a cumulative medium-run response rather than an annual elasticity (Section 4.2).[26] We therefore treat the positive reduced-form and IV relationships as the central evidence, and interpret the size of the 2SLS coefficient with these qualifications in mind.

[25] Appendix E examines the OLS–IV gap in detail. Appendix Figure C3 and Appendix Tables E2 and E3 characterize the investment margin shifted by the reform; Appendix Table E1 shows that classical measurement error in the composite regressor is unlikely to account for the gap; and Appendix Tables E4 and E5 show that the gap narrows substantially when OLS and IV are compared at a common, lower frequency.

[26] Appendix Figure D1 reports an unrestricted distributed-lag specification over five lags together with the implied cumulative response.

5.2 Comparison with Traditional Production Networks

This section tests whether FDI knowledge spillovers operate through the innovation network or through the production linkages conventionally used in the FDI literature. Because the KIO and PIO measures and instruments differ only in the network weights (Sections 3.4 and 4.1), the key question is not which measure yields the larger coefficient in isolation, but which retains explanatory power once the other is accounted for.

Table 3 runs a horse race between the two networks, entering each exposure measure separately and then jointly. In the OLS columns 1–4, PIO exposure alone yields a coefficient of 0.185, falling to 0.131 with controls, while KIO exposure alone yields 0.172. When both enter together, the KIO coefficient is 0.153 and the PIO coefficient 0.093, so the knowledge-network relationship is more robust to conditioning on the competing exposure.

[Table 3 here]

Columns 5–8 use the policy-induced variation. Instrumented separately, PIO and KIO exposure yield similar estimates of 2.107 and 2.000. Controlling for PIO exposure leaves the instrumented KIO estimate essentially unchanged at 1.974 (column 7). When both exposures are instrumented (column 8), the KIO coefficient remains positive and significant at 2.051, whereas the PIO coefficient falls to 0.741 and becomes insignificant. Because identification is weaker in this specification, with a Kleibergen–Paap rank F statistic of 8.19, we also report Anderson–Rubin confidence sets: the set for KIO excludes zero, whereas that for PIO includes it.

Why does PIO exposure produce a large estimate when instrumented on its own? Measurement error in industry input–output shares is an unlikely explanation: a production network rebuilt from pre-reform firm-to-firm transactions yields the same pattern (Appendix Table F1). Nor is instrument collinearity: the raw correlation of 0.975 between the two instruments largely reflects their common time structure and falls to 0.285 after partialling out the fixed effects and controls (Appendix Table F2).

The explanation lies in the instruments' cross-loadings (Appendix Table F3). The KIO instru-

ment predicts knowledge exposure but not production exposure (loading -0.004), whereas the PIO instrument predicts both, with loadings of 0.088 on production exposure and 0.058 on knowledge exposure. Instrumenting PIO alone therefore attributes to the production channel roughly two-thirds ($0.058/0.088$) of the knowledge-network effect: $0.741 + 0.67 \times 2.051 \approx 2.12$, close to the separate PIO estimate of 2.107. The large production-network coefficient thus reflects the knowledge exposure that production-linked shocks carry with them. Once the knowledge channel is accounted for, little independent production-network effect remains: FDI spillovers to Chinese innovation operate predominantly through the innovation network.

5.3 Additional Evidence on Identification

This subsection reports the evidence on the threats discussed in Section 4.2, examines whether the effect originates in upstream fields, and assesses the robustness of the identifying variation.

Non-FDI channels through upstream fields. Table 4 splits the instrument into the placebo component built from source economies without pre-reform FDI in China, IV^{KIO}_{Zero}, and the component built from the remaining economies, $IV^{KIO}_{Positive}$. In column 1, the reduced-form coefficient on IV^{KIO}_{Zero} is 0.035 and insignificant, whereas the coefficient on $IV^{KIO}_{Positive}$ is 0.109 and significant at the 1% level (column 2). Column 5 instruments $\ln KIOSPILL$ with $IV^{KIO}_{Positive}$ and includes IV^{KIO}_{Zero} as an exogenous regressor, so that the coefficient on the latter measures any direct effect on patenting beyond the embedded-technology channel. That coefficient is insignificant, while the effect of embedded technology remains positive and significant.[27]

[Table 4 here]

In the first stage (columns 3 and 4), IV^{KIO}_{Zero} does not raise embedded technology; its coefficient is negative and significant. Because these economies neither invested in China before the reform nor responded to it, and embedded technology is supplied almost entirely by baseline-FDI

[27] This comparison is meaningful only if IV^{KIO}_{Zero} contains independent variation. After partialling out the fixed effects and controls, the two components are correlated at 0.62, so about 61% of the residual variation in IV^{KIO}_{Zero} is independent of $IV^{KIO}_{Positive}$.

economies, the negative sign cannot reflect a causal effect through investment. It instead reflects a structural correlation: conditional on the fixed effects and controls, fields that draw more heavily on the knowledge of no-baseline-FDI economies receive less embedded technology from other economies.[28] This correlation does not affect our conclusions. It does not carry over into patenting, as columns 1 and 5 show, and these economies account for only 0.02% of the baseline instrument, so instrumenting with $IV^{KIO}_{Positive}$ alone reproduces the baseline estimate of 1.974.

Own-field effects. Columns 1 and 2 of Table 5 address direct effects of liberalization in the receiving field. Column 1 restricts the sample to the 338 fields whose own access status did not change and also removes the diagonal of the knowledge network, renormalizing the remaining weights, so that exposure originates entirely upstream. The estimate is 2.054. The restriction removes 41% of the sample and weakens the first stage, but the 95% Anderson–Rubin confidence set excludes zero. Column 2 retains the full sample and controls for the receiving field's own liberalization. The effect of FDI-embedded technology remains positive and significant at 1.702. The coefficient on own liberalization is also positive and significant (0.223), indicating that liberalization raises patenting in the receiving field directly, in addition to its effect through upstream knowledge spillovers.

[Table 5 here]

Upstream knowledge or same-field investment? Because FDI-embedded technology includes investment entering the receiving field itself, the estimated effect could reflect foreign entry into that field, through competition, capacity expansion, or direct technology transfer, rather than knowledge arriving from upstream fields. This possibility does not threaten the validity of the instrument, but it bears on the knowledge-network interpretation. Column 1 already excludes same-field exposure. Column 3 retains the full sample and splits exposure into a same-field component, the diagonal of the knowledge network, and a cross-field component. It instruments the

[28] The correlation between $\ln KIOSPILL$ and its baseline-FDI component is 1.000 to three decimal places. The first-stage coefficient on IV^{KIO}_{Zero} therefore records how exposure to the 2001 knowledge of no-baseline-FDI economies, propagated through the pre-reform network, covaries with the post-reform growth of embedded technology supplied by baseline-FDI economies once field and year fixed effects and controls are removed.

cross-field component with the corresponding cross-field instrument while conditioning on the same-field component. Cross-field exposure continues to predict patenting.

Robustness of the identifying variation. Appendix Table D1 varies the treatment definition, constructs an alternative shock weighted by pre-reform source-economy investment, limits the influence of extreme observations through winsorization and residual-based trimming, clusters standard errors at the coarser IPC3 level, and adds the later domestic innovation policies excluded from the baseline.[29] The estimate remains positive and significant throughout, although its magnitude is sensitive to the treatment of the tails. Appendix Figure D2 shows that every estimate remains positive and significant when one technology field or one source economy is left out at a time, and reform-induced exposure is broadly dispersed across fields rather than concentrated in a few.[30]

Finally, the binary treatment indicator discards information, because the 2002 revision moved industries by different numbers of steps along the four-category access ladder. Appendix Table C2 replaces it with a continuous measure of liberalization intensity (Appendix Figure C1). The first stages are stronger than in the baseline, with Kleibergen–Paap rank F statistics above 30, and the 2SLS estimates remain positive and significant, at 1.544 and 1.595.

5.4 Robustness Checks

Appendix D reports further checks on confounding investment motives and on the measurement choices underlying our key variables.

A first concern is that recorded FDI needs not carry foreign technology. A substantial share of China's inward investment is routed through offshore financial centers, and part of it reflects round-tripped domestic capital or tax planning rather than technology transfer. Because *KIOSPILL* weights investment by the source economy's knowledge stock, such flows misstate the origin and

[29] Adding source-economy $\times$ year fixed effects to the investment equation leaves the instrument numerically unchanged: the reform indicator and controls vary at the field $\times$ year level and the source-economy $\times$ field panel is balanced, so these fixed effects are orthogonal to the policy variation.

[30] Measuring each field's exposure change as the post-minus-pre change in the instrument, the top decile of the 576 fields accounts for 12.2% of the total change and the top quartile for 28.9%, close to the uniform benchmarks of 10% and 25%.

amount of the foreign knowledge that investment brings. Appendix Table D2 therefore splits embedded technology by whether the source economy is a tax haven, under five alternative haven lists, instrumenting the non-haven component and controlling for the haven component. The effect is carried by investment from non-haven economies, whereas the haven component is small and, if anything, negative.

A second concern is that the construction of *KIOSPILL* involves several choices that could drive the results. The industry–technology concordance is many-to-many, and dividing an industry's FDI equally among its fields is only one defensible allocation rule. Knowledge stocks depend on how quickly knowledge becomes obsolete and on how long citations are allowed to accumulate, and citation weighting may reflect citation practices as well as knowledge content. Appendix Table D3 assigns each industry's full FDI to every mapped field, raises the depreciation rate to 15% conventional for R&D stocks, extends the citation window to ten years, and builds knowledge stocks from patent counts rather than citations.

A third concern is that patent counts may measure quantity rather than quality. Chinese patenting expanded rapidly over the sample period, partly in response to policy incentives (Wei et al., 2017), so an increase in filings need not reflect more valuable inventions. Appendix Table D4 uses the inverse hyperbolic sine of patent counts, forward citations, which weight patents by their subsequent use, and patents granted outside China, which restrict attention to inventions whose owners bore the cost of protecting them abroad.

A final concern is that the mapping from industries to technology fields may misallocate FDI and hence exposure. Appendix Table D5 replaces the CNIPA concordance with a route through HS products using the probabilistic weights of Lybbert and Zolas (2014), reverses the direction of the mapping so that estimation takes place in industry space, and aggregates the analysis to the IPC3 level, which limits misallocation between neighboring fields.

Across these exercises, the estimated effect remains positive and statistically significant, although its magnitude varies with the construction, consistent with the caution in Section 5.1 about interpreting the size of the 2SLS coefficient.

6 Mechanisms

6.1 Knowledge Spillover Effects

If FDI-embedded technology raises innovation by transmitting foreign knowledge, domestic inventors should build more on earlier foreign inventions. Table 6 therefore replaces patenting with cross-border backward citations from Chinese-origin patent families to foreign patents in the same technology field (Jaffe et al., 1993; Jaffe and Trajtenberg, 1999). The citing side includes only Chinese-origin families filed at CNIPA, the cited side excludes families with a Chinese member, and citations are consolidated at the DOCDB-family level.[31]

[Table 6 here]

In column 1, both the IV and OLS estimates are positive and significant. Because citation counts contain more zeros and are more skewed than patent counts, we also report Poisson, winsorized Poisson, and extensive-margin estimates. The raw-count Poisson estimate is close to zero, whereas the winsorized estimate is positive and significant, indicating that the Poisson estimate is sensitive to a small number of citation-intensive field-years. The extensive-margin estimate is insignificant, as roughly three-quarters of field-years already contain cross-border citations.

Columns 2–4 examine whether the pattern reflects a change in who patents rather than learning by domestic inventors. Matching the FDI registry to patent applicants identifies 269,077 patent families affiliated with FIEs. Removing these families from the citing side, the cited side, or both changes the estimates little, so the citation response is not driven by foreign-affiliated patenting in China.[32]

We also consider whether liberalization simply accelerated China's access to global knowledge. Following Hegde et al. (2023), Appendix Table G2 measures diffusion speed by the lag between

[31] Country of origin is assigned by inventor residence, then by applicant country when inventor information is unavailable, and finally by the country of the patent office.

[32] Appendix Table G1 reports a more demanding test at the source-economy $\times$ field level: Chinese inventors disproportionately cite patents from the specific source economies whose technology reaches them through upstream fields.

a foreign patent filing and its first citation by a Chinese patent. We find no evidence that greater exposure shortens this lag, and controlling for citation lag leaves the effect on foreign citations essentially unchanged, consistent with evidence that the speed of cross-border knowledge flows has changed little in recent decades (Cotterlaz and Guillouzouic, 2025). The citation response thus reflects greater use of foreign knowledge rather than faster diffusion in general.

6.2 How Do Spillovers Occur?

Foreign knowledge can reach domestic firms through market transactions in existing intellectual property, such as patent assignments and licenses (Arora et al., 2001; Akcigit et al., 2016), or through interaction between foreign and domestic inventors (Singh, 2005). The distinction matters because an effect operating entirely through assignments or licenses would represent market-based technology transfer rather than a knowledge spillover (Keller, 2004).

Using the transaction records described in Section 3.1, we construct field-year counts of patents assigned or licensed by FIEs to domestic entities and of patents jointly applied for by an FIE and a domestic partner. Table 7 shows that collaboration responds most consistently: FDI-embedded technology increases joint patenting in the IV, OLS, and extensive-margin specifications, and the effect changes little after controlling for assignments and licenses. Assignments and licenses respond only in some specifications, and controlling for both leaves the baseline effect on domestic patenting economically and statistically intact, so market-based transfer cannot account for the innovation response.[33] Interaction between foreign and domestic inventors thus emerges as an important observable channel of transmission.

[Table 7 here]

[33] Because assignments and licenses are themselves outcomes of foreign entry, columns 4 and 5 provide a descriptive rather than causal decomposition. Coverage of transaction records also improves over the sample period; year fixed effects absorb improvements common across fields, although not differential improvements correlated with treatment.

6.3 Collaboration or Compliance?

Joint patenting may instead reflect a *quid pro quo*: technology in exchange for market access. Such requirements were sometimes imposed in China (Holmes et al., 2015; Bai et al., 2020), particularly on joint ventures, so a *quid pro quo* account predicts a stronger collaboration response to exposure carried by joint ventures than to exposure carried by wholly foreign-owned enterprises (WFOEs). Appendix Table G3 decomposes FDI-embedded technology by the ownership form of the investing FIE.[34] The evidence points the other way. The WFOE component yields the larger IV coefficient whether the two components enter separately (4.423 versus 1.913) or jointly (2.483 versus 1.680), and the same ordering holds in the OLS and extensive-margin specifications.

A second test uses the geographic scope of patent protection. Joint patents filed merely to satisfy a domestic transfer requirement have little reason to be protected abroad, whereas seeking costly foreign protection signals that the parties expected the invention to have value beyond the domestic regulatory setting. Of the 44,605 patent families underlying the collaboration measure, 27.2% include a member filed at a patent office other than CNIPA, and 25.5% filed at the U.S., European, or Japanese office.[35] When the collaboration outcome is restricted to these families, the IV and extensive-margin estimates remain positive and significant (Appendix Table G4, columns 1 and 2). Restricting it instead to families filed in the home jurisdiction of the FIE's foreign investors yields the same pattern (column 3).[36]

Taken together, the results are difficult to reconcile with a strong version of the mandated-transfer explanation, although they do not rule out that some collaboration within joint ventures is regulatory in origin.

[34] A firm with no domestic shareholder is classified as wholly foreign owned, and one with a domestic shareholder as a joint venture.

[35] We observe applications within the patent family rather than foreign grants, so this exercise establishes that protection was sought abroad, not that it was granted.

[36] For investors from Hong Kong, Macao, and Taiwan, whose principal market is the mainland, the home jurisdiction is taken to be CNIPA, so column 3 measures protection in the investor's principal market rather than protection abroad.

6.4 Heterogeneity

If FDI raises domestic innovation by supplying knowledge that domestic inventors lack, the gains should be larger where there is more to learn and where the receiving field is better positioned to absorb and propagate that knowledge. We examine these two dimensions using the technology gap and network centrality.

Theory predicts that the return to foreign knowledge rises with the gap to be closed (Findlay, 1978). We measure this gap in two ways. The first is China's distance from the global frontier in the field, the pre-sample log difference between rest-of-world and Chinese patenting. The second is specific to the investors. Because the knowledge arriving through FDI depends on who invests, we compute each source economy's technological lead over China in the field and weight these leads by pre-reform investment shares.[37] Both measures are built primarily from Patent Cooperation Treaty (PCT) applications, because counts pooled across offices confound inventive activity with differences in filing propensity and office coverage; the all-office versions are reported alongside.

Network position matters because a central field passes incoming knowledge on to more of the network. We measure it by eigenvector centrality in the citation matrix, following Liu and Ma (2021), computed over the full sample and over the pre-reform window used for the instrument.

Table 8 shows that the interaction with FDI-embedded technology is positive and significant for all six measures: spillovers are larger in fields farther from the global frontier, in fields whose investors hold a larger lead over China, and in fields more central to the knowledge network.[38] Foreign technology thus has larger effects where there is more scope for learning and where the knowledge network can propagate it further, consistent with the knowledge-spillover interpretation.

[Table 8 here]

[37] The field-level gap is $\ln(1+patnum_{ROW,i}) - \ln(1+patnum_{CN,i})$, standardized across fields. For the source lead, we compute $techgap_{ci} = \sum_{k \in i}(patnum_{ck} - patnum_{CN,k})/n_i$ over the n_i IPC main groups k in field i, and aggregate $gapsrc_i = \sum_c w_{ci} techgap_{ci}$ using source economies' shares of cumulative 1995–2001 FDI into the field. Because $techgap_{ci}$ is strongly right-skewed, $gapsrc_i$ enters as a standardized percentile rank. The pre-sample window is 1985–1995 for all-office counts and 1994–2001 for PCT applications, since China joined the PCT in 1994.

[38] Adding the interaction weakens the first stage, so we report Anderson–Rubin confidence sets for the interaction, which exclude zero in all six specifications.

7 Conclusion

Knowledge that arrives with foreign investment need not travel along the same routes as goods. Using a knowledge input–output table constructed from inter-field patent citations and policy-induced variation in FDI from China's 2002 liberalization of foreign investment access, we find that greater FDI-embedded intermediate technology raises domestic patenting. A production input–output analogue, built from the same investment variation and foreign knowledge stocks and differing only in its network weights, yields a large estimate when used on its own, but this estimate is largely accounted for by its correlation with knowledge exposure. Once knowledge exposure is taken into account, little independent production-network effect remains.

The mechanism evidence is consistent with domestic inventors learning from the foreign technology that arrives. FDI-embedded technology increases their cross-border backward citations, and this relationship is not explained by foreign-affiliated patenting in China or by a general acceleration in access to global knowledge. Among observable channels, collaboration between foreign affiliates and domestic entities provides the most consistent response, while market-based transfers and licenses do not account for the patenting effect. This collaboration response is inconsistent with a simple *quid pro quo* explanation: it is stronger for wholly foreign-owned than joint-venture exposure and remains positive for collaborations whose inventions were subsequently filed for protection abroad. Finally, spillovers are larger where the technological gap is greater and where the receiving field is more central in the knowledge network.

These findings bear on how the innovation effects of FDI are measured and evaluated. The benefits of opening a sector to foreign investment need not remain within that sector or follow its supply chain; they can propagate through technological relationships that connect otherwise distant fields. Measuring spillovers along production linkages alone may therefore misattribute where and how foreign technology affects domestic innovation. For host-country policy, the returns to liberalizing a sector depend not only on its production linkages but also on its position in the knowledge network.

Several limitations point to directions for future research. Our estimates are local to the invest-

ment margin shifted by a single liberalization and summarize a cumulative medium-run response, and the magnitude of the 2SLS coefficient should be interpreted with caution. Patent data capture codified knowledge but do not directly observe other channels, such as worker mobility, imitation, or training. Finally, we trace spillovers across technology fields but not across geographic space; integrating technological and geographic networks would provide a fuller account of how foreign knowledge diffuses within the host economy.

References

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

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

Akcigit, Ufuk, Murat Alp Celik, and Jeremy Greenwood. 2016. "Buy, Keep, or Sell: Economic Growth and the Market for Ideas." *Econometrica*, 84(3): 943–984.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P Vasquez. 2022. "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages." *The Quarterly Journal of Economics*, 137(3): 1495–1552.

Amiti, Mary, Cédric Duprez, Jozef Konings, and John Van Reenen. 2024. "FDI and Superstar Spillovers: Evidence from Firm-To-Firm Transactions." *Journal of International Economics*, 152: 103972.

Arora, Ashish, Andrea Fosfuri, and Alfonso Gambardella. 2001. *Markets for Technology: The Economics of Innovation and Corporate Strategy.* The MIT Press.

Ayerst, Stephen, Faisal Ibrahim, Gaelan MacKenzie, and Swapnika Rachapalli. 2023. "Trade and Diffusion of Embodied Technology: An Empirical Analysis." *Journal of Monetary Economics*, 137: 128–145.

Bai, Jie, Panle Jia Barwick, Shengmao Cao, and Shanjun Li. 2020. "Quid Pro Quo, Knowledge Spillover, and Industrial Quality Upgrading: Evidence from the Chinese Auto Industry." National Bureau of Economic Research Working Paper 27644.

Barrios, Salvador, Holger Görg, and Eric Strobl. 2011. "Spillovers Through Backward Linkages From Multinationals: Measurement Matters!" *European Economic Review*, 55(6): 862–875.

Bennedsen, Morten, and Stefan Zeume. 2018. "Corporate Tax Havens and Transparency." *The Review of Financial Studies*, 31(4): 1221–1264.

Blomström, Magnus, and Ari Kokko. 1998. "Multinational Corporations and Spillovers." *Journal of Economic Surveys*, 12(3): 247–277.

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

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.

Brambilla, Irene, Amit K. Khandelwal, and Peter K. Schott. 2010. "China's Experience under the Multi-Fiber Arrangement (MFA) and the Agreement on Textiles and Clothing (ATC)." National Bureau of Economic Research Working Paper 13346.

Brandt, Loren, Johannes Van Biesebroeck, Luhang Wang, and Yifan Zhang. 2017. "WTO Accession and Performance of Chinese Manufacturing Firms." *American Economic Review*, 107(9): 2784–2820.

Branstetter, Lee G. 2001. "Are Knowledge Spillovers International or Intranational in Scope?: Microeconometric Evidence From the U.S. and Japan." *Journal of International Economics*, 53(1): 53–79.

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

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.

Chen, Jiafeng, and Jonathan Roth. 2024. "Logs with Zeros? Some Problems and Solutions." *The Quarterly Journal of Economics*, 139(2): 891–936.

Che, Yi, and Lei Zhang. 2018. "Human Capital, Technology Adoption and Firm Performance: Impacts of China's Higher Education Expansion in the Late 1990s." *The Economic Journal*, 128(614): 2282–2320.

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

Cohen, Lauren, and Andrea Frazzini. 2008. "Economic Links and Predictable Returns." *The Journal of Finance*, 63(4): 1977–2011.

Cotterlaz, Pierre, and Arthur Guillouzouic. 2025. "The Percolation of Knowledge across Space." *Journal of International Economics*, 153: 104026.

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

Fons-Rosen, Christian, Sebnem Kalemli-Ozcan, Bent E. Sorensen, Carolina Villegas-Sanchez, and Vadym Volosovych. 2017. "Foreign Investment and Domestic Productivity: Identifying Knowledge Spillovers and Competition Effects." National Bureau of Economic Research Working Paper 23643.

Griliches, Zvi. 1979. "Issues in Assessing the Contribution of Research and Development to Productivity Growth." *Bell Journal of Economics*, 10(1): 92–116.

Grossman, Gene M., and Elhanan Helpman. 1991. "Quality Ladders and Product Cycles." *The Quarterly Journal of Economics*, 106(2): 557–586.

Handley, Kyle, and Nuno Limão. 2017. "Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States." *American Economic Review*, 107(9): 2731–2783.

Harrison, Ann, and Andrés Rodríguez-Clare. 2010. "Trade, Foreign Investment, and Industrial Policy for Developing Countries." In *Handbook of Development Economics*. Vol. 5 of *Handbooks in Economics*, ed. Dani Rodrik and Mark Rosenzweig, 4039–4214. Elsevier.

Havranek, Tomas, and Zuzana Irsova. 2011. "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics*, 85(2): 234–244.

Hegde, Deepak, Kyle Herkenhoff, and Chenqi Zhu. 2023. "Patent Publication and Innovation." *Journal of Political Economy*, 131(7): 1845–1903.

Hines, Jr., James R., and Eric M. Rice. 1994. "Fiscal Paradise: Foreign Tax Havens and American Business." *The Quarterly Journal of Economics*, 109(1): 149–182.

Holmes, Thomas J., Ellen R. McGrattan, and Edward C. Prescott. 2015. "Quid Pro Quo: Technology Capital Transfers for Market Access in China." *The Review of Economic Studies*, 82(3): 1154–1193.

Hsieh, Chang-Tai, and Zheng Michael Song. 2015. "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China." *Brookings Papers on Economic Activity*, 46(1): 295–366.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475.

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

Jaffe, Adam B., and Manuel Trajtenberg. 1999. "International Knowledge Flows: Evidence from Patent Citations." *Economics of Innovation and New Technology*, 8(1–2): 105–136.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics*, 108(3): 577–598.

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

Jiang, Kun, Wolfgang Keller, Larry D. Qiu, and William Ridley. 2024. "International Joint Ventures and Internal Technology Transfer vs. External Technology Spillovers: Evidence From China." *Journal of International Economics*, 150: 103939.

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

Keller, Wolfgang. 2021. "Knowledge Spillovers, Trade, and FDI." National Bureau of Economic Research Working Paper 28739.

Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei. 2013. "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters." *American Economic Review*, 103(6): 2169–2195.

Liu, Ernest, and Song Ma. 2021. "Innovation Networks and Innovation Policy." National Bureau of Economic Research Working Paper 29607.

Liu, Qing, and Larry D. Qiu. 2016. "Intermediate Input Imports and Innovations: Evidence from Chinese Firms' Patent Filings." *Journal of International Economics*, 103: 166–183.

Liu, Qing, Larry D. Qiu, and Chaoqun Zhan. 2024. "FDI Inflows and Export Quality: Domestic Competition and Within-Firm Adjustment." *Journal of Development Economics*, 170: 103293.

Lu, Yi, Zhigang Tao, and Lianming Zhu. 2017. "Identifying FDI Spillovers." *Journal of International Economics*, 107: 75–90.

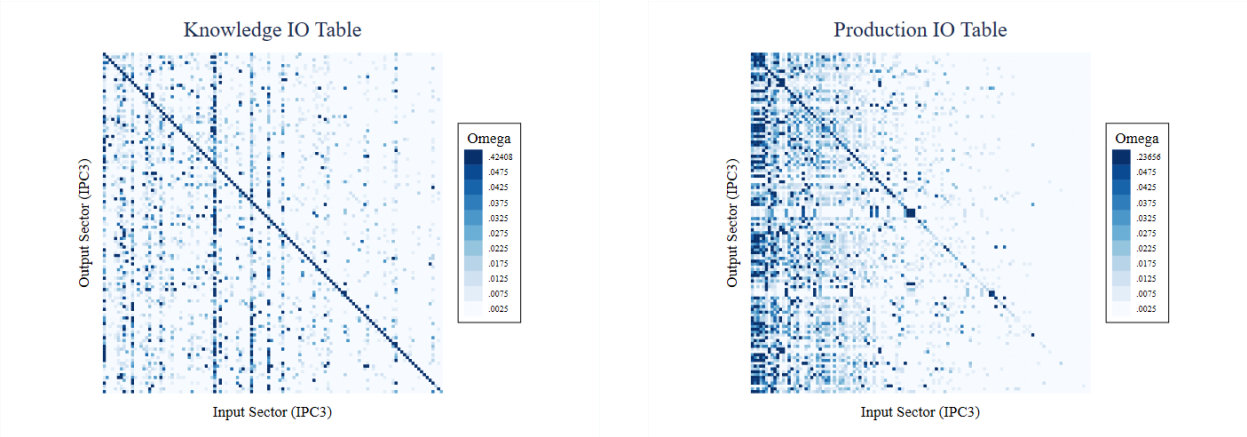
Lybbert, Travis J., and Nikolas J. Zolas. 2014. "Getting Patents and Economic Data to Speak to Each Other: An 'Algorithmic Links with Probabilities' Approach for Joint Analyses of Patenting and Economic Activity." *Research Policy*, 43(3): 530–542.

Ma, Xiao, and Yiran Zhang. 2025. "How Do Multinationals Impact China's Technology? The Role of Quid Pro Quo Policy and Technology Spillovers." *International Economic Review*, 66(5): 1933–1956.

Pierce, Justin R., and Peter K. Schott. 2016. "The Surprisingly Swift Decline of US Manufacturing Employment." *American Economic Review*, 106(7): 1632–1662.

Singh, Jasjit. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science*, 51(5): 756–770.

Wei, Shang-Jin, Zhuan Xie, and Xiaobo Zhang. 2017. "From 'Made in China' to 'Innovated in China': Necessity, Prospect, and Challenges." *Journal of Economic Perspectives*, 31(1): 49–70.

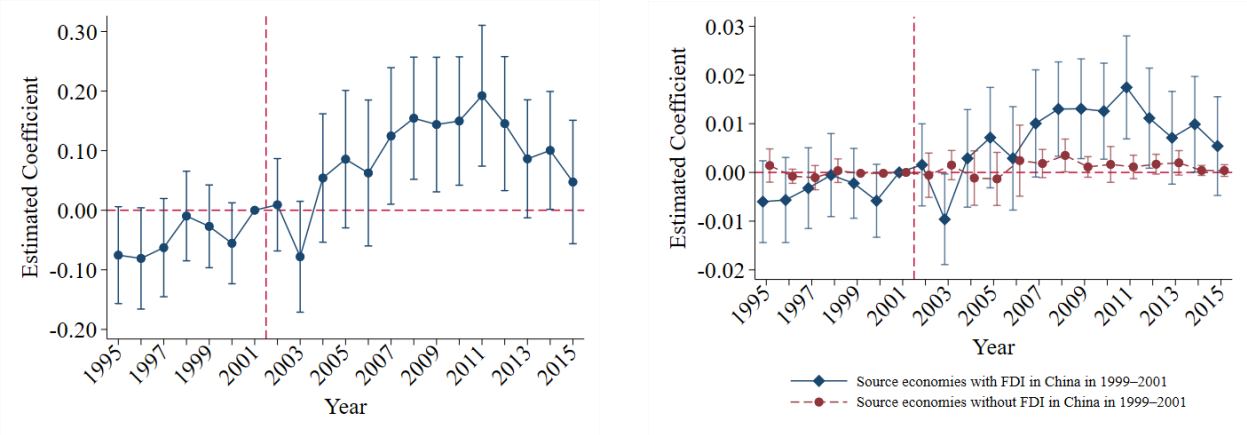


(a) Innovation network

(b) Production network

Figure 1: Innovation and Production Networks at the IPC3 Level

Notes: Heat maps of the two input–output matrices at the three-digit IPC level. Cell (i, j) is the weight of upstream field j in the inputs of using field i, so that rows sum to one; darker cells indicate larger weights. Panel (a) shows knowledge input shares ω_{ij}, the share of field i's backward citations directed to field j, computed from citations on U.S. patents over 1996–2016. Panel (b) shows production input shares ν_{ij} from the 1997 benchmark use table of the U.S. Bureau of Economic Analysis, defined over total intermediate inputs and mapped to technology fields using the concordance of Lybbert and Zolas (2014). Both panels cover the 120 three-digit fields for which the two networks are defined; the Pearson correlation between ω_{ij} and ν_{ij} across the 14,400 ordered field pairs is 0.32. Appendix Figure B1 repeats the comparison at the IPC4 level used in estimation.



(a) All source economies, $\ln(FDI+1)$

(b) By baseline FDI, extensive margin

Figure 2: Event-Study Estimates of the 2002 *Catalogue* Reform on FDI

Notes: Panel (a) reports the event-study counterpart of equation (4), estimated at the source-economy × technology-field × year level with $\ln(FDI_{cjt}+1)$ as the dependent variable. Panel (b) reports the extensive-margin counterpart, with an indicator for positive FDI as the dependent variable, estimated separately for source economies with and without FDI in China over 1999–2001. Both panels interact the *Catalogue* liberalization indicator with year dummies and include source-economy × technology-field fixed effects, year fixed effects, and the controls of Table 1. The omitted year is 2001. Markers denote point estimates and vertical bars 95% confidence intervals; the vertical dashed line marks the April 2002 reform. Standard errors are two-way clustered by source economy and technology field. Appendix Figure C3 decomposes the investment response into extensive and intensive margins.

Table 1: FDI Inflows Predicted by the 2002 *Catalogue* Reform

	(1) $\ln(FDI+1)$	(2) $\ln(FDI+1)$
Treat × *Post*	0.141***	0.189***
	(0.043)	(0.044)
Controls	Yes	Yes
1995 restricted category × year	Yes	Yes
Source economy × field FE	Yes	Yes
Year FE	Yes	Yes
R^2	0.642	0.642
Observations	2,467,584	2,467,584

Notes: The dependent variable is $\ln(FDI_{cjt}+1)$, foreign investment from source economy c into technology field j in year t. *Treat* indicates fields liberalized by the 2002 revision of the *Catalogue for the Guidance of Foreign Investment Industries*: fields whose share of liberalized industries exceeds the median in column (1), the baseline definition, and fields with any positive share in column (2). *Post* equals one from 2003 onward and $3/4$ in 2002, reflecting the April implementation of the revised *Catalogue*. Controls are 1998 field characteristics, 2001 input-, output-, and partner-weighted tariffs, the 2001 SOE share, and the 1995 restricted category, each interacted with year dummies; the PNTR and TRIPS exposures, each interacted with an indicator for years from 2001 onward; and the textile and clothing quota fill rate, which falls to zero as quotas are removed. All columns include source-economy × technology-field and year fixed effects. Standard errors, in parentheses, are two-way clustered by source economy and technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table 2: FDI-Embedded Technology and China's Innovation Output

	(1) OLS $\ln patnum$	(2) OLS $\ln patnum$	(3) First stage $\ln KIOSPILL$	(4) 2SLS $\ln patnum$	(5) Reduced form $\ln patnum$	(6) PPML $patnum$
$\ln KIOSPILL$	0.173***	0.172***		2.000***		0.066**
	(0.035)	(0.035)		(0.756)		(0.027)
$\ln IV^{KIO}$			0.054***		0.108***	
			(0.015)		(0.033)	
Controls	No	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.911	0.923	0.941		0.922	
Observations	12,096	12,096	12,096	12,096	12,096	12,096
KP-rk LM statistic				8.772		
KP-rk F statistic				12.460		

Notes: The dependent variable is $\ln(1 + patnum)$, the log of one plus the number of granted Chinese patent families in the technology field, dated by application year, except in column (3), where it is $\ln KIOSPILL$, and column (6), where it is the patent count. $\ln KIOSPILL$ is the log of FDI-embedded intermediate technology in equation (1), and $\ln IV^{KIO}$ the log of the knowledge-network instrument in equation (5); both are lagged one year. Column (4) instruments $\ln KIOSPILL$ with $\ln IV^{KIO}$, column (5) reports the reduced form, and column (6) estimates a Poisson pseudo-maximum-likelihood model. The baseline controls, included from column (2) onward, are log patent applications outside China and log imports in the field; 1998 field characteristics, 2001 input-, output-, and partner-weighted tariffs, and the 2001 SOE share, each interacted with year dummies; the PNTR and TRIPS exposures, each interacted with an indicator for years from 2001 onward; and the textile and clothing quota fill rate, which falls to zero as quotas are removed. All columns include technology-field and year fixed effects. KP-rk LM and F denote the Kleibergen–Paap rank LM statistic for underidentification and Wald F statistic for weak identification. The sample is a balanced panel of 576 IPC4 fields over 1996–2016. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table 3: Knowledge versus Production Networks: Horse Race

	(1) OLS ln *patnum*	(2) OLS ln *patnum*	(3) OLS ln *patnum*	(4) OLS ln *patnum*	(5) 2SLS ln *patnum*	(6) 2SLS ln *patnum*	(7) 2SLS ln *patnum*	(8) 2SLS ln *patnum*
ln *KIOSPILL*			0.172***	0.153***		2.000***	1.974***	2.051***
			(0.035)	(0.031)		(0.756)	(0.733)	(0.786)
ln *PIOSPILL*	0.185***	0.131***		0.093**	2.107***		−0.364*	0.741
	(0.048)	(0.048)		(0.044)	(0.673)		(0.203)	(0.736)
Instrumented	none	none	none	none	*PIOSPILL*	*KIOSPILL*	*KIOSPILL*	both
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,096	12,096	12,096	12,096	12,096	12,096	12,096	12,096
KP-rk LM statistic					18.828	8.772	10.524	11.446
KP-rk F statistic					23.423	12.460	15.124	8.186
AR 95% set, ln *KIOSPILL*						[0.79, 4.80]	[0.80, 4.39]	[0.58, 5.74]
AR 95% set, ln *PIOSPILL*					[0.96, 3.99]			[−1.56, 3.04]

Notes: The dependent variable is $\ln(1 + patnum)$, defined as in Table 2. ln *KIOSPILL* and ln *PIOSPILL* aggregate the same FDI-embedded foreign knowledge through knowledge input shares and production input shares, respectively (equations (1) and (3)); both are lagged one year. Columns (1)–(4) are OLS estimates with production exposure alone, without and with controls, knowledge exposure alone, and both together. Columns (5)–(8) are 2SLS estimates that instrument production exposure alone, knowledge exposure alone, knowledge exposure with production exposure as an exogenous control, and both exposures. The production-network instrument in equation (6) uses the same policy-induced investment shock and 2001 knowledge stocks as the knowledge-network instrument, replacing citation shares with production input shares. Anderson–Rubin (AR) confidence sets are valid regardless of instrument strength; in column (8), each set is the projection of the joint set onto the named coefficient and is therefore conservative. With two endogenous regressors and two instruments the relevant Stock–Yogo critical value is 7.03 at the 10% level. All columns include technology-field and year fixed effects, and columns (2)–(8) the baseline controls of Table 2. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table 4: Placebo Test: Source Economies without Pre-Reform Investment

	(1) Reduced form $\ln patnum$	(2) Reduced form $\ln patnum$	(3) First stage $\ln KIOSPILL$	(4) First stage $\ln KIOSPILL$	(5) 2SLS $\ln patnum$	(6) 2SLS $\ln patnum$
$\ln KIOSPILL$					1.154***	0.909**
					(0.359)	(0.418)
$\ln IV^{KIO}_{Zero}$	0.035		−0.118***		0.172	
	(0.135)		(0.030)		(0.119)	
$\ln IV^{KIO}_{Positive}$		0.109***		0.055***		
		(0.032)		(0.014)		
Instruments					Positive	Both
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.922	0.923	0.944	0.944		
Observations	12,096	12,096	12,096	12,096	12,096	12,096
KP-rk LM statistic					32.312	42.849
KP-rk F statistic					68.381	44.036

Notes: The table splits the knowledge-network instrument by source economies' FDI in China over 1999–2001: IV^{KIO}_{Zero} sums over the 41 source economies with no FDI in China in any year of this period, and $IV^{KIO}_{Positive}$ over the remaining economies. Each component is constructed as in equation (5) and lagged one year. The dependent variable is $\ln(1+patnum)$, defined as in Table 2, except in columns (3) and (4), where it is $\ln KIOSPILL$. Columns (1) and (2) report reduced forms and columns (3) and (4) first stages. Column (5) instruments $\ln KIOSPILL$ with $\ln IV^{KIO}_{Positive}$ and includes $\ln IV^{KIO}_{Zero}$ as an exogenous regressor, whose coefficient measures any direct effect on patenting beyond embedded technology. Column (6) uses both components as instruments. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table 5: Separating Upstream Knowledge Spillovers from Own-Field Effects

	(1) Upstream-only exposure ln *patnum*	(2) Own liberalization controlled ln *patnum*	(3) Same-field FDI controlled ln *patnum*
Panel A. 2SLS			
ln *KIOSPILL*	2.054**	1.702**	1.055*
	(1.010)	(0.689)	(0.604)
ln *KIOSPILL*, same field			0.006
			(0.004)
Treat × *Post*		0.223***	
		(0.083)	
Observations	7,098	12,096	12,096
KP-rk LM statistic	4.140	10.244	11.579
KP-rk F statistic	7.216	14.807	17.507
AR 95% confidence set	[0.64, 8.62]		
Panel B. OLS			
ln *KIOSPILL*	0.133***	0.151***	0.155***
	(0.037)	(0.031)	(0.032)
ln *KIOSPILL*, same field			0.008**
			(0.004)
Treat × *Post*		0.251***	
		(0.075)	
R^2	0.919	0.923	0.923
Observations	7,098	12,096	12,096

Notes: The dependent variable is $\ln(1 + patnum)$, defined as in Table 2. Column (1) restricts the sample to the 338 fields whose own *Catalogue* status did not change, removes the diagonal of the knowledge network, and renormalizes the remaining weights, so that exposure originates entirely in upstream fields. Column (2) adds the receiving field's own liberalization, *Treat* × *Post*. Column (3) splits exposure into a cross-field component, which omits the diagonal of the knowledge network, and a same-field component, the diagonal alone; the cross-field component is instrumented by the corresponding cross-field instrument, and the same-field component enters as a control. For column (1), whose first stage is weak, the table reports the 95% Anderson–Rubin confidence set; the Anderson–Rubin p-value at zero is 0.003. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table 6: Mechanism: Cross-Border Backward Citations

	(1) Baseline *backcite*	(2) Excluding FIE citing families *backcite*	(3) Excluding FIE cited families *backcite*	(4) Excluding both *backcite*
Panel A. 2SLS				
$\ln KIOSPILL$	4.349***	3.887***	4.346***	3.884***
	(1.318)	(1.232)	(1.317)	(1.231)
Observations	12,096	12,096	12,096	12,096
KP-rk LM statistic	10.524	10.524	10.524	10.524
KP-rk F statistic	15.124	15.124	15.124	15.124
Panel B. Log OLS				
$\ln KIOSPILL$	0.170***	0.156***	0.170***	0.156***
	(0.041)	(0.040)	(0.041)	(0.040)
R^2	0.866	0.865	0.866	0.865
Observations	12,096	12,096	12,096	12,096
Panel C. PPML, raw counts				
$\ln KIOSPILL$	−0.010	−0.021	−0.010	−0.021
	(0.031)	(0.032)	(0.031)	(0.032)
Observations	12,033	12,033	12,033	12,033
Panel D. PPML, counts winsorized at p99				
$\ln KIOSPILL$	0.078**	0.063**	0.078**	0.063**
	(0.034)	(0.031)	(0.034)	(0.031)
Observations	12,033	12,033	12,033	12,033
Panel E. Extensive margin				
$\ln KIOSPILL$	0.014	0.014	0.014	0.014
	(0.012)	(0.012)	(0.012)	(0.012)
R^2	0.509	0.511	0.509	0.511
Observations	12,096	12,096	12,096	12,096

Notes: The dependent variable is the number of cross-border backward citations from Chinese-origin patent families filed at CNIPA to foreign families with no Chinese member, entering as $\ln(1+\cdot)$ in Panels A and B, as a raw count in Panel C, as a count winsorized at the 99th percentile in Panel D, and as an indicator for a positive count in Panel E. Column (1) uses all citation links. Columns (2), (3), and (4) exclude links involving patent families affiliated with registered FIEs on the citing side, the cited side, and both sides, respectively. $\ln KIOSPILL$ is defined as in Table 2, instrumented in Panel A and entered directly elsewhere. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table 7: Channels of Knowledge Spillover

	(1) Transfers *transfer*	(2) Licenses *license*	(3) Collaborations *collaborate*	(4) Collaborations *collaborate*	(5) Patents *patnum*
Panel A. 2SLS					
$\ln KIOSPILL$	2.139***	0.443***	5.102***	4.087***	1.582**
	(0.671)	(0.161)	(1.542)	(1.238)	(0.677)
Observations	12,096	12,096	12,096	12,096	12,096
KP-rk LM statistic	10.524	10.524	10.524	10.426	10.426
KP-rk F statistic	15.124	15.124	15.124	14.986	14.986
Panel B. Log OLS					
$\ln KIOSPILL$	0.022	−0.007	0.075***	0.068***	0.149***
	(0.017)	(0.009)	(0.025)	(0.018)	(0.030)
R^2	0.443	0.231	0.649	0.699	0.924
Observations	12,096	12,096	12,096	12,096	12,096
Panel C. PPML, raw counts					
$\ln KIOSPILL$	0.200	−0.114	−0.094	−0.089	0.044
	(0.139)	(0.199)	(0.093)	(0.085)	(0.028)
Observations	6,560	2,115	9,954	9,954	12,096
Panel D. PPML, counts winsorized at p99					
$\ln KIOSPILL$	0.077	0.001	−0.013	−0.031	0.076***
	(0.097)	(0.192)	(0.068)	(0.068)	(0.024)
Observations	6,560	2,115	9,954	9,954	12,096
Panel E. Extensive margin					
$\ln KIOSPILL$	0.020**	−0.005	0.037***	0.036***	0.002
	(0.009)	(0.006)	(0.010)	(0.010)	(0.009)
R^2	0.402	0.217	0.519	0.524	0.439
Observations	12,096	12,096	12,096	12,096	12,096

Notes: The dependent variables are patent assignments from FIEs to domestic entities in column (1), licenses in column (2), patents co-applied by an FIE and a domestic entity in columns (3) and (4), and *patnum* in column (5). They enter as $\ln(1+\cdot)$ in Panels A and B, as raw counts in Panel C, as counts winsorized at the 99th percentile in Panel D, and as an indicator for a positive count in Panel E. $\ln KIOSPILL$ is defined as in Table 2, instrumented in Panel A and entered directly elsewhere. Columns (4) and (5) additionally control for assignments and licenses. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table 8: Heterogeneity Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	H: China–ROW gap		H: Source-lead gap		H: ln *centrality*	
	PCT	All offices	PCT	All offices	1996–2016	1992–2001
	ln *patnum*	ln *patnum*	ln *patnum*	ln *patnum*	ln *patnum*	ln *patnum*
Panel A. 2SLS						
ln *KIOSPILL*	−0.195	1.923***	0.110	0.413	0.492	0.523
	(0.485)	(0.661)	(0.550)	(0.622)	(0.606)	(0.633)
ln *KIOSPILL* × H	0.343***	0.308***	0.191***	0.271***	0.156***	0.163***
	(0.020)	(0.060)	(0.026)	(0.035)	(0.013)	(0.013)
Observations	12,096	12,096	11,865	11,865	12,096	12,096
KP-rk LM statistic	11.328	10.534	10.881	8.164	10.484	9.367
KP-rk F statistic	8.350	7.356	8.424	5.692	7.893	6.655
AR 95% confidence set	[0.29, 0.39]	[0.12, 0.46]	[0.13, 0.26]	[0.18, 0.37]	[0.12, 0.20]	[0.14, 0.22]
Panel B. OLS						
ln *KIOSPILL*	0.165***	0.128***	0.156***	0.148***	0.936***	0.930***
	(0.030)	(0.030)	(0.030)	(0.031)	(0.063)	(0.062)
ln *KIOSPILL* × H	0.210***	0.191***	0.127***	0.164***	0.093***	0.093***
	(0.011)	(0.016)	(0.015)	(0.014)	(0.006)	(0.006)
R^2	0.931	0.928	0.927	0.929	0.930	0.931
Observations	12,096	12,096	11,865	11,865	12,096	12,096

Notes: The dependent variable is $\ln(1 + patnum)$, defined as in Table 2. H is a pre-determined, time-invariant field characteristic. In columns (1) and (2), H is the China–rest-of-world technology gap, the standardized log difference between rest-of-world and Chinese filings in the field over 1985–1995 (1994–2001 for PCT filings, as China joined the Patent Cooperation Treaty in 1994). In columns (3) and (4), H is the source lead, the pre-2002 FDI-share-weighted technological lead of China's source economies over China in the field, converted to standardized percentile ranks because the raw measure is strongly right-skewed. In each of these pairs, the first column uses PCT families, which are comparable across offices, and the second uses counts pooled across all offices. In columns (5) and (6), H is log eigenvector centrality in the knowledge network, computed over the full sample and over the pre-reform window used for the instrument. In Panel A, ln *KIOSPILL* and its interaction are instrumented by the baseline instrument and its interaction with H; Anderson–Rubin confidence sets for the interaction, valid regardless of instrument strength, are projections of the joint set and therefore conservative. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Online Appendix for "Identifying FDI Knowledge Spillovers Using Innovation Networks: Evidence from China"

(Not for publication)

Li Jiang, Canlei Liu, Yue Lv, Jing Zhang, Laixun Zhao*

November 2026

Contents

Appendix A	Data Construction	A-2
Appendix B	Innovation Network	A-19
Appendix C	The 2002 *Catalogue* Reform	A-24
Appendix D	Robustness of the Baseline	A-28
Appendix E	The OLS–IV Gap	A-35
Appendix F	Why Production-Network Spillovers Do Not Survive	A-38
Appendix G	Additional Results	A-43

*Contact authors: Li Jiang, School of Economics, Renmin University of China, Beijing 100872, China, email: jiangli0029@163.com; Canlei Liu, School of Economics, University of International Business and Economics, Beijing 100029, China, email: canlei_liu@163.com; Yue Lv, School of International Trade and Economics, University of International Business and Economics, Beijing 100029, China, email: lvyue@uibe.edu.cn; Jing Zhang, School of Economics and Statistics, Guangzhou University, Guangzhou 510405, China, email: zh_jing19@163.com; Laixun Zhao, Research Institute for Economics & Business, Kobe University, 2-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan, email: zhao@rieb.kobe-u.ac.jp.

A Data Construction

A.1 FIEs in China

The registration file from the Ministry of Commerce (MOFCOM) records each foreign-invested enterprise (FIE) with its name, address, establishment date, four-digit industry code, investment type and ownership structure, together with the name, jurisdiction and contributed capital of each investor. Registration was mandatory: since the 1992 *Regulations on the Administration of Registration for Foreign Enterprises* the industrial and commercial registration system and MOFCOM have jointly collected and approved information on every foreign-invested project, and until the 2016 *Interim Measures for Record-Filing Administration* projects were approved case by case rather than merely filed. The file therefore covers the population of registered FIEs rather than a sample of them.

We clean the file in four steps. First, we drop records with a missing establishment date or a missing investor jurisdiction. Second, we standardize investor jurisdiction names, which appear in several transliterations, to ISO3 codes. Third, we assign each firm a single source economy, the jurisdiction of its largest shareholder by contributed capital, following Ma and Zhang (2025). Fourth, we classify ownership form from the capital-structure field: a firm is wholly foreign owned (WFOE) when no domestic investor appears among its shareholders and a joint venture (JV) otherwise. That split is what divides embedded technology into its two components in Table G3. Investment amounts are converted to U.S. dollars at the year's average exchange rate using Penn World Table 10.0.

The cleaned file contains 824,565 FIEs established before 2019, with investors from 205 economies across 1,364 CIC4 industries. We aggregate the firm-level data to the country–industry–year level for empirical analysis. During this process, Chinese capital contributions are excluded, and only the actual foreign-invested amounts are retained to more accurately measure the intensity and industry distribution of foreign investment.

Figure A1 compares our sample data with publicly available records from the MOFCOM on

the number of newly established FIEs in China.[1] The two track each other closely in both level and timing, which is the external check available for a registry of this kind. In certain years (e.g., 1992–1994), MOFCOM statistics differ from our sample data, likely due to early-stage data collection biases such as underreporting or retroactive corrections; however, these discrepancies do not alter the overall trend conclusions. Table A1 reports the distribution of entrants by source economy, Table A2 by industry and Table A3 by technology field.

A.2 Global Patent Data

Global patent data come from the Autumn 2023 release of PATSTAT Global. We process the data as follows. (1) We retain applications filed through 2022 to mitigate right truncation arising from publication delays, as patent applications are typically published within 18 months after filing. (2) We retain invention patents and exclude utility models, designs, and other patent types. (3) Following European Patent Office conventions, we de-duplicate application records. The resulting data contain 92.3 million application records belonging to 60.2 million DOCDB families. We use the patent family as the statistical unit so that the same invention filed in multiple jurisdictions is counted once, and date each family by its earliest application year. (4) We use citations at the family level and link citing and cited families to construct the inter-field citation network. Before imposing the citation-window restrictions described below, the network contains 14.28 million citing and 15.00 million cited families. (5) For the baseline KIO network, we restrict citing families to those filed between 1996 and 2016 and subsequently granted within the observation window, allowing sufficient time for citations to accumulate. After these restrictions, the citation network contains 6.49 million citing and 5.92 million cited families. For the pre-reform network used in the instrument, we apply the same restrictions to citing families filed between 1992 and 2001.

Three additional conventions govern country and time attribution. A family's country of origin is determined from inventor residence, with applicant country used when inventor information is

[1] The MOFCOM data are sourced from the Commerce Data Center, https://data.mofcom.gov.cn/lywz/inmr.shtml.

unavailable and the patent-office country used as a final fallback. When inventors reside in multiple countries, the family is allocated fractionally across them. Families are dated by application rather than grant year because application timing more closely reflects when the underlying invention was formalized, whereas grant timing additionally reflects examination delays that vary across offices and filing cohorts.

A.3 Patent Transactions and Collaboration by FIEs in China

Patent assignments, licenses, and joint applications involving FIEs in China come from Chinese patent records in the IncoPat Global Patent Database for 1985–2022. The assignment and licensing records cover transactions registered with the China National Intellectual Property Administration (CNIPA). Because the patent and FIE registries do not share a common firm identifier, we link them by applicant names.

We first standardize firm names by removing punctuation and other non-informative symbols. For initially unmatched observations, we apply a second-stage normalization that removes Chinese administrative and geographic terms (e.g., the characters for "province," "autonomous region," "city," "district," and "county") and common legal-form or organizational terms (e.g., "limited," "liability," "shares," "company," and "factory"). Ambiguous matches are manually reviewed.

For patent assignments, we retain granted invention patents with corporate applicants and identify transactions by their assignment execution dates. We classify an assignment as FIE-to-domestic when the assignor matches an FIE in the MOFCOM registry and the assignee is a non-FIE Chinese entity. Each assigned patent is counted once and aggregated to the IPC4 field-year level. Licensing records are processed analogously, requiring an FIE licensor and a non-FIE licensee. For collaboration, we identify patent applications jointly filed by an FIE and at least one domestic non-FIE entity. These definitions yield 10,668 assignments from FIEs to domestic entities, 1,138 licenses, and 44,885 joint applications.

The same linkage supports two additional exercises. First, we identify 269,077 FIE-affiliated DOCDB families. Columns (2)–(4) of Table 6 exclude these families from the citing side, cited

side, or both sides of the cross-border backward citation measure to assess whether the citation response reflects changes in patenting composition. Second, the 44,605 unique DOCDB families underlying the collaboration measure are linked back to PATSTAT to identify the patent offices in which family members subsequently appear. Table G4 uses this information to restrict collaboration to inventions for which protection was subsequently sought abroad, including in the foreign investor's home jurisdiction.

A.4 The CIC4-to-IPC4 Concordance

Foreign investment is recorded by industry and the analysis is conducted by technology field, so the two classifications must be linked. The procedure is as follows. First, patent data are classified according to the International Patent Classification and retain their original four-digit IPC (IPC4) codes. Second, industry data are categorized under China's National Economic Industry Classification (CIC) using its four-digit codes (CIC4), with the 2017 version serving as the baseline for FDI statistics. Third, the linkage between FDI industry information and patent technology fields is constructed using the *Reference Table between International Patent Classification and National Economic Industry Classification (2018)* published by the CNIPA, compiled by the Planning and Development Department of CNIPA. It was produced to connect patent statistics to national economic statistics, and it is built by subject-matter experts who check, entry by entry, whether the technical content of an IPC heading corresponds to the economic activity described in the notes to the industry classification. Entries whose technical content matches completely are flagged.

The raw table lists 17,999 IPC-group $\times$ CIC4 entries, of which 4,112 carry the complete-correspondence flag. We process the table in three steps. First, we truncate IPC groups to four-character IPC subclasses (IPC4) and collapse duplicate CIC4–IPC4 pairs created by this aggregation, yielding 3,737 pairs, of which 793 inherit the complete-correspondence flag. Second, we match these IPC4 codes to the IPC scheme used in PATSTAT, dropping subclasses retired in intervening revisions. Third, we restrict the concordance to industries and technology fields appearing in the estimation sample. The resulting concordance contains 3,736 CIC4–IPC4 pairs linking 819

industries to 633 technology fields. Table A7 reports the corresponding coverage statistics.

The mapping is many-to-many. A CIC4 industry maps to 4.56 technology fields on average (median 3), while an IPC4 field maps to 5.90 industries on average (median 4); Table A7 reports the full distribution. When an industry maps to multiple technology fields, we divide its FDI equally across the matched fields, preserving the industry's total investment while avoiding additional assumptions about its technological composition. Table D3 reports robustness to assigning the full investment amount to each matched field.

Alternative concordance. As an alternative, we construct an industry–technology mapping through HS2002 products, combining the industry–product correspondence of Brandt et al. (2017) with the probabilistic product–technology concordance of Lybbert and Zolas (2014). This route covers only 305 CIC4 industries because the Harmonized System classifies traded goods and therefore provides limited coverage of services and other non-traded activities.

Sensitivity. Table D5 re-estimates the baseline using the HS-based mapping, the reverse mapping with the analysis conducted in CIC4 rather than IPC4 space, and aggregation to the coarser IPC3 level. Where an alternative concordance covers only manufacturing, we hold the comparison sample fixed accordingly.

A.5 Contemporaneous Policy Controls

China's WTO accession coincided with several reforms that could differentially affect innovation across technology fields. The baseline therefore controls for five potentially confounding policy dimensions: tariff concessions, state-owned-enterprise (SOE) reform, U.S. permanent normal trade relations (PNTR), the removal of textile and clothing quotas, and changes in intellectual-property protection associated with TRIPS. We also allow pre-reform industry characteristics to generate differential time paths across technology fields. Except for the quota fill rate, which follows the phase-out schedule, exposure is measured before the relevant policy change and then interacted with year indicators or a post-policy indicator, so that the controls absorb differential trends associ-

ated with predetermined exposure rather than contemporaneous outcomes. The PNTR and TRIPS controls take the common form $Exposure_i \times \mathbf{1}[t \geq 2001]$, switching on in the year of accession.

Trade liberalization. China's WTO accession was accompanied by substantial and uneven tariff reductions across industries. To account for the possibility that fields exposed to greater trade liberalization experienced different innovation dynamics, we construct three tariff measures, all fixed at their 2001 values. The output tariff is the applied import tariff on an industry's own products. The input tariff is the input-share-weighted average tariff on the products supplied to that industry. The partner-weighted foreign tariff measures the tariffs faced by Chinese exporters in destination markets, weighted by the industry's pre-reform export shares. Tariff data come from Brandt et al. (2017) and the World Integrated Trade Solution. The measures are constructed at the adjusted four-digit industry level, mapped to CIC4 industries and then to IPC4 technology fields through the concordance described in Appendix A.4, and interacted with year indicators. The log value of Chinese imports enters the baseline separately to account for variation in access to foreign technology through trade.

State-owned-enterprise reform. The restructuring of China's state sector during the late 1990s and early 2000s could also have generated differential innovation trends across industries (Hsieh and Song, 2015). We measure pre-reform state ownership using the SOE share of industry assets in 2001, calculated from the Annual Survey of Industrial Enterprises (ASIE). The measure is constructed across 483 manufacturing industries, mapped to IPC4 technology fields through the CIC4–IPC4 concordance, and interacted with year indicators. Fixing the measure at its pre-reform value avoids conditioning on subsequent changes in state ownership that could themselves respond to foreign entry or innovation performance.

Permanent normal trade relations. The granting of permanent normal trade relations (PNTR) to China by the United States reduced the possibility that Chinese exports would face substantially higher non-NTR tariffs and thereby lowered trade-policy uncertainty (Pierce and Schott, 2016; Handley and Limão, 2017). We measure field-level exposure using the NTR gap, defined as the

difference between the non-NTR tariff and the applied NTR tariff. The gap is measured at the U.S. HS8 level in 1999, the last complete year before PNTR legislation, averaged to HS6, converted from HS1996 to HS2002 using the United Nations Statistics Division concordance, and mapped to IPC4 fields using the probabilistic product–technology weights of Lybbert and Zolas (2014). Negative gaps are set to zero. The resulting control is

$$NTRGap_{i,1999} \times \mathbf{1}[t \geq 2001].$$

Removal of textile and clothing quotas. The Agreement on Textiles and Clothing (ATC) gradually eliminated the quotas imposed under the Multi-Fibre Arrangement (MFA), with the final phase taking effect in January 2005 (Khandelwal et al., 2013). We measure quota bindingness using product-level fill rates from Brambilla et al. (2010), defined as actual exports relative to the quota ceiling, so that a higher fill rate indicates a more binding quota. MFA product categories are linked to HS products using the correspondence employed by Pierce and Schott (2016), converted across HS vintages, and mapped to IPC4 fields using the probabilistic concordance of Lybbert and Zolas (2014). The control is the field-level fill rate in each year, set to zero once the corresponding quotas are removed: from 2002 for products in the first three phases, which China's accession brought forward, and from 2005 for products in the final phase.

TRIPS compliance. China's WTO accession also required compliance with the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), which strengthened formal intellectual-property protection. To allow the effect of this change to differ with a field's reliance on intellectual property, we construct a measure of IP intensity using the 2013 joint report of the European Patent Office and the Office for Harmonization in the Internal Market.[2] The report classifies 321 four-digit NACE Rev.2 industries according to their intellectual-property intensity. We map these industries

[2]Two considerations justify using a 2013 classification to control for a shock dated 2001. First, it is the earliest systematic classification of intellectual-property intensity we can find, so the alternative is not an earlier vintage but no measure at all. Second, the report itself characterizes intellectual-property intensity as a structural feature of an industry's technology rather than a cyclical one, and documents that it moves little over time; if the attribute is close to fixed, a later measurement of it is an acceptable proxy for its value in the earlier period.

to technology fields using the probabilistic concordance of Lybbert and Zolas (2014), and classify a technology field as IP intensive when more than half of its mapped industry weight comes from IP-intensive industries. The resulting indicator is interacted with an indicator for years from 2001 onward. Although the classification is measured after China's accession, it is used as a proxy for a relatively persistent technological characteristic of industries rather than as a contemporaneous policy outcome.

Pre-reform industry characteristics. In addition to these policy controls, the baseline allows technology fields with different initial industry structures to follow different time paths. We construct four characteristics from the 1998 ASIE: new-product intensity, export intensity, the number of firms, and average firm age. Each variable is measured before the reform, mapped from industries to IPC4 technology fields through the CIC4–IPC4 concordance, and interacted with year indicators. The baseline also includes the 2001 SOE share and tariff measures described above in the same interacted form.

Domestic innovation policies. We separately consider four domestic policies that could affect innovation later in the sample period: the R&D super-deduction, preferential corporate income taxation for high- and new-technology enterprises (HNTEs), development zones, and the expansion of higher education. The first two became substantially more important following the 2008 Enterprise Income Tax Law. We construct industry-level exposure to the R&D super-deduction, HNTE relief, and government subsidies using the National Tax Survey for 2008–2016. Numerators and denominators are first aggregated separately across firms, converted to CIC2017 industries, mapped to IPC4 fields, and then recombined at the field level so that the ratio structure is preserved. Development-zone exposure is constructed from the State Council's list of approved development zones and allocated to technology fields using pre-reform provincial output shares. Exposure to the higher-education expansion is measured by field-level human-capital intensity interacted with an indicator for years from 2003 onward, following Che and Zhang (2018).

These later domestic innovation policies are not included in the baseline specification. Most were introduced substantially after the 2002 *Catalogue* reform, while several may themselves re-

spond to innovation performance or to subsequent industrial development. Conditioning on them in the baseline could therefore absorb part of the post-reform innovation process rather than remove a contemporaneous confound. Panel B of Table D1 instead introduces these controls sequentially as robustness checks.

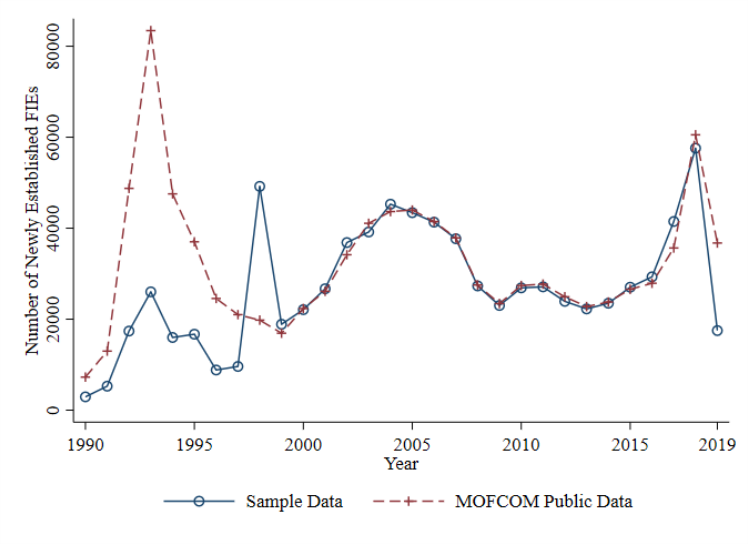


Figure A1: Data Comparison of Newly Established FIEs in China

Notes: The figure presents the annual number of newly established FIEs in China from 1990 to 2019, comparing our sample data with official statistics released by the MOFCOM via the Commerce Data Center. The vertical axis represents the number of FIEs, and the horizontal axis represents the year.

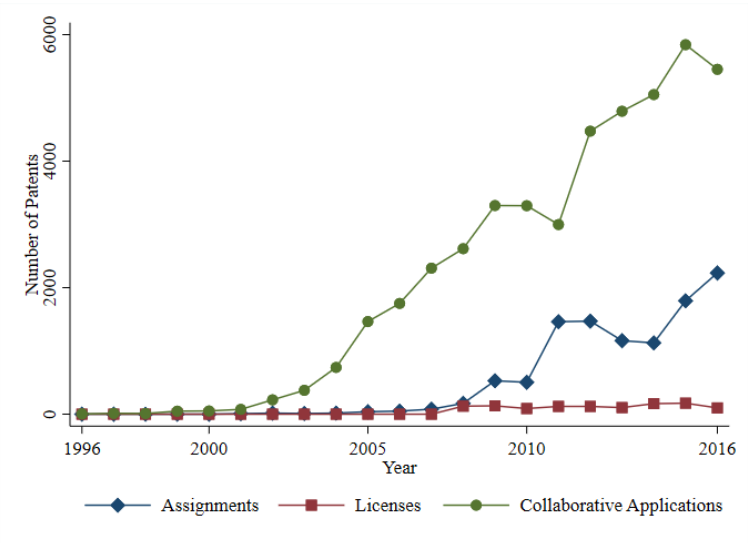


Figure A2: Patent Transactions and Collaboration by FIEs, 1996–2016

Notes: The figure plots the annual number of patents assigned by FIEs to domestic entities, the number licensed by FIEs to domestic entities, and the number applied for jointly by an FIE and at least one domestic entity. FIEs are identified by matching the MOFCOM registry to applicant names in CNIPA records from IncoPat (Appendix A). Counts are based on DOCDB patent families. Assignments and licenses are dated by the year in which the transaction was executed, and joint applications by application year.

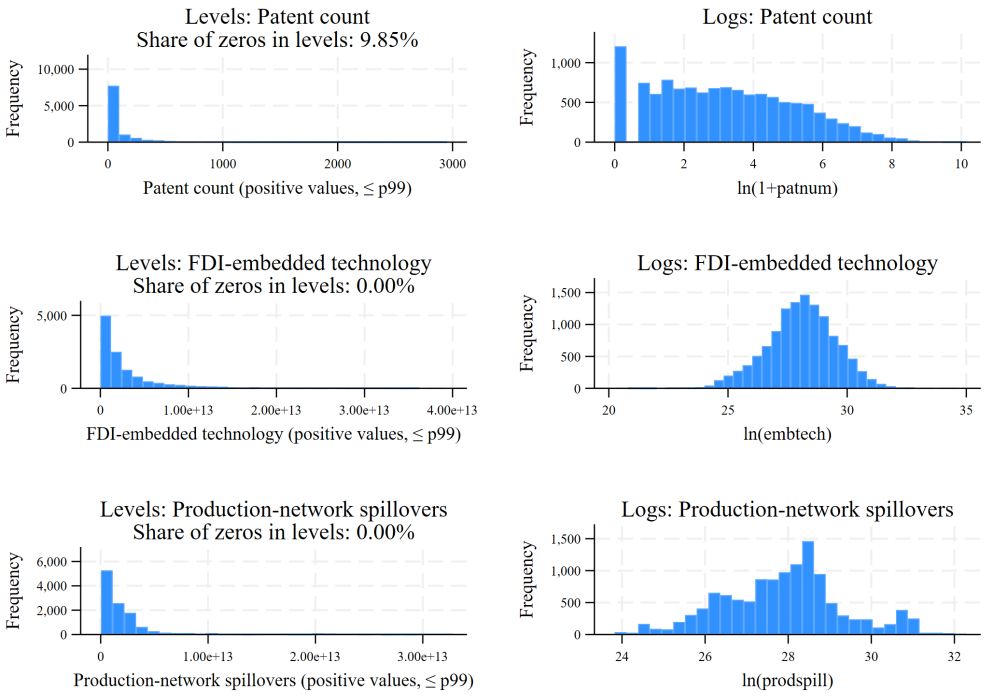


Figure A3: Distributions of the Dependent Variable and the Key Regressors

Notes: The figure plots the distributions of *patnum*, *KIOSPILL*, and *PIOSPILL* in levels (left) and logs (right) for the 12,096 field-year observations of the baseline sample. Level panels report the share of zeros and are truncated at the 99th percentile of positive values. Patent counts are strongly right-skewed, but only 9.85% are zero, and both exposure measures are strictly positive and approximately symmetric in logs. These features motivate the log specification as the baseline, with Poisson estimates reported alongside. Figure A4 reports the corresponding distributions for the mechanism outcomes.

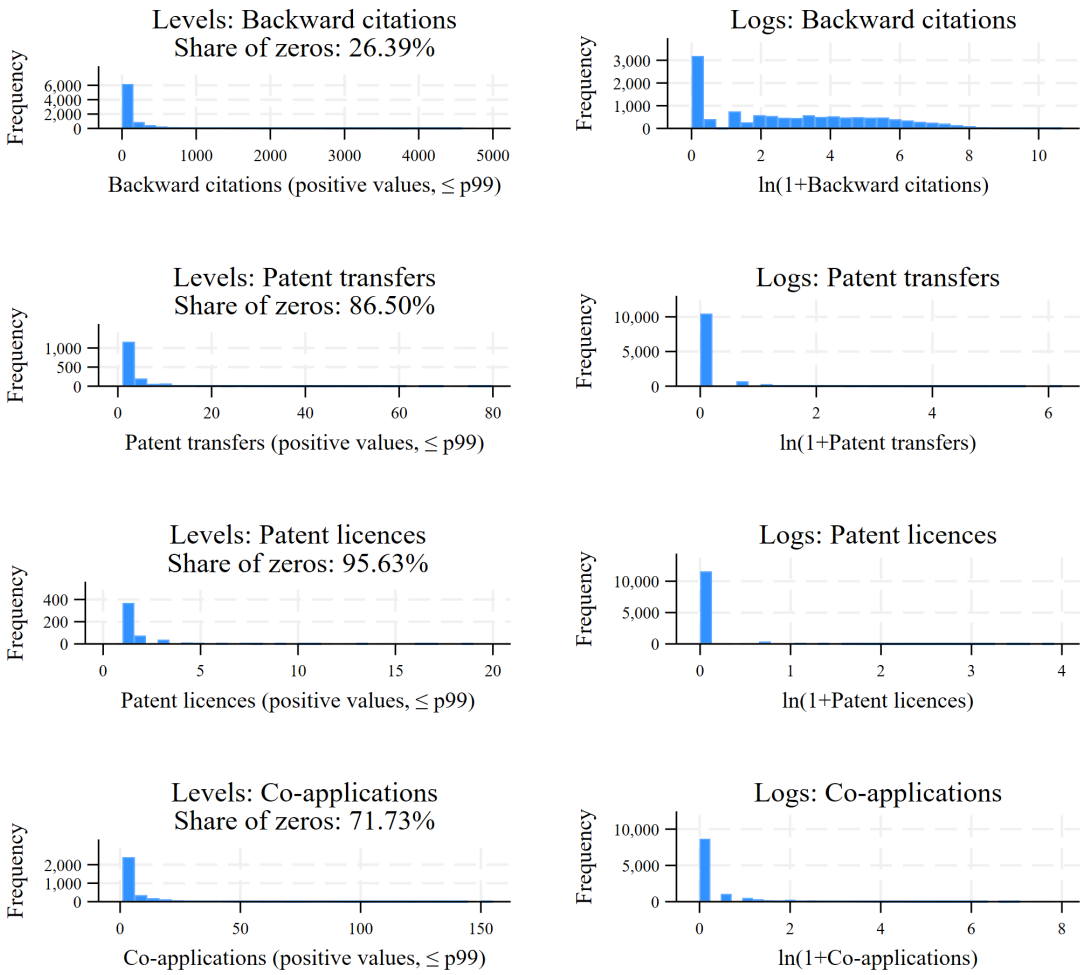


Figure A4: Distributions of the Mechanism Outcomes

Notes: The figure plots the distributions of the four mechanism outcomes, namely cross-border backward citations, collaborations, assignments, and licenses, in levels (left) and logs (right) for the 12,096 field-year observations of the baseline sample. Level panels report the share of zeros and are truncated at the 99th percentile of positive values. The share of zeros rises from 26% for cross-border citations to 96% for licenses, so Tables 6 and 7 report Poisson and extensive-margin estimates alongside the log specification. Table A6 reports the corresponding statistics.

Table A1: Geographic Distribution of Newly Established FIEs in Mainland China, 2000–2015

Source	Number	Share
Hong Kong, China	177,797	41.27%
Taiwan, China	45,331	10.52%
South Korea	42,169	9.79%
United States	29,777	6.91%
Japan	24,594	5.71%
British Virgin Islands	19,517	4.53%
Singapore	12,033	2.79%
Samoa	7,680	1.78%
Macao, China	7,296	1.69%
Canada	6,382	1.48%
Germany	5,395	1.25%
Australia	5,077	1.18%
United Kingdom	4,407	1.02%

Notes: The table reports the source economies of FIEs newly established in mainland China over 2000–2015. Each firm is assigned to the economy of its largest shareholder by contributed capital (Ma and Zhang, 2025). Number is the count of firms from each source, and Share is that count as a percentage of all new FIEs. Only sources accounting for more than 1% of firms are listed, so shares do not sum to 100%.

Table A2: Industry Distribution of Newly Established FIEs in Mainland China, 2000–2015

CIC4	Industry	Number	Share
7249	Other professional consulting and surveying services	23,604	5.48%
1819	Other woven apparel manufacturing	23,533	5.46%
5199	Other unspecified wholesale trade	13,432	3.12%
7010	Real estate development and operation	11,586	2.69%
6513	Application software development	10,989	2.55%
3981	Manufacture of resistors, capacitors and inductors	10,653	2.47%
5179	Wholesale of other machinery, equipment and electronic products	10,338	2.40%
3670	Manufacture of motor vehicle parts and accessories	7,148	1.66%
2929	Manufacture of plastic parts and other plastic products	6,664	1.55%
7499	Other unspecified professional and technical services	6,449	1.50%
5132	Wholesale of apparel	6,024	1.40%
7243	Socio-economic consulting services	5,942	1.38%
6299	Other unspecified food and beverage services	5,469	1.27%
3990	Manufacture of other electronic equipment	5,453	1.27%
3525	Manufacture of molds	4,888	1.13%
5146	Wholesale of jewelry, handicrafts and collectibles	4,586	1.06%

Notes: The table reports the industry distribution of FIEs newly established in mainland China over 2000–2015, by the CIC4 code recorded at registration. Number is the count of firms in the industry, and Share is that count as a percentage of all new FIEs. Only industries accounting for more than 1% of firms are listed, so shares do not sum to 100%.

Table A3: Distribution of Technology Fields for Newly Established FIEs in China, 2000–2015

IPC4	Technology field	Amount	Share
A23L	Foods, foodstuffs or non-alcoholic beverages	309.78	2.53%
A41D	Outerwear; protective garments; accessories	291.93	2.38%
A41B	Shirts; underwear; baby linen; handkerchiefs	263.88	2.15%
E04F	Finishing work on buildings	191.24	1.56%
H01F	Magnets; inductances; transformers	177.32	1.45%
H01G	Capacitors; electrolytic rectifiers, detectors and switching devices	174.88	1.43%
H01L	Semiconductor devices; electric solid-state devices	167.37	1.37%
C12N	Microorganisms or enzymes; compositions thereof	161.22	1.32%
H02K	Dynamo-electric machines	155.66	1.27%
G06F	Electric digital data processing	152.50	1.25%
H01M	Direct conversion of chemical energy into electrical energy, e.g. batteries	151.26	1.23%
H01P	Waveguides; resonators, lines or other waveguide-type devices	147.28	1.20%
H01C	Resistors	142.19	1.16%
C10L	Fuels; natural gas; synthetic natural gas	125.67	1.03%

Notes: The table reports the distribution of registered foreign investment by FIEs newly established over 2000–2015 across IPC4 technology fields. Industries are mapped to fields using the concordance in Appendix A.4, with an industry's investment divided equally among its fields. Amount is registered foreign investment in hundreds of millions of U.S. dollars, and Share is that amount as a percentage of the total. Only fields accounting for more than 1% of the total are listed, so shares do not sum to 100%.

Table A4: Summary Statistics

Variable	N	Mean	SD	Min	Max
Panel A. Technology field × year					
$\ln(1+patnum)$	12,096	3.167	2.064	0.000	10.146
$\ln(1+backcite)$	12,096	2.896	2.441	0.000	10.670
$\ln(1+transfer)$	12,096	0.177	0.543	0.000	6.230
$\ln(1+license)$	12,096	0.040	0.213	0.000	3.912
$\ln(1+collaborate)$	12,096	0.472	0.941	0.000	7.073
$\ln KIOSPILL$	12,096	28.068	1.438	20.803	32.840
cross-field component	12,096	28.028	1.429	20.803	32.672
same-field component	12,096	19.504	6.355	0.000	32.816
network diagonal removed	12,096	28.378	1.409	20.936	32.677
$\ln PIOSPILL$	12,096	27.878	1.451	23.823	32.287
$\ln IV^{KIO}$	12,096	5.460	3.963	0.000	11.121
$\ln IV^{PIO}$	12,096	5.589	4.011	0.000	10.501
$\ln globalpat$	12,096	5.325	1.683	0.000	10.475
$\ln import$	12,096	18.894	2.092	6.848	26.033
PNTR shock	12,096	0.223	0.179	0.000	1.255
Textile quota fill rate	12,096	0.009	0.065	0.000	1.000
TRIPS shock	12,096	0.285	0.452	0.000	1.000
R&D super-deduction intensity	12,096	0.002	0.007	0.000	0.327
HNTE tax-relief intensity	12,096	0.046	0.079	0.000	0.576
Government subsidy intensity	12,096	0.001	0.003	0.000	0.095
Higher-education expansion shock	12,096	0.076	0.069	0.000	0.279
Panel B. Technology field (pre-determined)					
Catalogue liberalization indicator	576	0.413	0.493	0.000	1.000
Openness intensity (continuous)	576	0.264	0.421	0.000	2.000
Network policy exposure	576	0.471	0.248	0.000	0.962
New product intensity, 1998	576	0.095	0.065	0.000	0.460
Export intensity, 1998	576	0.207	0.163	0.000	0.790
Number of firms, 1998	576	310.988	270.962	9.000	2438.620
Average firm age, 1998	576	16.187	4.638	5.300	39.656
State-owned share, 2001	576	0.429	0.187	0.009	0.989
Input tariff, 2001	576	9.863	1.892	3.997	23.897
Output tariff, 2001	576	14.524	5.730	3.429	47.761
Partner-weighted tariff, 2001	576	3.913	2.445	0.230	25.004
Development-zone intensity, 2001	576	0.007	0.003	0.001	0.026
China–ROW technology gap, PCT	576	0.082	0.919	−2.815	2.066
Source-economy lead, PCT	565	0.055	0.977	−1.728	1.728
China–ROW technology gap, all offices	576	0.088	0.889	−4.605	2.446
Source-economy lead, all offices	565	0.055	0.985	−1.728	1.728
$\ln centrality$, 1996–2016	576	−8.161	1.979	−14.285	−2.156
$\ln centrality$, 1992–2001	576	−8.099	2.033	−14.982	−2.247
Panel C. Source economy × technology field × year					
$\ln KIOSPILL$	2,322,432	4.889	7.512	0.000	32.757
$\ln IV^{KIO}$	2,322,432	0.325	1.074	0.000	10.425
$\ln(1+backcite)$	2,322,432	0.088	0.483	0.000	10.003

Notes: The table reports summary statistics for the variables used in the analysis. Panel A covers the baseline estimating sample of 576 IPC4 technology fields observed annually over 1996–2016. Panel B reports the pre-determined field characteristics, which are constant within a field. Panel C covers the source economy × field × year panel used in Table G1, which covers 192 source economies. Zero shares and the distribution of the positive values of the counts underlying the logged outcomes are reported in Table A6.

Table A5: Tax Haven Lists

Country/Region	ISO3	List 1	List 2	List 3	List 4	List 5
Andorra	AND	1	1	0	1	1
Anguilla	AIA	1	1	1	1	1
Antigua and Barbuda	ATG	1	1	1	1	1
Aruba	ABW	1	1	1	1	1
Bahamas	BHS	1	1	1	1	1
Bahrain	BHR	1	1	1	1	1
Barbados	BRB	1	0	1	1	1
Belize	BLZ	1	1	1	1	1
Bermuda	BMU	1	0	1	1	1
British Virgin Islands	VGB	1	1	1	1	1
Cayman Islands	CYM	1	1	1	1	1
Channel Islands	/	1	1	1	1	1
Cook Islands	COK	1	1	1	1	1
Costa Rica	CRI	1	1	1	0	1
Cyprus	CYP	1	0	1	1	0
Dominica	DMA	1	1	1	1	1
Gibraltar	GIB	1	1	1	1	1
Grenada	GRD	1	1	1	1	1
Guatemala	GTM	0	0	0	0	1
Hong Kong, China	HKG	1	0	1	1	0
Ireland	IRL	0	0	0	1	0
Isle of Man	IMN	1	0	1	1	1
Jordan	JOR	0	1	0	1	0
Lebanon	LBN	0	1	0	1	0
Liberia	LBR	1	1	1	1	1
Liechtenstein	LIE	1	1	1	1	1
Luxembourg	LUX	0	0	0	1	0
Macao, China	MAC	0	0	0	1	1
Maldives	MDV	1	1	0	1	0
Malta	MLT	1	0	1	1	0
Marshall Islands	MHL	1	1	0	1	1
Mauritius	MUS	1	0	0	0	1
Monaco	MCO	1	1	0	1	1
Montserrat	MSR	1	1	0	1	1
Nauru	NRU	1	1	1	0	0
Niue	NIU	1	1	0	0	0
Panama	PAN	1	1	1	1	1
Samoa	WSM	1	1	1	1	1
San Marino	SMR	1	0	0	1	1
Seychelles	SYC	1	0	0	1	1
Singapore	SGP	1	1	1	1	0
St. Kitts and Nevis	KNA	1	1	1	1	1
St. Lucia	LCA	1	1	1	1	1
St. Vincent and the Grenadines	VCT	1	1	1	1	1
Tonga	TON	1	1	0	0	0
Turks and Caicos Islands	TCA	1	1	1	1	1
U.S. Virgin Islands	VIR	1	0	0	1	0
Uruguay	URY	0	0	0	0	1
Vanuatu	VUT	1	1	0	1	1

Notes: The table reports the classification of tax havens based on five lists. List 1 covers jurisdictions identified by entering a Tax Information Exchange Agreement under the OECD Harmful Tax Practices framework; List 2 is the OECD Grey List (as of August 17, 2009); List 3 corresponds to the "Stop Tax Haven Abuse Act" (not enacted); List 4 refers to the original OECD Tax Haven List; List 5 follows the classification by Hines and Rice (1994). The data are obtained from Bennedsen and Zeume (2018).

Table A6: Zeros and the Distribution of the Positive Values

Variable	Share of zeros (%)	N	Mean	Median	P99
patnum	9.85	12,096	205.9	27	2,950
backcite	26.39	12,096	374.7	45	4,583
collaborate	71.73	12,096	12.72	3	155
transfer	86.50	12,096	6.17	2	80
license	95.63	12,096	2.04	1	19

Notes: For each count variable the table reports the share of observations equal to zero and, among the positive values, the mean, the median and the 99th percentile. *patnum* is the number of granted Chinese patent families dated by application year, *backcite* the number of citations from Chinese patents to foreign patents, and *collaborate*, *transfer* and *license* the counts of co-applied, assigned and licensed patents involving foreign-invested enterprises. The sample is the baseline estimating sample of 12,096 field–year observations. The dependent variable has a limited zero mass whereas the mechanism outcomes are progressively sparser. The two exposure measures, *KIOSPILL* and *PIOSPILL*, contain no zeros.

Table A7: Coverage and Density of the Benchmark Concordance

Panel A. Coverage

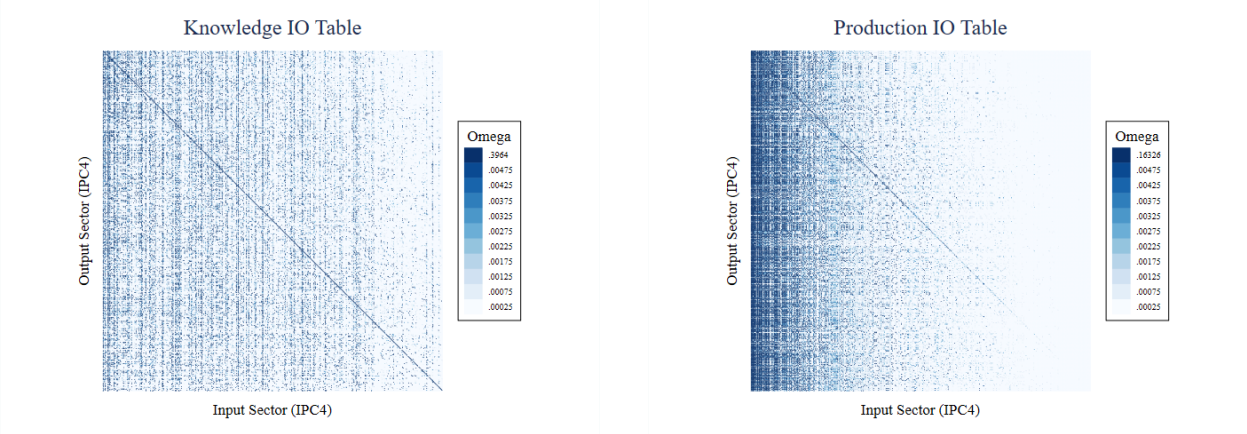
Statistic	Value	Source
Raw entries (IPC group × CIC4)	17,999	CNIPA 2018, Sheet 1
of which flagged as complete correspondences	4,112	group level
IPC4 × CIC4 pairs after aggregation to subclasses	3,737	before IPC 2018 filter
of which complete correspondences	793	
Unique CIC4 industries, before filter	819	
Unique IPC4 subclasses, before filter	634	
Unique CIC4 industries, working concordance	819	
Unique IPC4 subclasses, working concordance	633	matched to PATSTAT
Total IPC4 × CIC4 pairs, working concordance	3,736	final file

Panel B. Density

Mapping density	Mean	Median	P10	P25	P75	P90	Min	Max
IPC4 fields per CIC4 industry	4.56	3	1	2	5	9	1	111
CIC4 industries per IPC4 field	5.90	4	1	2	8	13	1	50

Notes: The table documents the concordance used to map four-digit China Industry Classification (CIC4) industries to four-digit International Patent Classification (IPC4) technology fields. Panel A reports how the published *Reference Table* of the China National Intellectual Property Administration (CNIPA, 2018 edition) is reduced to the working file, and Panel B the density of the resulting mapping. A complete correspondence is an entry the source flags as one-to-one at the IPC group level. The final step retains only IPC4 subclasses present in PATSTAT.

B Innovation Network



(a) Innovation Network (b) Production Network

Figure B1: Innovation and Production Networks at IPC4 Level

Notes: The figure repeats Figure 1 at the IPC4 level used in estimation. Cell (i, j) is the weight of upstream field j (horizontal axis) in the inputs of using field i (vertical axis); darker cells indicate larger weights. Panel (a) shows knowledge input shares ω_{ij} computed from citations on U.S. patents over 1996–2016, and panel (b) shows production input shares ν_{ij} from the 1997 benchmark use table of the U.S. Bureau of Economic Analysis (BEA), mapped to technology fields using Lybbert and Zolas (2014). In both panels, fields are ordered by their importance as upstream sources in the production network, measured by off-diagonal column sums. Both panels cover the 603 fields for which the two networks are defined; the Pearson correlation between ω_{ij} and ν_{ij} across the 363,609 ordered field pairs is 0.17, compared with 0.32 at the IPC3 level.

Table B1: Top 20 Upstream Fields of Battery Technology (H01M) in the Two Networks

	Ranked by the innovation network			Ranked by the production network		
Rank	IPC4	ω	ν	IPC4	ω	ν
1	H01M	0.576101	0.067276	C22C	0.002783	0.108387
2	H02J	0.031250	0.010071	H01M	0.576101	0.067276
3	C01B	0.024048	0.006076	H01L	0.014484	0.060989
4	B01J	0.018128	0.003467	E06B	0.000160	0.054517
5	H01G	0.016429	0.000503	B29C	0.004220	0.047629
6	B01D	0.014633	0.002041	C22F	0.000200	0.038071
7	H01L	0.014484	0.060989	B60C	0.000108	0.026326
8	C01G	0.013789	0.001445	B24D	0.000040	0.022648
9	G01R	0.012118	0.001267	B21B	0.000154	0.021029
10	G06F	0.009466	0.003255	C03B	0.000342	0.020195
11	C25B	0.009346	0.001632	B65D	0.001180	0.017358
12	C23C	0.008959	0.013854	C04B	0.005252	0.016798
13	H01B	0.008525	0.005954	B32B	0.007790	0.015877
14	G01N	0.008046	0.000013	B22F	0.003028	0.015769
15	B32B	0.007790	0.015877	B22D	0.000781	0.015137
16	A61N	0.007214	0.000000	C23C	0.008959	0.013854
17	B05D	0.007128	0.000152	F21V	0.000416	0.013154
18	C08J	0.007037	0.004413	F16B	0.000684	0.013091
19	B60K	0.006239	0.000016	F16L	0.000821	0.012361
20	A61B	0.006079	0.000015	B24B	0.000177	0.011994

Notes: The table lists the 20 largest upstream fields of battery technology (H01M) in each network. The left block ranks fields by the knowledge input share ω in equation (2), the share of H01M's backward citations directed to the field, computed from citations on U.S. patents over 1996–2016. The right block ranks fields by the production input share ν underlying equation (3), the share of H01M's intermediate inputs supplied by the field, computed from the 1997 BEA benchmark use table and mapped to technology fields using Lybbert and Zolas (2014). Both weights are reported in each block, and H01M's weights in each network sum to one across all upstream fields. H01M itself appears in both rankings.

Table B2: Stability of the Innovation Network over Sample Periods

	1996–2016	1996–2006	2001–2011	2006–2016	1992–2001
1996–2016	1				
1996–2006	0.976	1			
2001–2011	0.988	0.974	1		
2006–2016	0.982	0.935	0.969	1	
1992–2001	0.958	0.969	0.947	0.926	1

Notes: Entries are Pearson correlations, across ordered field pairs, between knowledge input shares ω_{ij} computed over different windows from the citations of invention patents from all countries and regions. The three overlapping windows within 1996–2016 test stability over the sample period, and the 1992–2001 window is the pre-reform network used to construct the instrument (Section 4.1). The four windows within 1996–2016 share 92,649 field pairs, of which 74,826 also appear in the 1992–2001 window; correlations involving that window are computed on the pairs it shares with the other window. The pre-reform window is the least correlated with later windows, as the longer time gap implies, but all its correlations exceed 0.92.

Table B3: Similarity of Innovation Networks Across Regions

	World	JP	CN	US	KR	DE	RU	TW	FR	GB	AU
World	1										
JP	0.954	1									
CN	0.950	0.940	1								
US	0.932	0.878	0.862	1							
KR	0.932	0.930	0.931	0.873	1						
DE	0.936	0.930	0.917	0.899	0.907	1					
RU	0.890	0.887	0.908	0.831	0.881	0.883	1				
TW	0.843	0.845	0.833	0.844	0.851	0.838	0.810	1			
FR	0.912	0.900	0.893	0.886	0.892	0.922	0.873	0.838	1		
GB	0.883	0.872	0.859	0.876	0.864	0.889	0.851	0.838	0.889	1	
AU	0.856	0.850	0.834	0.855	0.846	0.868	0.841	0.830	0.859	0.865	1

Notes: Entries are Pearson correlations, across ordered field pairs, between knowledge input shares computed from the citations made by patents originating in each economy, for the world as a whole and for the ten economies with the most patents over 1996–2016. All years in 1996–2016 are pooled.

Table B4: Most and Least Important Upstream Fields, IPC3 Level

Innovation network		Production network	
IPC3	Description	IPC3	Description
Panel A. Five most important fields			
A61	Medical or veterinary science; hygiene	H01	Basic electric elements
H01	Basic electric elements	B21	Mechanical metal-working without removing material
G06	Computing; calculating; counting	F41	Weapons
G01	Measuring; testing	C22	Metallurgy; ferrous or non-ferrous alloys
B01	Physical or chemical processes or apparatus	E04	Building
Panel B. Five least important fields			
C14	Skins; hides; pelts; leather	G12	Instrument details
G12	Instrument details	E03	Water supply; sewerage
B68	Saddlery; upholstery	C40	Combinatorial technology
C13	Sugar industry	G03	Photography; cinematography; electrography
C40	Combinatorial technology	G07	Checking devices

Notes: A field's importance as an upstream source is the sum of its weights in the inputs of all other fields, excluding its own diagonal weight. The innovation network is computed from citations on U.S. patents over 1996–2016, and the production network from the 1997 BEA benchmark use table (Ayerst et al., 2023), mapped to technology fields using Lybbert and Zolas (2014). Rankings cover fields defined in both networks and exclude IPC residual classes for subject matter not otherwise provided for. Panel A orders fields from most to least important, and Panel B from least to most important.

Table B5: Most and Least Important Upstream Fields, IPC4 Level

Innovation network		Production network	
IPC4	Description	IPC4	Description
Panel A. Five most important fields			
G06F	Electric digital data processing	H01L	Semiconductor devices; electric solid-state devices
H01L	Semiconductor devices; electric solid-state devices	B21B	Rolling of metal
B01D	Separation	F16B	Devices for fastening or securing constructional elements
A61B	Diagnosis; surgery; identification	B60C	Vehicle tires; tire inflation; tire changing
A61K	Preparations for medical, dental or toilet purposes	B29C	Shaping or joining of plastics
Panel B. Five least important fields			
F21W	Indexing scheme for uses or applications of lighting devices	G06D	Digital fluid-pressure computing devices
C06F	Matches; manufacture of matches	G06C	Digital computers with mechanical computing elements
B41G	Apparatus for bronze printing or line printing	B62C	Vehicles drawn by animals
G06D	Digital fluid-pressure computing devices	C12F	Recovery of by-products of fermented solutions
C23D	Enameling of metals	G03H	Holographic processes or apparatus

Notes: The table repeats Table B4 at the IPC4 level, with the same definitions and sample restrictions.

C The 2002 *Catalogue* Reform

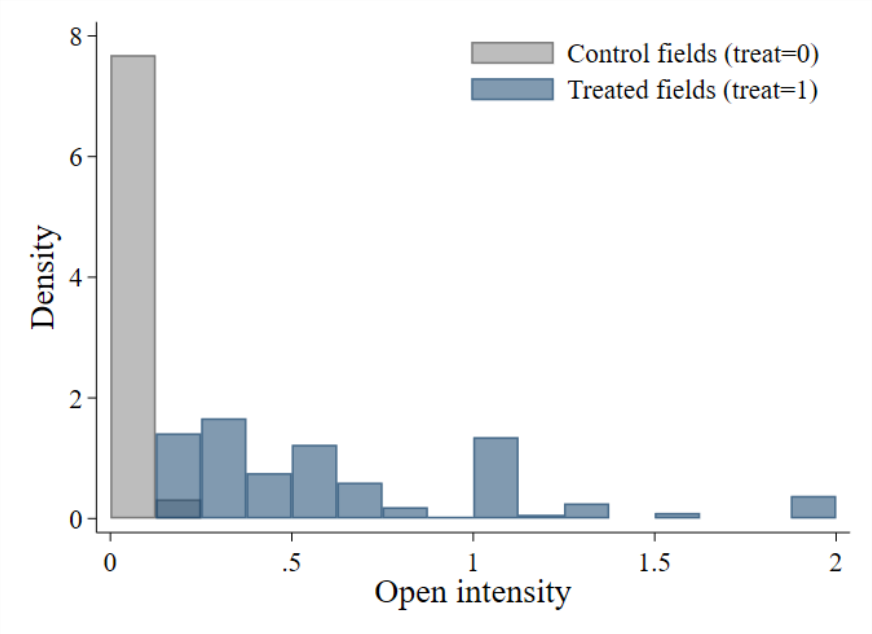
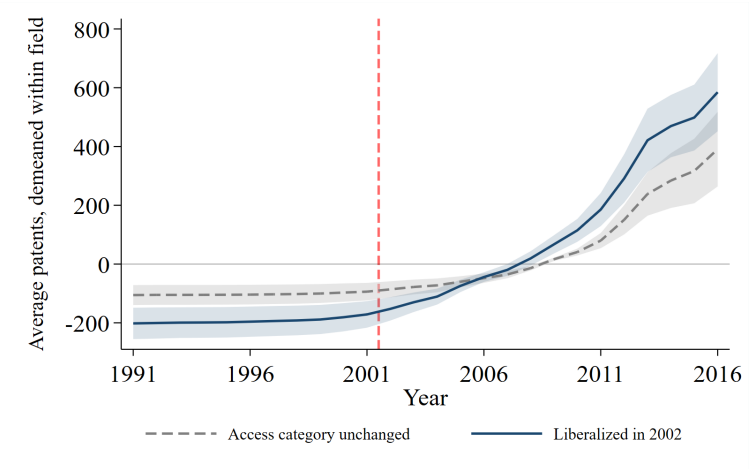
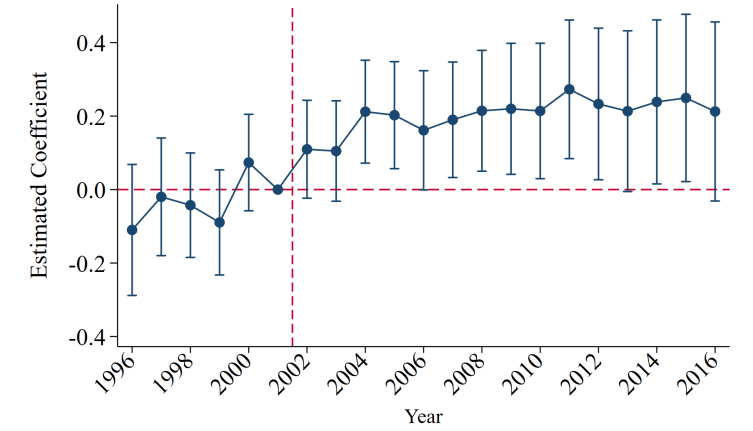


Figure C1: Distribution of the Continuous Openness Intensity Measure

Notes: Openness intensity is the exposure-weighted average number of categories by which a field's CIC4 industries moved toward greater access along the four-step ladder (prohibited, restricted, permitted, encouraged) between the 1997 and 2002 editions of the *Catalogue*. A value of zero indicates no movement, and a value of two an average movement of two categories. Bars are frequencies across the 576 IPC4 fields.



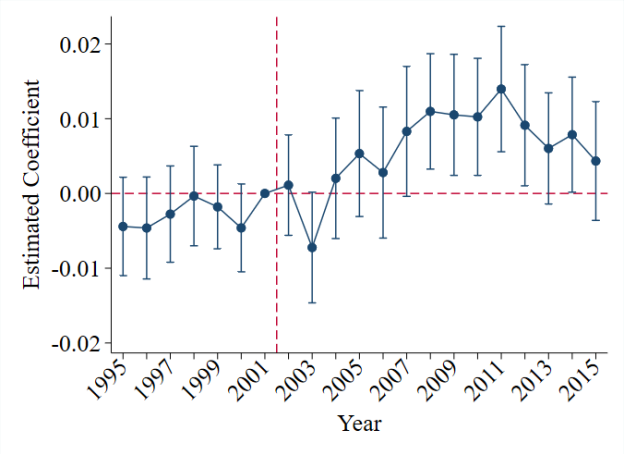
(a) Average patenting, demeaned within field



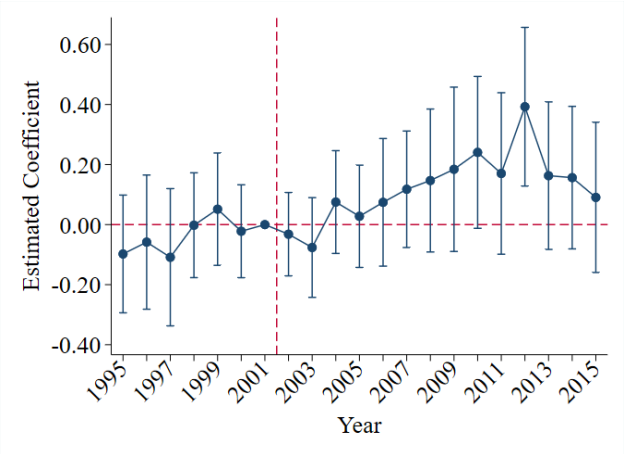
(b) Event-study coefficients

Figure C2: Pre-Reform Patenting Dynamics by *Catalogue* Liberalization Status

Notes: Panel (a) plots the average patent count, demeaned within field, over 1991–2016 for fields liberalized in 2002 (solid line) and fields whose access category was unchanged (dashed line); shaded bands are 95% confidence intervals across fields. Panel (b) reports coefficients from a field-level regression of $\ln(1 + patnum_{it})$ on the liberalization indicator interacted with year dummies over 1996–2016, with technology-field and year fixed effects and the controls of Figure 2. The omitted year is 2001. Dots denote point estimates and vertical bars 95% confidence intervals; the vertical dashed line marks the April 2002 reform. Standard errors are clustered by technology field. Liberalized and unchanged fields differ in levels but show no clear sustained divergence before the reform.



(a) Extensive margin



(b) Intensive margin

Figure C3: The FDI Response to the Reform, by Margin

Notes: The figure decomposes the event study in panel (a) of Figure 2 into two margins, using the same specification and controls. In panel (a), the dependent variable is an indicator for positive FDI in the source economy × field × year cell; in panel (b), it is $\ln(FDI)$ conditional on positive investment. The omitted year is 2001. Dots denote point estimates and vertical bars 95% confidence intervals; the vertical dashed line marks the April 2002 reform. Standard errors are two-way clustered by source economy and technology field. Participation rose after liberalization, while the scale of existing investment relationships did not systematically contract. Because panel (b) conditions on positive investment, its sample is selected, and it is reported descriptively.

Table C1: Changes in China's Foreign Investment Access Categories, 1997–2002

		2002 category				
		Encouraged	Permitted	Restricted	Prohibited	Total
1997 category	Encouraged	257	8	0	0	265
	Permitted	73	534	5	0	612
	Restricted	184	68	156	10	418
	Prohibited	1	2	11	37	51
	Total	515	612	172	47	1,346

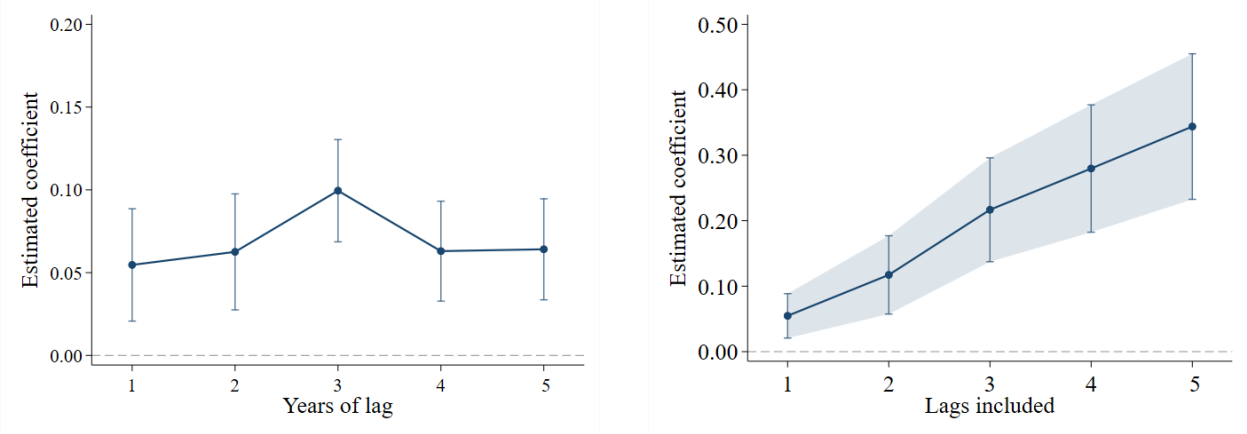
Notes: The table cross-tabulates the 1,346 four-digit industries (2017 CIC vintage) by their foreign-access category in the 1997 (rows) and 2002 (columns) editions of the *Catalogue for the Guidance of Foreign Investment Industries*. With categories ordered from encouraged to prohibited, cells below the diagonal represent liberalization and cells above it tightening. Of the 1,346 industries, 339 (25.2%) were liberalized, 984 (73.1%) were unchanged, and 23 (1.7%) were tightened. Among the liberalized industries, 152 moved by one category, 186 by two, and one by three; Figure C1 and Table C2 exploit this variation. The 23 tightened industries, which account for 0.5% of total technology-field weight, are excluded from the analysis. Within manufacturing, 110 of 569 industries (19.3%) were liberalized.

Table C2: Finer Variation in the Instrument: Continuous Openness Intensity

	Openness intensity		Restricted sectors only	
	(1) First stage $\ln KIOSPILL$	(2) 2SLS $\ln patnum$	(3) First stage $\ln KIOSPILL$	(4) 2SLS $\ln patnum$
$\ln KIOSPILL$		1.544***		1.595***
		(0.482)		(0.421)
$\ln IV^{KIO}$	0.091***		0.102***	
	(0.017)		(0.016)	
Observations	12,096	12,096	12,096	12,096
KP-rk LM statistic		29.232		40.530
KP-rk F statistic		30.157		41.232

Notes: The table replaces the binary treatment indicator with the continuous openness intensity of Figure C1, so that the instrument reflects the extent of liberalization rather than only its occurrence. The dependent variable is $\ln KIOSPILL$ in the first-stage columns and $\ln(1+patnum)$ in the 2SLS columns. Columns (1) and (2) use openness intensity for all fields; columns (3) and (4) use it only for industries that were restricted in 1997, for which the reform represented the largest change in access. Because openness intensity is defined relative to the 1997 category, the 1995 restricted category interacted with year dummies is dropped from the investment equation (4). Other specification choices follow the baseline. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

D Robustness of the Baseline



(a) Distributed-lag estimates

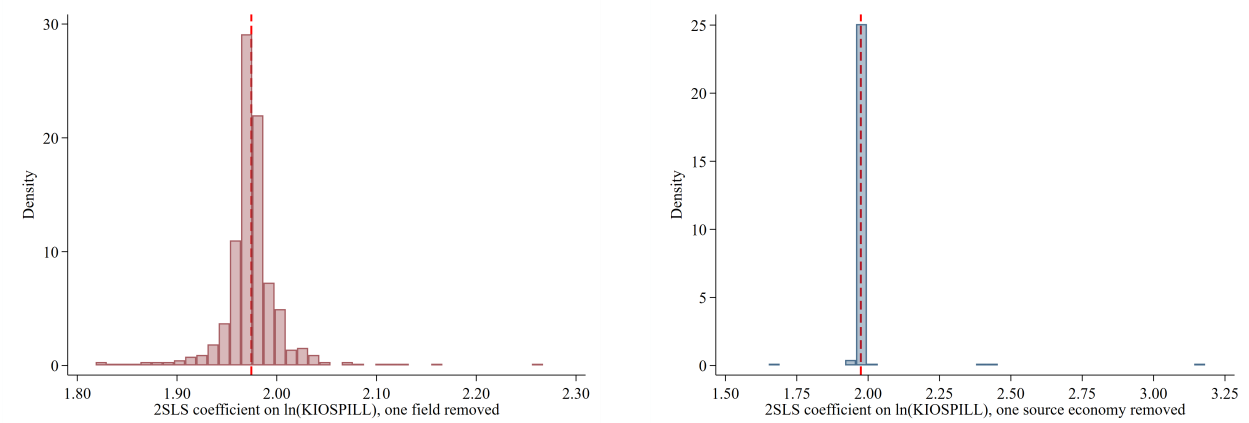
(b) Cumulative effect

Figure D1: Distributed-Lag Estimates and the Cumulative Response

Notes: The figure reports OLS estimates of the unrestricted distributed-lag counterpart of equation (7),

$$\ln patnum_{it} = \alpha + \sum_{s=1}^{5} \beta_s \ln KIOSPILL_{i,t-s} + \gamma' X_{i,t-1} + \mu_i + \delta_t + \varepsilon_{it},$$

on the 9,792 field-year observations over 2000–2016 for which all five lags are available. Panel (a) plots $\hat{\beta}_1$ to $\hat{\beta}_5$ from this single regression, and panel (b) the cumulative response $\sum_{s=1}^{S} \hat{\beta}_s$ for $S = 1, \ldots, 5$, so its horizontal axis is the number of lags included. Dots denote point estimates and vertical bars 90% confidence intervals; the shaded band in panel (b) connects these intervals. $X_{i,t-1}$ contains the baseline controls of Table 2 and production-network spillovers, and the regression includes technology-field and year fixed effects. Standard errors are clustered by technology field. The estimates are associations: because the instrument is a permanent step, lagged instruments are nearly collinear after fixed effects and controls, and an instrumented distributed lag is not identified (Kleibergen–Paap rank F statistic below one).



(a) One technology field removed

(b) One source economy removed

Figure D2: Leave-One-Out Estimates of the Baseline Effect

Notes: The figure plots the distribution of 2SLS estimates of the coefficient on ln *KIOSPILL* when the baseline specification in column (7) of Table 3 is re-estimated leaving out one unit at a time: one technology field in panel (a), giving 576 estimates, and one source economy, removed from both embedded technology and the instrument, in panel (b), giving 192 estimates. Bars are densities, and the vertical dashed line marks the baseline estimate of 1.974. Estimates range from 1.818 to 2.267 in panel (a) and from 1.652 to 3.180 in panel (b), and all are significant at the 5% level. Each regression includes the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects, with standard errors clustered by technology field.

Table D1: Instrumental Variable Robustness

Panel A. Alternative instruments and outliers	(1) Any positive exposure ln *patnum*	(2) Weighted shock ln *patnum*	(3) Winsorize y,x ln *patnum*	(4) Winsorize y,x,z ln *patnum*	(5) Trim 1.95 s.d. ln *patnum*
ln *KIOSPILL*	1.855**	2.006***	0.727**	0.766**	2.629***
	(0.732)	(0.764)	(0.303)	(0.302)	(0.679)
Observations	12,096	12,096	12,096	12,096	11,411
KP-rk LM statistic	18.247	8.937	23.700	23.479	11.597
KP-rk F statistic	18.780	14.598	35.705	35.820	18.545
First stage: instrument	0.061***	0.052***	0.101***	0.106***	0.045***
	(0.014)	(0.014)	(0.017)	(0.018)	(0.010)
First-stage R^2	0.944	0.944	0.933	0.933	0.966
Panel B. Inference and domestic policy controls	(6) Cluster by IPC3 ln *patnum*	(7) Innovation policy ln *patnum*	(8) Development zones ln *patnum*	(9) Higher-education expansion ln *patnum*	(10) All domestic controls ln *patnum*
ln *KIOSPILL*	1.974**	1.980***	1.932***	2.236***	2.207***
	(0.904)	(0.728)	(0.745)	(0.762)	(0.784)
Observations	12,096	12,096	12,096	12,096	12,096
KP-rk LM statistic	4.490	10.616	9.823	11.004	9.993
KP-rk F statistic	5.956	15.312	14.120	17.023	15.546
AR 95% confidence set	[0.62, 9.84]				
First stage: instrument	0.055**	0.055***	0.054***	0.060***	0.058***
	(0.023)	(0.014)	(0.014)	(0.014)	(0.015)
First-stage R^2	0.944	0.944	0.945	0.944	0.945

Notes: Each column changes one element of the baseline 2SLS specification in column (7) of Table 3. The dependent variable is $\ln(1+patnum)$ and the endogenous regressor is $\ln KIOSPILL$, both defined as in Table 2. The lower rows of each panel report the first-stage coefficient on the instrument constructed as named in the column heading. Column (1) defines treated fields as those with any positive share of liberalized industries rather than a share above the median. Column (2) replaces $Treat_j \times Post_t$ with the weighted shock $\ln(FDI_{c,1999\text{–}2001} \times Treat_j \times Post_t + 1)$. Columns (3) and (4) winsorize at the 5th and 95th percentiles, first the dependent and endogenous variables and then also the instrument. Column (5) trims observations whose baseline residual exceeds 1.95 standard deviations. Column (6) keeps the baseline specification and clusters standard errors by the 120 IPC3 classes instead of the 576 IPC4 fields. Columns (7)–(10) add domestic policy controls: R&D super-deduction intensity, HNTE tax relief, and government subsidies in column (7); development-zone intensity in column (8); the 1999 higher-education expansion in column (9); and all of them in column (10). All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field except in column (6). For column (6), whose first stage is weak, the table reports the 95% Anderson–Rubin confidence set, which is valid whatever the strength of the instrument; the Anderson–Rubin p-value at a coefficient of zero is below 0.01. The winsorized specifications, which have the strongest first stages, yield the smallest second-stage estimates, consistent with the sensitivity of the 2SLS magnitude to the scale of the first stage. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table D2: Confounding Investment Motives: Tax-Haven Sources

	(1) List 1 ln *patnum*	(2) List 2 ln *patnum*	(3) List 3 ln *patnum*	(4) List 4 ln *patnum*	(5) List 5 ln *patnum*
Panel A. 2SLS					
ln *KIOSPILL*,	2.889**	2.237**	2.889**	2.870**	3.501
non-haven sources	(1.446)	(0.878)	(1.446)	(1.437)	(2.194)
ln *KIOSPILL*,	−0.418	−0.113*	−0.417	−0.413	−0.467
haven sources	(0.278)	(0.060)	(0.278)	(0.277)	(0.353)
Observations	12,096	12,096	12,096	12,096	12,096
KP-rk LM statistic	4.679	9.101	4.681	4.701	2.498
KP-rk F statistic	5.754	12.240	5.757	5.791	3.007
AR 95% confidence set	[1.01, 17.06]	[0.92, 5.58]	[1.01, 17.06]	[1.00, 16.66]	unbounded
Panel B. OLS					
ln *KIOSPILL*,	0.137***	0.150***	0.137***	0.137***	0.124***
non-haven sources	(0.030)	(0.031)	(0.030)	(0.030)	(0.028)
ln *KIOSPILL*,	0.109***	0.024	0.109***	0.112***	0.055
haven sources	(0.030)	(0.016)	(0.030)	(0.029)	(0.041)
R^2	0.923	0.923	0.923	0.923	0.923
Observations	12,096	12,096	12,096	12,096	12,096

Notes: The dependent variable is $\ln(1 + patnum)$, defined as in Table 2. ln *KIOSPILL* is split by whether the source economy is a tax haven under each of the five lists in Table A5. In Panel A, the non-haven component is instrumented by the corresponding non-haven instrument, and the haven component enters as a control. Hong Kong is a haven in Lists 1, 3, and 4 but not in Lists 2 and 5. Lists 1 and 3 differ only in small jurisdictions with little investment in China, so their estimates nearly coincide. Because several first stages are weak, the table reports 95% Anderson–Rubin confidence sets, which are valid regardless of instrument strength. In column (5), the set is unbounded and disconnected but excludes zero (Anderson–Rubin p-value 0.004), so that column indicates which component carries the effect rather than its magnitude. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table D3: Alternative Constructions of FDI-Embedded Technology

	(1) FDI allocation: full assignment ln *patnum*	(2) Depreciation rate 15% ln *patnum*	(3) Citation window 10 years ln *patnum*	(4) Knowledge stock: patent counts ln *patnum*
Panel A. 2SLS				
ln *KIOSPILL*	1.252***	2.132***	2.081***	2.023**
	(0.447)	(0.790)	(0.761)	(0.826)
Observations	12,096	12,096	12,096	12,096
KP-rk LM statistic	20.541	10.983	10.433	7.791
KP-rk F statistic	27.127	13.683	13.690	12.483
Panel B. OLS				
ln *KIOSPILL*	0.083***	0.145***	0.152***	0.159***
	(0.032)	(0.032)	(0.033)	(0.034)
R^2	0.923	0.923	0.923	0.923
Observations	12,096	12,096	12,096	12,096

Notes: The dependent variable is $\ln(1 + patnum)$, defined as in Table 2. Each column changes one element of ln *KIOSPILL*, holding the rest at baseline. Column (1) assigns each CIC4 industry's full investment to every mapped field rather than dividing it equally; column (2) raises the depreciation rate of knowledge stocks from 5% to 15%; column (3) extends the citation window from five to ten years; and column (4) builds knowledge stocks from patent counts rather than citation-weighted patents. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table D4: Alternative Measures of Innovation Output

	(1) Patent count (IHS) IHS $patnum$	(2) Forward citations $\ln fwdcite$	(3) Patents granted abroad $\ln patnum^{abroad}$
Panel A. 2SLS			
$\ln KIOSPILL$	1.397**	3.577***	5.509***
	(0.679)	(1.199)	(1.601)
Observations	12,096	12,096	12,096
KP-rk LM statistic	10.524	10.524	10.524
KP-rk F statistic	15.124	15.124	15.124
Panel B. OLS			
$\ln KIOSPILL$	0.160***	0.156***	0.107***
	(0.033)	(0.041)	(0.026)
R^2	0.921	0.896	0.805
Observations	12,096	12,096	12,096

Notes: The table replaces the dependent variable, holding the specification fixed. In column (1), it is the inverse hyperbolic sine of the patent count. In column (2), it is the log of one plus the forward citations received by the field's patents, which weights output by subsequent use. In column (3), it is the log of one plus the number of the field's patent families granted by offices other than CNIPA, which restricts attention to inventions whose owners bore the cost of protecting them abroad. $\ln KIOSPILL$ is defined as in Table 2. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table D5: Alternative Industry–Technology Concordances

	(1) CNIPA, manufacturing ln *patnum*	(2) HS route, manufacturing ln *patnum*	(3) Reverse mapping, estimated in CIC4 ln *patnum*	(4) IPC3 aggregation ln *patnum*
Panel A. 2SLS				
ln *KIOSPILL*	2.403**	2.730***	1.835***	0.670**
	(1.029)	(0.959)	(0.362)	(0.329)
Observations	12,096	11,739	12,411	2,499
KP-rk LM statistic	11.083	12.511	25.801	12.658
KP-rk F statistic	10.595	10.468	32.510	16.810
Panel B. OLS				
ln *KIOSPILL*	0.070***	0.122***	0.099***	0.174***
	(0.025)	(0.037)	(0.015)	(0.063)
R^2	0.922	0.925	0.969	0.972
Observations	12,096	11,739	12,411	2,499

Notes: The dependent variable is $\ln(1+patnum)$ and the main explanatory variable is $\ln KIOSPILL$, both defined as in Table 2. All columns restrict the analysis to manufacturing, which the HS-based route covers, so that the comparison reflects the concordance rather than the sample. Column (1) applies the baseline CNIPA concordance, and column (2) routes CIC4 industries to HS2002 products and then to IPC4 fields using the probabilistic weights of Lybbert and Zolas (2014). Column (3) reverses the mapping, carrying patents and the knowledge network into CIC4 space and estimating there; because production-network spillovers cannot be constructed consistently at the CIC4 level, this column omits that control. Column (4) aggregates the analysis to the IPC3 level. All columns include the baseline controls of Table 2 and fixed effects for the unit of analysis and year. Standard errors, in parentheses, are clustered at the unit of analysis. ***, ** and * denote significance at the 1%, 5% and 10% levels.

E The OLS–IV Gap

Table E1: Component-Wise Attenuation Diagnostics for the Composite Regressor

	(1) OLS	(2)	(3)	(4)	(5)	(6)	(7)
		2SLS, instrument varies one component					
		Benchmark IV^{KIO}	FDI allocation	Depreciation 15%	Citation window	Patent counts	Network global
	$\ln patnum$	$\ln patnum$	$\ln patnum$	$\ln patnum$	$\ln patnum$	$\ln patnum$	$\ln patnum$
$\ln KIOSPILL$	0.153***	1.974***	0.100***	0.144***	0.153***	0.167***	0.157***
	(0.031)	(0.733)	(0.036)	(0.032)	(0.033)	(0.035)	(0.034)
Observations	12,096	12,096	12,096	12,096	12,096	12,096	12,075
Reliability ratio λ			0.830	0.990	1.001	0.889	0.850
Disattenuated OLS			0.185	0.155	0.153	0.172	0.180
KP-rk LM statistic		10.524	265.371	279.327	280.776	285.471	292.423
KP-rk F statistic		15.124	12,629	290,092	305,444	14,763	6,031

Notes: The table asks whether classical measurement error in the composite regressor can explain the gap between the OLS estimate in column (1) and the policy-instrumented estimate in column (2). $\ln KIOSPILL$ combines three objects measured with error: the allocation of industry-level investment across fields, source-economy knowledge stocks, and citation weights. Under classical error, an independently constructed measure of the same variable is a valid instrument for it, so if attenuation explained the gap, instrumenting the baseline measure with an alternative construction should move the estimate toward column (2). Columns (3)–(7) vary one component at a time: the FDI allocation rule, the depreciation rate, the citation window, knowledge stocks built from patent counts, and a global rather than U.S. citation network. The estimates range from 0.100 to 0.167, close to the OLS estimate and an order of magnitude below 1.974. λ is the implied within-field reliability ratio $\mathrm{Cov}(x_1, x_k)/\mathrm{Var}(x_1)$ after absorbing fixed effects and controls, and the disattenuated coefficient is the OLS estimate divided by λ. Reconciling columns (1) and (2) through classical attenuation alone would require $\lambda = 0.078$, that is, 92.3% of the within-field variation being noise. The exercise bounds the attenuation from one component at a time but does not remove error common to the two constructions; λ above one in column (5) indicates that the error is not purely classical, and columns (3)–(7) are not causal estimates. Their Kleibergen–Paap statistics are very large because the first-stage coefficient, not reported, is close to one; this reflects agreement between two constructions of the same variable and is not comparable with the weak-identification diagnostic in column (2). All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table E2: Source Economies Carrying the Policy-Induced FDI

Economy	Share of FDI (%)	Share of policy-induced growth (%)
Hong Kong, China	48.41	1.08
Taiwan, China	8.74	1.86
South Korea	7.72	1.55
British Virgin Islands	5.98	1.81
United States	4.99	1.55
Japan	3.81	1.34
Samoa	2.54	2.32
Singapore	2.50	1.69
Canada	1.22	1.89
Germany	1.20	1.43
All other economies	12.89	83.49

Notes: The unit is a source economy × IPC4 field cell. The first column reports each economy's share of the level of FDI. The second reports its share of policy-induced log growth, measured as the within-economy difference-in-differences in mean log investment between 1996–2001 and 2003–2015, that is, the change in liberalized fields minus the same economy's average change in non-liberalized fields, truncated below at zero. This is an empirical proxy for complier intensity, not an identification of latent compliance types. Because shares of log growth do not weight economies by size, the comparison concerns each economy's role relative to its share of the level of FDI. Hong Kong and the British Virgin Islands, whose prominence in the level of FDI reflects round-tripping and conduit investment, account for much smaller shares of policy-induced growth. Table D2 reports the corresponding decomposition by tax-haven status.

Table E3: Characteristics of the Policy-Induced FDI Response, by Margin

	(1)	(2)	(3)	(4)
	H: knowledge stock		H: PCT lead over China	
	Intensive $\Delta \ln FDI$	Extensive $1(FDI > 0)$	Intensive $\Delta \ln FDI$	Extensive $1(FDI > 0)$
H	0.071**	0.047***	−0.010	−0.010
	(0.028)	(0.008)	(0.006)	(0.015)
$H \times Treat$	0.007	0.038***	0.005	0.194***
	(0.025)	(0.010)	(0.007)	(0.046)
R^2	0.452	0.347	0.424	0.345
Observations	20,267	100,494	18,428	69,747

Notes: The unit is a source economy × technology field cell. In columns (1) and (3), the intensive margin, the dependent variable is the change in mean $\ln(FDI)$ between 1996–2001 and 2003–2015, and the sample is cells with positive pre-reform investment. In columns (2) and (4), the extensive margin, the dependent variable indicates positive investment after 2002, and the sample is cells with no pre-reform investment. H is the source economy's standardized 2001 knowledge stock in the field in columns (1) and (2), and its standardized PCT lead over China in the field in columns (3) and (4). *Treat* indicates a field liberalized in 2002; its main effect is absorbed by the field fixed effects. All columns include source-economy and technology-field fixed effects. Standard errors, in parentheses, are clustered by technology field. New entrants account for most affected cells but a small share of policy-induced investment by value; the results therefore show that liberalization tilted new entry toward technologically stronger sources, not that it selectively attracted high-technology investment overall. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table E4: The OLS Association by Differencing Horizon

	(1) $k=2$ $\Delta_k \ln patnum$	(2) $k=4$ $\Delta_k \ln patnum$	(3) $k=6$ $\Delta_k \ln patnum$	(4) $k=8$ $\Delta_k \ln patnum$
$\Delta_k \ln KIOSPILL$	-0.005	0.042^{***}	0.084^{***}	0.195^{***}
	(0.013)	(0.015)	(0.023)	(0.037)
R^2	0.047	0.036	0.031	0.036
Observations	10,944	9,792	8,640	7,488

Notes: Each column regresses the k-year difference in $\ln(1+patnum)$ on the k-year difference in $\ln KIOSPILL$, with year fixed effects; differencing removes the field fixed effect. The estimates are OLS associations that describe how the relationship varies with the horizon over which the data are differenced. The slope is indistinguishable from zero at a two-year horizon and rises monotonically with k, consistent with a relationship that operates at low frequency. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table E5: The Two Estimators at Annual and Long-Difference Frequency

	Annual panel		Long difference	
	(1) OLS $\ln patnum$	(2) 2SLS $\ln patnum$	(3) OLS $\Delta \ln patnum$	(4) 2SLS $\Delta \ln patnum$
$\ln KIOSPILL$	0.153^{***}	1.974^{***}	0.350^{***}	1.013^{*}
	(0.031)	(0.733)	(0.119)	(0.553)
Observations	12,096	12,096	576	576
KP-rk LM statistic		10.524		12.851
KP-rk F statistic		15.124		14.610
IV/OLS ratio	12.9		2.90	

Notes: The dependent variable is $\ln(1+patnum)$ and the main explanatory variable is $\ln KIOSPILL$, both defined as in Table 2. Columns (1) and (2) use the annual panel with the baseline controls, production-network spillovers, and technology-field and year fixed effects; the OLS estimate is computed on the 2SLS estimating sample, and standard errors are clustered by technology field. Columns (3) and (4) use long differences between the 1996–2001 and 2003–2016 means for each of the 576 fields, instrumenting with the field's exposure to the policy-induced investment shock, with heteroskedasticity-robust standard errors. Because the instrument is a permanent step, its variation is almost entirely a low-frequency contrast, so comparing the annual 2SLS estimate with an annual within-field OLS association compares them at different frequencies. The long-difference 2SLS estimate is imprecise, so the narrowing of the IV/OLS ratio from 12.9 to 2.9 indicates magnitude rather than a precise decomposition. ***, ** and * denote significance at the 1%, 5% and 10% levels.

F Why Production-Network Spillovers Do Not Survive

The production-network estimate is large when considered on its own but becomes statistically insignificant once knowledge-network exposure is included. This appendix examines three possible explanations for this pattern: measurement error in the production network, collinearity between the two instruments, and cross-loading of the instruments on the two exposure measures.

F.1 An Alternative Production Network

The baseline production network is constructed from an industry input–output table. Such industry-level input shares may be a noisy proxy for the production relationships of individual firms (Barrios et al., 2011; Alfaro-Ureña et al., 2022). To assess whether this measurement error explains the disappearance of the production-network effect, we construct an alternative network from observed firm-to-firm transactions.

The data come from the Compustat Segment Customer files obtained through Wharton Research Data Services. These files report major customers disclosed by listed firms and the sales associated with each relationship. U.S. accounting rules require public firms to disclose customers accounting for at least 10% of their revenues, and customer names are linked to firm identifiers following the approach used in the supply-chain literature (Cohen and Frazzini, 2008). Industry codes are subsequently mapped to technology fields using the probabilistic NAICS 2002 concordance of Lybbert and Zolas (2014).

We construct the network in five steps. First, we retain supplier–customer links dated between 1992 and 2001, restrict the sample to corporate customers with positive reported sales, and obtain 28,696 links. Using only pre-reform relationships ensures that the network weights cannot respond to the 2002 policy shock. Second, suppliers are assigned four-digit NAICS codes from the Segment Customer files. Third, customers are assigned four-digit NAICS codes using, in order of preference, bilateral matches in the Segment Customer files, Historical Segment Fundamentals, and the global Compustat firm file. The resulting customer match rate is 85.2%. Fourth, reported sales are

aggregated to supplier-industry × customer-industry cells and normalized by total disclosed purchases of the customer industry, yielding buyer-side linkage shares across 2,992 NAICS industry pairs. Fifth, both sides of each pair are mapped to IPC4 technology fields, and the linkage share is multiplied by the corresponding concordance probabilities and summed across industry pairs. This is the same aggregation procedure used to map the benchmark input–output table into technology space. The resulting transaction-based network contains 195,384 technology-field pairs, compared with 306,427 pairs in the baseline production network.

The transaction-based network has different measurement errors from the input–output measure. It covers U.S.-listed firms rather than Chinese producers; disclosure is limited to major customers accounting for at least 10% of supplier revenue; listed firms are disproportionately concentrated in manufacturing; and reported sales measure the value of commercial relationships rather than the technological content of the underlying inputs. The alternative network should therefore not be interpreted as a more accurate measure of production linkages. Its value for this exercise is precisely that its measurement errors differ from those of the benchmark input–output network.

Table F1 repeats the horse race using this transaction-based network. The same qualitative pattern emerges: production-network exposure generates a positive estimate when considered separately, but its independent effect weakens once knowledge-network exposure is included. The disappearance of the production-network effect is therefore unlikely to be driven solely by measurement error in the benchmark input–output weights.

F.2 Instrument Correlation

A second possibility is that the production-network effect becomes imprecisely estimated simply because the KIO and PIO instruments are highly collinear. Table F2 examines this possibility. The raw correlation between the two instruments is high, reflecting the fact that both propagate the same policy-induced investment shock and therefore share substantial common time-series variation. Once we residualize the instruments on the fixed effects and baseline controls used in the estimating equation, however, the correlation falls substantially. This indicates that simple

collinearity in their common time structure is unlikely to be the main reason why the production-network coefficient loses significance in the joint specification.

F.3 First-Stage Cross-Loadings

Table F3 reports the matrix of first-stage cross-loadings, obtained by regressing each exposure measure separately on each instrument while including the same fixed effects and controls as in the baseline specification. The pattern is asymmetric. The KIO instrument strongly predicts knowledge-network exposure but has little predictive power for production-network exposure. The PIO instrument, by contrast, predicts both production-network and knowledge-network exposure. The PIO-only IV specification therefore does not generate variation that is specific to the production channel: the same identifying variation that raises production-network exposure also shifts knowledge-network exposure.

This asymmetry helps explain why the PIO-only IV estimate can be large even though the PIO coefficient becomes statistically insignificant once both exposures are instrumented jointly. A regression that instruments production exposure alone can attribute to production linkages an innovation response associated with the knowledge exposure that the same instrument also shifts. Once the two exposures are allowed to enter jointly, the variation specific to knowledge-network exposure continues to predict patenting, whereas we find no statistically precise independent production-network effect.

Taken together, these results indicate that the large production-network estimate obtained in isolation reflects the knowledge exposure that production-linked shocks also shift, rather than measurement error or collinearity.

Table F1: An Alternative Production Network Built from Firm-to-Firm Transactions

	(1) Input–output network alone ln *patnum*	(2) Transaction network alone ln *patnum*	(3) Knowledge network alone ln *patnum*	(4) Input–output + knowledge ln *patnum*	(5) Transaction + knowledge ln *patnum*
Panel A. 2SLS					
ln *KIOSPILL*			2.000***	2.051***	1.571**
			(0.756)	(0.786)	(0.723)
ln *PIOSPILL*	2.107***	0.988***		0.741	0.445
	(0.673)	(0.284)		(0.736)	(0.372)
Observations	12,096	11,571	12,096	12,096	11,571
KP-rk LM statistic	18.828	52.151	8.772	11.446	8.022
KP-rk F statistic	23.423	25.873	12.460	8.186	5.946
AR 95% set, ln *KIOSPILL*			[0.79, 4.80]	[0.58, 5.74]	[0.22, 5.87]
AR 95% set, ln *PIOSPILL*	[0.96, 3.99]	[0.48, 1.70]		[−1.56, 3.04]	[−1.18, 1.38]
Panel B. OLS					
ln *KIOSPILL*			0.172***	0.153***	0.158***
			(0.035)	(0.031)	(0.034)
ln *PIOSPILL*	0.131***	0.142***		0.093**	0.116***
	(0.048)	(0.035)		(0.044)	(0.033)
R^2	0.922	0.924	0.923	0.923	0.924
Observations	12,096	11,571	12,096	12,096	11,571

Notes: The dependent variable is $\ln(1 + patnum)$, defined as in Table 2. ln *PIOSPILL* is built from the 1997 BEA benchmark use table in columns (1) and (4) and from pre-reform firm-to-firm transactions in columns (2) and (5). The transaction network uses Compustat Segment Customer disclosures over 1992–2001, linked to supplier and customer industries as described in Appendix F, and its measurement errors differ from those of the input–output table. If attenuation from mismeasured input–output shares explained the pattern in Table 3, the transaction-based measure should retain an independent effect alongside ln *KIOSPILL*; it does not (column 5). Anderson–Rubin confidence sets are valid regardless of instrument strength; with two instrumented regressors, each is the projection of the joint set and is therefore conservative. All columns include the baseline controls of Table 2 and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table F2: Correlation between the Knowledge- and Production-Network Instruments

	Pearson correlation	Observations
Panel A. Levels, as constructed		
All field–years	0.9745	12,096
Panel B. Levels, by period		
Pre-reform years, $t \leq 2002$	/	4,032
Post-reform years, $t > 2002$	0.3781	8,064
Panel C. Residualized: the variation the estimator uses		
Field fixed effects only (variation over time)	0.9913	12,096
Year fixed effects only (variation across fields)	0.3739	12,096
Field and year fixed effects	0.3739	12,096
Field and year fixed effects plus controls	0.2847	12,096

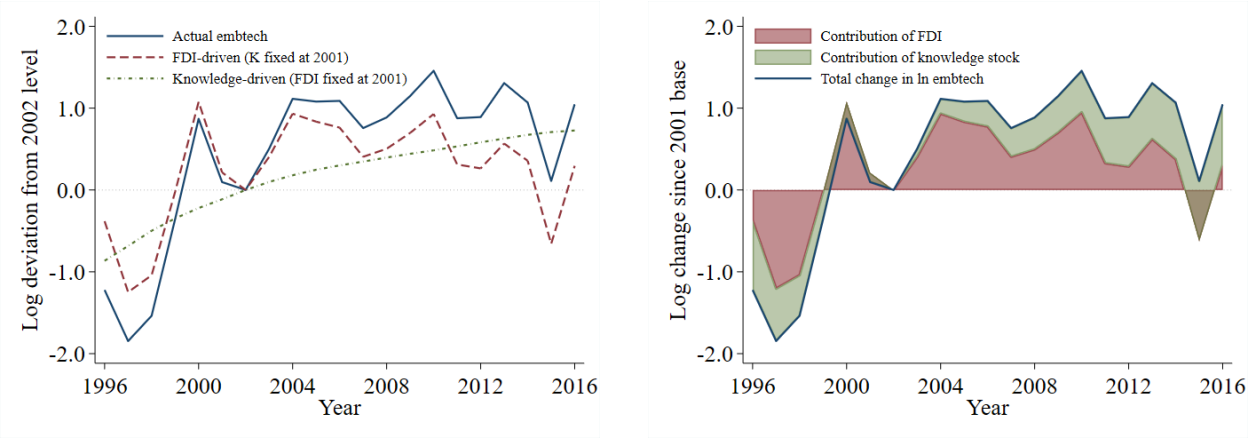
Notes: The table reports the Pearson correlation between the knowledge-network instrument $\ln IV^{KIO}$ of equation (5) and the production-network instrument $\ln IV^{PIO}$ of equation (6), in raw form and after absorbing the fixed effects and controls of the estimating equation. Panel A shows that the raw correlation is high. Panel B shows why: both instruments are zero for every field before 2002 by construction, so the pre-reform correlation is undefined, and the post-reform correlation is 0.378. The raw figure is therefore dominated by the common zero block rather than by any similarity between the two networks. Panel C strips the two dimensions in turn. Removing only field effects leaves the common 2002 step, which the two instruments share almost exactly. Removing only year effects leaves the cross-field variation, where the two networks genuinely differ. The last line is the correlation of the variation the two-stage estimator actually uses.

Table F3: First-Stage Cross-Loadings of the Two Instruments

	(1) $\ln KIOSPILL$	(2) $\ln PIOSPILL$	(3) $\ln KIOSPILL$	(4) $\ln PIOSPILL$	(5) $\ln KIOSPILL$	(6) $\ln PIOSPILL$
$\ln IV^{KIO}$	0.054^{***}	-0.004			0.046^{***}	-0.023
	(0.015)	(0.014)			(0.017)	(0.015)
$\ln IV^{PIO}$			0.058^{***}	0.088^{***}	0.040^{**}	0.097^{***}
			(0.016)	(0.018)	(0.017)	(0.019)
R^2	0.005	0.000	0.003	0.009	0.007	0.010
Observations	12,096	12,096	12,096	12,096	12,096	12,096

Notes: Each column regresses an exposure measure on one or both instruments, after residualizing all variables on the fixed effects and controls of the estimating equation, so that the coefficients are first-stage loadings. The dependent variable is $\ln KIOSPILL$ in odd columns and $\ln PIOSPILL$ in even columns. Columns (1) and (2) include only the knowledge-network instrument, columns (3) and (4) only the production-network instrument, and columns (5) and (6) both. The ratio of the loadings in columns (3) and (4), 0.67 before rounding, is the share of the knowledge-network effect that a production-only IV specification attributes to the production channel. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

G Additional Results



(a) Evolution of ln *KIOSPILL* and its components

(b) Contributions to the change

Figure G1: Decomposing the Sources of Variation in Embedded Technology

Notes: The figure decomposes ln *KIOSPILL* into an investment-driven component, which holds source-economy knowledge stocks at their 2001 values, and a knowledge-driven component, which holds investment at its 2001 values. Panel (a) plots the cross-field averages of the actual measure (solid line) and of the investment-driven (dashed line) and knowledge-driven (dash-dotted line) components. Panel (b) plots each component's contribution to the average log change since 2001 (shaded areas) and their total (solid line). In the raw time path, the knowledge-driven component contributes more, because knowledge stocks grow smoothly and largely in parallel across fields. In the within-field variation used in the regressions, after field and year fixed effects absorb this common growth, the investment-driven component tracks the actual measure almost exactly. The decomposition is descriptive; it shows that the variation used in estimation, and shifted by the instrument, is driven by investment.

Table G1: Cross-Border Backward Citations at the Source Economy × Field Level

	(1) Baseline *backcite*	(2) Excluding FIE citing families *backcite*	(3) Excluding FIE cited families *backcite*	(4) Excluding both *backcite*
Panel A. 2SLS				
ln *KIOSPILL*	5.353***	5.118***	5.351***	5.117***
	(0.492)	(0.472)	(0.492)	(0.471)
Observations	2,322,432	2,322,432	2,322,432	2,322,432
KP-rk LM statistic	126.147	126.147	126.147	126.147
KP-rk F statistic	153.269	153.269	153.269	153.269
Panel B. Log OLS				
ln *KIOSPILL*	0.0038***	0.0036***	0.0038***	0.0036***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
R^2	0.589	0.581	0.589	0.581
Observations	2,322,432	2,322,432	2,322,432	2,322,432
Panel C. PPML, raw counts				
ln *KIOSPILL*	0.007**	0.008**	0.007**	0.008**
	(0.004)	(0.003)	(0.004)	(0.003)
Observations	439,341	431,592	439,257	431,550
Panel D. PPML, counts winsorized at p99				
ln *KIOSPILL*	0.017***	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	439,341	431,592	439,257	431,550
Panel E. Extensive margin				
ln *KIOSPILL*	0.0021***	0.0020***	0.0021***	0.0020***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
R^2	0.479	0.473	0.479	0.473
Observations	2,322,432	2,322,432	2,322,432	2,322,432

Notes: The table repeats Table 6 at the source economy × technology field × year level for 192 source economies. The dependent variable is the number of citations from Chinese patents in the field to patents from the source economy, entering as $\ln(1+\cdot)$ in Panels A and B, as a raw count in Panel C, as a count winsorized at the 99th percentile in Panel D, and as an indicator for a positive count in Panel E. ln *KIOSPILL* is measured at the source economy × field level and instrumented in Panel A. Columns (2), (3), and (4) exclude links involving FIE-affiliated patent families on the citing side, the cited side, and both sides, respectively. All columns include source economy × field and year fixed effects, the baseline controls of Table 2, and production-network spillovers. Poisson samples are smaller because cells with no positive outcome in any year are dropped. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table G2: Speed of Cross-Border Knowledge Diffusion

	(1) $\ln(1+citationlag)$	(2) $\ln(1+backcite)$
Panel A. 2SLS		
$\ln KIOSPILL$	0.337^{***}	4.445^{***}
	(0.110)	(1.307)
$\ln(1+citationlag)$		-0.240
		(0.234)
Observations	12,093	12,093
KP-rk LM statistic	10.606	10.877
KP-rk F statistic	15.243	15.549
Panel B. OLS		
$\ln KIOSPILL$	-0.001	0.169^{***}
	(0.005)	(0.041)
$\ln(1+citationlag)$		-0.273^{**}
		(0.122)
R^2	0.943	0.866
Observations	12,093	12,093
Panel C. PPML, counts winsorized at p99		
$\ln KIOSPILL$	-0.001	0.074^{**}
	(0.005)	(0.034)
$\ln(1+citationlag)$		-1.577^{***}
		(0.312)
Observations	12,093	12,031

Notes: In column (1), the dependent variable is the log of one plus the mean lag, in years, between the filing of a foreign patent and its first citation by a Chinese patent in the field (Hegde et al., 2023); a negative coefficient would indicate faster diffusion. In column (2), the dependent variable is $\ln(1+backcite)$, and the citation lag is added as a control. $\ln KIOSPILL$ is defined as in Table 2. The OLS and Poisson estimates in column (1) are negligible and insignificant, and the 2SLS estimate is positive, the opposite of the sign that faster diffusion would imply. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table G3: Wholly Foreign-Owned versus Joint-Venture Sources of Embedded Technology

	(1) WFOE *collaborate*	(2) JV *collaborate*	(3) Both *collaborate*
Panel A. 2SLS			
ln *KIOSPILL*, WFOE	4.423***		2.483***
	(0.835)		(0.589)
ln *KIOSPILL*, JV		1.913***	1.680***
		(0.354)	(0.394)
Observations	12,096	12,096	12,096
KP-rk LM statistic	32.653	34.547	26.631
KP-rk F statistic	38.401	48.978	18.890
Panel B. Log OLS			
ln *KIOSPILL*, WFOE	0.067***		0.047**
	(0.023)		(0.020)
ln *KIOSPILL*, JV		0.054***	0.044**
		(0.020)	(0.019)
R^2	0.648	0.649	0.649
Observations	12,096	12,096	12,096
Panel C. PPML, counts winsorized at p99			
ln *KIOSPILL*, WFOE	0.073		0.082
	(0.077)		(0.075)
ln *KIOSPILL*, JV		−0.038	−0.044
		(0.035)	(0.034)
Observations	9,954	9,954	9,954
Panel D. Extensive margin			
ln *KIOSPILL*, WFOE	0.040***		0.034***
	(0.011)		(0.010)
ln *KIOSPILL*, JV		0.021***	0.014*
		(0.008)	(0.008)
R^2	0.519	0.518	0.519
Observations	12,096	12,096	12,096

Notes: The dependent variable is the number of patents co-applied by an FIE and a domestic entity, entering as $\ln(1+\cdot)$ in Panels A and B, as a count winsorized at the 99th percentile in Panel C, and as an indicator for a positive count in Panel D. ln *KIOSPILL* is split by the ownership form of the source FIE: a firm with no domestic shareholder is classified as wholly foreign owned (WFOE) and one with a domestic shareholder as a joint venture (JV). Columns (1) and (2) enter each component alone and column (3) both together. Because transfer requirements applied to JV entry but not to WFOEs, a *quid pro quo* account predicts a larger JV coefficient. Embedded technology is instrumented in Panel A and entered directly elsewhere. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table G4: Collaborative Innovation and the *Quid Pro Quo* Hypothesis

	(1) Foreign office *collaborate*	(2) US, European or Japanese office *collaborate*	(3) Home-country office *collaborate*
Panel A. 2SLS			
ln *KIOSPILL*	3.329***	3.217***	2.213***
	(1.035)	(1.001)	(0.735)
Observations	12,096	12,096	12,096
KP-rk LM statistic	10.524	10.524	10.524
KP-rk F statistic	15.124	15.124	15.124
Panel B. Log OLS			
ln *KIOSPILL*	0.046***	0.042**	0.023
	(0.017)	(0.017)	(0.018)
R^2	0.567	0.561	0.502
Observations	12,096	12,096	12,096
Panel C. PPML, counts winsorized at p99			
ln *KIOSPILL*	−0.203*	−0.275**	0.166**
	(0.107)	(0.115)	(0.070)
Observations	6,405	6,174	8,442
Panel D. Extensive margin			
ln *KIOSPILL*	0.027***	0.024***	0.026***
	(0.008)	(0.008)	(0.009)
R^2	0.450	0.444	0.422
Observations	12,096	12,096	12,096

Notes: The dependent variable is the number of patents co-applied by an FIE and a domestic entity, restricted to families with a member filed at a patent office other than CNIPA (column 1; 12,137 families), at the USPTO, EPO, or JPO (column 2; 11,364 families), or at an office of one of the FIE's foreign investors (column 3). Column (3) counts the EPO for investors from its member states and CNIPA for investors from Hong Kong, Macao, and Taiwan, and is inclusive when a firm has investors from several jurisdictions. A collaboration undertaken only to satisfy a domestic requirement would have little reason to seek protection abroad. Family membership records applications rather than grants. The outcome enters as $\ln(1+\cdot)$ in Panels A and B, as a count winsorized at the 99th percentile in Panel C, and as an indicator for a positive count in Panel D. ln *KIOSPILL* is instrumented in Panel A and entered directly elsewhere. All columns include the baseline controls of Table 2, production-network spillovers, and technology-field and year fixed effects. Standard errors, in parentheses, are clustered by the technology field. ***, ** and * denote significance at the 1%, 5% and 10% levels.

Table G5: Most and Least Central Technology Fields, IPC4 Level, 1996–2016 Window

IPC4	Description
Panel A. Five most central fields	
G06F	Electric digital data processing
A61B	Diagnosis; surgery; identification
H01L	Semiconductor devices; electric solid-state devices
A61K	Preparations for medical, dental or toilet purposes
H04L	Transmission of digital information
Panel B. Five least central fields	
F24T	Geothermal collectors; geothermal systems
B41G	Apparatus for bronze printing or line printing
D06G	Mechanical or pressure cleaning of carpets, hides and textile articles
C10F	Drying or working-up of peat
D04G	Knotting of filamentary material; knotted nets and carpets

Notes: The table lists the five most and five least central IPC4 fields in the innovation network computed over 1996–2016. Centrality is the left eigenvector of the row-normalized citation matrix associated with the unit eigenvalue (Liu and Ma, 2021), so a field is central when it is drawn on by fields that are themselves central. Panel A orders fields from most to least central, and Panel B from least to most central. IPC residual subclasses for subject matter not otherwise provided for are excluded. Table G6 repeats the ranking for the 1992–2001 window.

Table G6: Most and Least Central Technology Fields, IPC4 Level, 1992–2001 Window

IPC4	Description
Panel A. Five most central fields	
G06F	Electric digital data processing
A61B	Diagnosis; surgery; identification
H01L	Semiconductor devices; electric solid-state devices
H04L	Transmission of digital information
A61F	Filters implantable into blood vessels; prostheses
Panel B. Five least central fields	
B82Y	Specific uses or applications of nanostructures
F24S	Solar heat collectors; solar heat systems
B41D	Apparatus for the mechanical reproduction of printing surfaces
E02C	Ship-lifting devices
D04G	Knotting of filamentary material; knotted nets and carpets

Notes: The table repeats Table G5 for the 1992–2001 window used to construct the instrument. Four of the five most central fields coincide across the two windows, so the pre-reform network does not have an unrepresentative structure.