Global R&D, Local Knowledge: Evidence from Patent Text Similarity

Kose John[a], Kyeong Hun Lee[b], Zhou Lu[c], and Emma Q. Xu[d]

March 2026

Abstract

This paper examines knowledge spillovers between R&D laboratories within multinational firms. Using deep-learning-based patent text similarity, we document significant innovation differences between R&D labs located in a firm's home country and those abroad. These differences are larger when headquarters engages in AI-related innovation and when foreign labs operate in countries with weak IP protection, suggesting firms limit technology sharing to mitigate IP risk. Internal competition among divisional managers, who are incentivized to differentiate their innovation, further contributes to within-firm differences. The results are not driven by firms adapting innovation to local market needs, cultural distance between countries, or firms seeking foreign technologies unavailable in their home countries.

We thank Abraham Ravid and participants at the 2026 AEA annual meeting and the 2026 MMM conference for their constructive comments.

[a] Department of Finance, New York University, 44 West Fourth St. 9-98, New York, NY 10012. Email: kj1@stern.nyu.edu
[b] Department of Economics, Finance, and Legal Studies, University of Alabama, 361 Stadium Dr, Tuscaloosa, AL 35487. Email: klee102@ua.edu.
[c] School of Finance, Nankai University. Email: joelu@nankai.edu.cn
[d] Department of Finance, Insurance and Real Estate, University of Memphis, 3675 Central Ave., Memphis, TN 38152. Email: qxu2@memphis.edu.

1. Introduction

Global research and development (R&D) operations have expanded rapidly in recent decades. For example, in 2019, approximately 530 multinational firms operated R&D centers in Israel. By 2022, Shanghai alone hosted around 510 such centers established by foreign firms.[1] These operations can provide firms with strategic advantages, such as access to a broader talent pool that includes highly skilled scientists and engineers who may be scarce in their home countries. Moreover, foreign R&D may facilitate cross-border knowledge spillovers, enabling firm employees in home and host countries to integrate their expertise and skills in novel and complementary ways.

In this paper, we provide, to the best of our knowledge, the first comprehensive study of global R&D activities and the resulting knowledge spillovers within multinational firms. A central contribution of our study is that we develop a measure of within-firm knowledge spillover based on deep-learning-based textual analysis of a large, global sample of patents.[2] Moreover, we investigate several mechanisms through which knowledge diffuses across national borders within firms.

Using a sample of 6,681 firms from 39 countries and their global R&D labs over the period 1990–2014, we document significant differences in innovation between R&D labs located in a firm's headquarters country and those abroad.[3] This finding is surprising because, within a unified firm boundary, prior research shows that divisions' technologies tend to converge through technology transfers and inventor collaboration (e.g., Bloom, Sadun, and Van Reenen, 2012; Li and Wang, 2023). The effect is also economically meaningful: on average, patent similarity

[1] For the Israel case, see Startup Nation Central, *Forbes*, May 27, 2019 (link). For the Shanghai case, see *China Daily*, October 18, 2022 (link).
[2] Our main analysis relies on Sentence-Bidirectional Encoder Representations from Transformer (SBERT) for generating text embeddings (Reimers and Gurevych, 2019). For robustness, we additionally use PatentSBERTa, a version of SBERT fine-tuned on patent documents (Bekamiri, Hain, and Jurowetzki, 2024).
[3] Because our dependent variables are measured over a three-year lead period, they include innovation outcomes realized up to 2017.

between a firm's headquarters and its foreign R&D labs is 16% lower than the average similarity between patents produced within the headquarters location or between two foreign (non-headquarters) R&D labs.

While our findings are stronger for U.S.-headquartered firms, they hold across a global sample of multinational corporations. Moreover, the effects are particularly pronounced for firms whose headquarters focus on artificial intelligence (AI)-related innovation, a technological domain of growing strategic importance for firms and economies.

We examine several plausible, non-mutually exclusive channels underlying our results. First, we investigate whether firms are reluctant to share key technologies with foreign divisions when intellectual property (IP) risks are particularly pronounced in host countries. Host country governments may expropriate foreign firms' innovative outcomes, while local firms may exploit new technologies through various means, such as leaking sensitive information via local employees or engaging in unauthorized reverse engineering of foreign products and technologies. In such situations, courts in host countries often fail to treat foreign firms equitably, thereby undermining the firms' ability to profit from their innovation. Consequently, weak IP-rights protection can discourage firms from fully transferring core technologies to their foreign divisions, even when such transfers would be technologically beneficial.

To empirically test the IP risk channel, we use the signing of a bilateral investment treaty (BIT) as a proxy for stronger IP protection between country pairs. BITs safeguard cross-border investments and provide neutral arbitration mechanisms that mitigate the risk of unfair treatment by host-country institutions. We find that patent similarity between headquarters and foreign R&D laboratories increases significantly following BIT adoption, consistent with headquarters being more willing to share core technologies when IP risks decline. Although we cannot establish a

definitive causal relationship, BIT adoption is plausibly exogenous to individual firms, allowing us to provide suggestive evidence linking IP protection to within-firm knowledge spillovers.

In addition, IP risks may be harder to mitigate when foreign R&D laboratories are located in countries with low political alignment with the headquarters country. Consistent with this idea, our results are stronger when foreign laboratories are located in such countries.

Next, we examine competition among divisional managers as an additional channel. Internal competitive pressures may discourage cooperation across divisions and encourage managers to differentiate their innovation efforts to secure corporate resources and advance their internal career prospects. We proxy for internal R&D competition using two measures: the within-firm concentration of innovation output relative to human capital input, following Seru (2014), and the variation in tax-induced R&D competition across divisions. Consistent with this idea, patent dissimilarity between headquarters and foreign R&D laboratories is more pronounced when firms exhibit greater internal R&D competition among divisional managers.

Another potential explanation is that firms may adapt their innovation to local consumer preferences, leading foreign R&D laboratories to pursue different types of innovation. Such local customization of R&D often co-locates with manufacturing facilities (Fan, 2025). To test this channel, we examine foreign R&D segments located in countries where the firm has manufacturing capabilities and where local customization is more likely. However, we find no evidence that these foreign R&D laboratories engage in less similar innovation relative to headquarters. In contrast, our findings indicate that foreign R&D labs in manufacturing locations exhibit a greater degree of knowledge spillover with headquarters, suggesting that headquarters transfer their technologies to these manufacturing-intensive locations.

Further, we examine cultural differences between R&D laboratories as a potential driver of limited knowledge spillovers. When cultural gaps are large, employees in foreign laboratories and those at headquarters may face greater communication and collaboration barriers, potentially leading to less coherent innovation portfolios and greater patent dissimilarity. Although the effect is stronger for culturally distant countries, it is not statistically significant.

Additionally, firms may locate foreign R&D laboratories to access technologies that are not present at headquarters. If so, country-level technological differences could mechanically generate differences in innovation between headquarters and foreign laboratories. Our results are not driven by cross-country technological differences.

Our paper contributes to the literature on innovation and knowledge spillovers. The diffusion of innovative knowledge is critical for economic growth (e.g., Griliches, 1979), prompting extensive research on how such knowledge spreads. Jaffe (1986) shows that firms operating in technologically related fields tend to be more innovative. Jaffe, Trajtenberg, and Henderson (1993), Carrino and Kerr (2014), and Jiang, John, Lee, and Xu (2026) highlight the importance of geographic location and associated features—such as agglomeration economies—for facilitating knowledge spillovers. In addition, knowledge can diffuse across national borders through international trade and foreign direct investment (Teece, 1977; Coe and Helpman, 1995; Bloom et al., 2012).

Our study focuses on within-firm, cross-border knowledge spillovers. While prior studies often rely on patent citations or technological class overlap to measure knowledge spillovers, substantial knowledge sharing within firms may occur across inventors and patents that are neither explicitly cited nor classified in the same technological classes. We therefore develop a novel

measure of within-firm knowledge spillovers based on deep-learning-based textual analysis and examine spillover patterns and their underlying mechanisms.

Our paper is also closely related to the literature on corporate global diversification. Internationally diversified firms often trade at valuation discounts despite the potential benefits of sharing knowledge and resources across borders, much like firms that diversify domestically (Lang and Stulz, 1994; Denis, Denis, and Yost, 2002). Relatedly, Stiebale (2016) shows that target firms in cross-border acquisitions experience declines in innovation output, suggesting limited—or even absent—knowledge spillovers. Our findings indicate that limited within-firm knowledge spillover in multinational firms, driven by weak IP protection in foreign countries and internal competition among divisional managers, may help explain the somewhat depressed performance of globally diversified firms.

Our paper relates to the broader literature on the institutional determinants of innovation, particularly the role of intellectual property rights. Lerner (2002) shows that stronger IP protection increases innovation by foreign entities. Closely related to our study, Branstetter, Fisman, and Foley (2006) examine U.S. multinationals and show that internal technology transfer rises when host countries strengthen intellectual property rights. Our paper complements this literature by examining a global sample of multinational firms and developing deep-learning-based textual measures to quantify internal knowledge sharing across borders.

2. Sample and Methodology

2.1. Main Sample

Our sample consists of R&D units of all firms listed in Compustat Global and Compustat North America during the 1990-2014 period. We exclude financial (SIC 6000-6900) and utility (SIC

4900-4999) industries. The dependent variable is measured over years t+1 to t+3, corresponding to the period through 2017. We define foreign R&D units based on firms' patenting activities. To identify patents produced by firms, we utilize the firm-patent crosswalk developed by Bena, Ferreira, Matos, and Pires (2017). This crosswalk links firm data to patents registered in the USPTO's PatentsView database. Using this crosswalk, we identify 6,681 unique publicly listed firms headquartered across 39 countries. We aggregate patents and their inventors at the firm-country level according to their location at the time of the patent application. Using this aggregation, we construct unordered country-pairs within each firm, as described in the following paragraph.

2.2. Measure of Knowledge Spillover

To measure knowledge spillovers between R&D laboratories, we apply a deep-learning-based natural language processing approach to machine-read patent claims. Specifically, we convert each patent claim produced by a firm within a rolling three-year window into a dense vector representation using a neural embedding model and compute cosine similarity scores for all possible patent pairs within the firm. These patent-level similarities capture the degree of technological relatedness between inventions. We then aggregate the pairwise similarities to unordered country pairs by calculating the average similarity between patents originating from each pair of countries in a given year, yielding a country-pair measure of within-firm technological proximity.

Patent Claims

We focus on patent claims because they define the legal scope and technological boundaries of an invention. Patent claims specify the novel features that distinguish an invention from prior art and therefore provide the most precise description of the underlying technology. Compared with patent classifications or citation links—which reflect broad technological categories or examiner- and inventor-driven referencing behavior—claim text contains richer and more granular information about an invention's functional components and mechanisms. As a result, analyzing the semantic content of patent claims allows us to capture technological relatedness that may not be observable through patent class overlap or citation-based measures alone, enabling a more comprehensive assessment of potential knowledge spillovers.[4]

SBERT Embedding

To extract semantic information from patent claims, we employ Sentence-BERT (SBERT), a deep-learning language model that converts sentences into dense vector representations while preserving their semantic meaning. For example, the sentences "the device converts sunlight into electrical power" and "the apparatus generates electricity from solar radiation" use different words but convey nearly identical meanings; SBERT places these sentences close together in vector space. This approach offers important advantages over traditional text representations such as one-hot encoding or bag-of-words, which ignore semantic relationships between words, and over word-level embeddings like Word2Vec, which capture similarity at the individual word level but do not directly represent the meaning of entire sentences. By embedding full patent claims using SBERT, we obtain a scalable and semantically rich measure of technological similarity across patents.

[4] As a robustness check, we reproduce our main analysis by embedding the full patent descriptions rather than patent claims. The results are qualitatively similar (Table 8).

To illustrate the advantage of this approach, consider two patents developed by Volvo's R&D laboratories in Sweden and the United States. U.S. Patent #9,361,575 describes deep-learning–based automotive safety systems, while Swedish Patent #8,755,998 focuses on real-time collision-avoidance systems. Although both inventions emphasize vehicle and human safety, the patents neither cite each other nor share common patent classes and would therefore appear unrelated under traditional spillover measures. In contrast, our SBERT-based embedding identifies their underlying technological relatedness, assigning a cosine similarity score of 57 (out of 100).

Tech Similarity: From Patent-Pair Similarity to Segment-Pair Similarity

Our measure of technological similarity, *Tech similarity*, is derived from SBERT embeddings of patent claims over years t+1 through t+3. For each pair of geographic segments within a firm—including both cross-country and within-country pairs—we compute the average cosine similarity between patents produced in the two segments during this three-year window.

Specifically, we collect all patents produced by segment a of firm f during the three-year window and obtain their claim text from PatentsView, maintained by the U.S. Patent and Trademark Office (USPTO). We apply the SBERT model to generate a high-dimensional vector representation for each patent (Reimers and Gurevych, 2019). Let $c_p$denote the text of patent claim $p$produced by segment a. We convert each claim into a dense vector representation using SBERT:

$$\mathbf{e}_p = f_{\text{SBERT}}(c_p) \in \mathbb{R}^d.$$

These embeddings place patents in a d-dimensional semantic vector space that captures the underlying meaning of their technological content. We repeat this procedure for patent claim q produced by segment b of the same firm f and obtain its dense vector representation $\mathbf{e}_q$.

Using these vector representations, we compute cosine similarity scores between $\mathbf{e}_p$ and $\mathbf{e}_q$ for all patent pairs (p, q) across segments a and b:[5]

$$\text{Cosine Similarity}(p, q) = \frac{\mathbf{e}_p \cdot \mathbf{e}_q}{\sqrt{\mathbf{e}_p \cdot \mathbf{e}_p}\,\sqrt{\mathbf{e}_q \cdot \mathbf{e}_q}}$$

Appendix B reports the distribution of patent-pair-level cosine similarity for all within-firm patent pairs in 1990. In that year alone, our procedure computes 1,112,292,195 such similarity measures, and the number of patent pairs increases over time. The resulting patent-to-patent cosine similarity scores are centered around 0.276. While the vast majority of the values are positive, a small number of patent pairs exhibit negative cosine similarity, reflecting highly dissimilar patent claims in the embedding space.

We aggregate these patent-to-patent similarities to the unordered country-pair level within a firm by computing the mean similarity score for each country-pair in year *t*, averaged over the three-year window (*t*+1 to *t*+3), which we label *Tech similarity*:

$$Tech\ similarity_{a,b,f,t} = \frac{1}{|S_{ab,f,t}|} \sum_{(p,q) \in S_{ab,f,t}} Cosine\ Similarity(p, q)$$

where $S_{ab,f,t}$ denotes the set of patent pairs between segments a and b of firm *f* during the three-year window. Cosine similarity ranges from −1 to 1. For ease of interpretation, we multiply the measure by 100 so that *Tech similarity* ranges from −100 to 100. The vast majority of the similarity scores fall between 0 and 100. Higher values indicate greater semantic similarity between patents and, consequently, greater potential knowledge spillover. Importantly, our measure identifies relatedness in segments' innovation activities even when their patents do not

[5] Alternatively, one can aggregate embedded patent claims to the firm-segment level so that each segment is represented by a single vector summarizing the patent claims within the time window. Cosine similarity can then be computed directly between country-pairs within a firm. This approach is computationally faster and yields the same similarity measure. However, it does not allow us to recover the distribution of similarities at the patent-pair level.

overlap under traditional spillover measures such as technology-class proximity or cross-citation patterns.[6]

2.3. Main Regression Model

Our primary analysis is based on dyadic regressions using all unique dyads of geographic R&D segments within a firm over the period 1990-2014, such that each unordered pair (e.g., USA-South Korea) enters the sample only once. For example, if a U.S. firm maintains R&D operations in three countries—USA, South Korea, and Denmark—we construct six unique segment pairs for that year: USA-South Korea, USA-Denmark, South Korea-Denmark, USA-USA, South Korea-South Korea, and Denmark-Denmark.

We regress the technological similarity between segments a and b within firm f on an indicator variable, *Foreign-Home*, equal to one if the pair consists of an R&D segment in the headquarters country and a segment in a foreign country, and zero otherwise. Formally, we estimate the following ordinary least squares (OLS) specification:

$$Tech\ similarity_{a,b,f,t} = \alpha + \beta Foreign\text{-}Home_{a,b,f,t} + \gamma X_{f,t} + FE + \epsilon_{a,b,f,t}. \quad (1)$$

where $X_{f,t}$ denotes firm- and industry-level controls that we define below. All regressions include firm and year fixed effects, and standard errors are clustered by country-pair.

We note that the coefficient β on *Foreign-Home* captures how similar patents produced in a firm's headquarters country are to those produced in a foreign country, relative to two benchmarks: (1) the similarity between patents produced within the same country (whether

[6] Relatedly, Jaffe, Trajtenberg, and Fogarty (2000) note that roughly half of the citations in their sample do not reflect true knowledge spillovers, highlighting the noisiness of citation-based measures.

headquarters or foreign), and (2) the similarity between patents produced across two (non-headquarters) foreign countries. The second comparison ensures that β reflects differences specific to headquarters *versus* non-headquarters locations, rather than differences that simply arise from comparing patents across any two countries.

We control for firm and industry characteristics in our regressions. *Log assets* is defined as the natural logarithm of a firm's total assets. *Sales growth* is a firm's annual sales growth from year t-1 to year t. *R&D* is the ratio of a firm's R&D expenditure to total assets. *CAPEX* is the ratio of a firm's capital expenditure to total assets. *PPE* is the ratio of a firm's net plant, property, and equipment to total assets. *Leverage* is the sum of long-term and short-term debt, net of cash and cash equivalents, scaled by total assets. Our regressions also control for industry-level market competition. *Tech concentration* is the Herfindahl index of a firm's patent output across technological classes based on patents produced over the past three years. We define technological classes using IPC patent classes and subclasses. *Industry concentration* is the firm sales-based Herfindahl index for each two-digit SIC industry within a given host country and year.

Although most specifications include country-pair fixed effects, we additionally control for several time-varying country-pair characteristics. *Country-level patent similarity* is the cosine similarity between the patent class distributions of two countries, where each element of the vector represents the share of patents in a particular class among all patents filed in that country over the previous three years. $|\Delta GDP\ growth|$ is the absolute difference in GDP growth rates between the two countries. $|\Delta Corporate\ tax|$ is the absolute difference in corporate income tax rates between the two countries.

3. Main Results

Table 1 presents the summary statistics for the key variables in our main sample. Our sample consists of 496,496 unordered country-pair-year observations within firms. The mean *Tech similarity* between R&D labs in the sample is 4.43.[7] 15.4% of the observations represent pairs consisting of a firm's headquarters and a foreign R&D lab.

3.1. Baseline Estimation

We estimate the baseline specification in equation (1) and report the results in Table 2. Column 1 includes firm and year fixed effects. Our key independent variable, *Foreign-Home*, captures the difference in innovation similarity between a firm's headquarters and its foreign R&D laboratories relative to similarity within headquarters or among foreign (non-headquarters) R&D laboratories. Consistent with our prediction, the coefficient on *Foreign-Home* is significantly negative, indicating lower technological similarity between headquarters and foreign R&D laboratories.

In column 2, we include unordered country-pair fixed effects to account for time-invariant characteristics between countries. The coefficient on *Foreign-Home* remains significantly negative. In column 3, we further include industry-by-year fixed effects to absorb time-varying industry trends in patenting activity and innovation outcomes (e.g., Lerner and Seru, 2022). Industries are defined using two-digit SIC codes. Our results remain robust to these additional controls.

The magnitude of the effect is also economically meaningful. Patent similarity between a firm's headquarters and a foreign R&D laboratory is approximately 16.3% lower than the sample mean (−0.723/4.425). Overall, these findings indicate that technological similarity is significantly

[7] The relatively small mean value reflects the aggregation from patent-pair similarities to the country-pair level. While each patent pair receives equal weight in the patent-pair-level distribution, the aggregated measure assigns equal weight to each country pair regardless of the number of patent combinations it contains.

lower between headquarters and foreign laboratories, suggesting limited knowledge spillovers between them.

3.2. U.S. versus Non-U.S. Firms

Because a large share of patents in our sample is produced by US firms, one concern is that our results may be disproportionately driven by these firms. To address this issue, we split the sample into two subsamples based on headquarters location—US-headquartered firms and non-US-headquartered firms—and report the results in Table 3. We find that our main findings hold in both subsamples, indicating that the results are not driven solely by US firms. The coefficient magnitude is significantly larger for US-headquartered firms ($t = -2.61$, $p < 0.01$). This may reflect the fact that US firms are global leaders in innovation and therefore possess more valuable intellectual property, making them particularly sensitive to intellectual property risks in foreign markets.

3.3. AI-related Innovation

In addition, we focus on artificial intelligence (AI)-related patents. AI technologies are increasingly central to firms' competitive advantage, and leakage of core AI knowledge can impose substantial costs on firms. To identify AI-related patents, we use the USPTO Artificial Intelligence Patent Dataset (AIPD). Specifically, we use their 50% (86%) [93%] thresholds for identifying AI patents and define *AI patents 50* (*86*) [*93*] as an indicator variable equal to one if at least 20% of the patents produced by the firm's headquarters over the previous three years are AI-related, and zero otherwise. Higher thresholds correspond to stricter definitions of AI-related patents, while the USPTO recommends a threshold of 86, which balances false positives and false negatives in classification. We estimate the following model and present the results in Table 4.

$$Tech\ similarity_{a,b,f,t} = \alpha + \beta_1 AI\ patents_{f,t} \times Foreign\text{-}Home_{a,b,f,t} + \beta_2 Foreign\text{-}Home_{a,b,f,t} + \beta_3 AI\ patents_{f,t} + \gamma X_{f,t} + FE + \epsilon_{a,b,f,t} \quad (2)$$

We find that β_1 is significantly negative across alternative thresholds used to define AI patents (columns 1-3). This is consistent with the idea that firms may be less willing to share their core AI-related technologies with foreign labs due to the risk of IP leakage.

3.4. Potential Channels

This section examines the potential channels underlying the results reported in Table 2. Although an overseas R&D operation may offer attractive growth opportunities, a firm's headquarters may not fully share its core technologies with the overseas operation for several reasons. We investigate the channels related to intellectual property risk and internal R&D competition in Table 5 and those related to local customization, cultural proximity, and access to foreign technologies unavailable at home in Table 6.

3.4.1. IP Risk

A firm's headquarters may be reluctant to share its core technologies with foreign R&D units due to concerns about weak IP protection. Employees in foreign subsidiaries may misappropriate or leak the firm's key trade secrets by joining local competitors or founding new ventures. Foreign governments may also expropriate the firm's resources for political or strategic reasons. In such cases, bias in local judicial and enforcement systems can make it challenging for headquarters to safeguard its intellectual property once it has been misappropriated.

To proxy for IP risk concerns between two geographic segments within a firm, we first use the presence of a bilateral investment treaty (BIT) between the corresponding countries. BITs are formal agreements that enhance investor rights by reducing expropriation risk, improving contract enforceability, and strengthening intellectual property protections. These treaties generally provide foreign investors with access to neutral dispute-resolution mechanisms such as international arbitration. We therefore expect IP risk concerns to be mitigated once the headquarters country and the host country of the foreign R&D lab have signed a BIT. Specifically, we estimate the following equation and report the result in column 1 of Table 5:

$$Tech\ similarity_{a,b,f,t} = \alpha + \beta_1 BIT_{a,b,t} \times Foreign\text{-}Home_{a,b,f,t} + \beta_2 Foreign\text{-}Home_{a,b,f,t} + \beta_3 BIT_{a,b,t} + \gamma X_{f,t} + FE + \epsilon_{a,b,f,t}. \quad (3)$$

BIT is a dummy equal to one if a bilateral investment treaty was signed between the pair of countries, and zero otherwise. Consistent with the IP risk channel, the signing of a BIT increases patent similarity between a firm's headquarters and its foreign R&D labs, suggesting greater within-firm knowledge spillovers when IP risk is reduced.

Additionally, IP risks may also be greater when foreign R&D laboratories are located in countries that are politically distant from the headquarters country. Political frictions can weaken institutional cooperation and increase uncertainty about the protection of foreign investors' intellectual property. To capture this dimension, we use the political alignment measure of Gartzke (1998), which reflects the degree of similarity in United Nations General Assembly voting patterns between country pairs. Specifically, we define *Alignment* between two countries as follows:

$$Alignment_{i,j,t} = 1 - \frac{\sum_{k=1}^{N_t} |v_{i,k,t} - v_{j,k,t}|}{2N_t}$$

where i and j denote countries, t denotes year, and k indexes United Nations General Assembly votes. N_t represents the total number of votes in year t, and $v_{i,k,t}$ denotes the vote of country i on vote k. Votes are coded as 1 = Yes, 0 = Abstain, -1 = No. Based on this measure, we construct a dummy variable *Low alignment*, equal to one when the political alignment between the headquarters country and the host country of the foreign R&D laboratory falls below the sample median. Consistent with the IP-risk channel, we find that our results are stronger when foreign laboratories operate in countries with low political alignment with the headquarters country (column 2 of Table 5).

Overall, the results in columns 1 and 2 of Table 5, together with the evidence from AI-related patents, provide consistent evidence that IP risk concerns play an important role in shaping knowledge spillovers across national boundaries within firms.

3.4.2. R&D Competition among Divisions

Next, we examine whether the observed dissimilarity in innovation is driven by internal competition among divisional managers, which incentivizes them to pursue more distinctive innovation and discourages them from sharing their competitive advantages.

We use two firm-level measures of internal competition among divisions. First, following Seru (2014), we proxy for internal competition using the concentration of patent output relative to the allocation of innovative human capital (inventors) across divisions. Specifically, we define *Internal competition* as the inverse of the Herfindahl index based on each segment's patent output over the previous three years, with segment shares weighted by the number of unique inventors in that segment during the same period. This variable is larger when multiple segments contribute to innovation, indicating stronger internal competition.

In column 3 of Table 5, we find that patent dissimilarity between a firm's headquarters and its foreign R&D laboratories is more pronounced when the firm exhibits greater internal R&D competition. The findings are consistent with divisional managers attempting to stand out through distinctive innovation while withholding key advantages from others.

Our second measure captures internal competition arising from cross-country differences in R&D tax incentives across a firm's divisions. When country-level R&D tax incentives diverge, the after-tax marginal return to R&D differs across the firm's divisions. This creates stronger incentives for headquarters to reallocate R&D resources toward locations with more favorable tax treatment, thereby intensifying competition among divisional managers for budgets, projects, and internal influence. In contrast, in the extreme case where all countries offer identical R&D tax incentives, there would be no tax-driven differences in after-tax R&D returns across divisions and thus no internal competition arising from tax incentives.

We measure within-firm divergence in R&D tax incentives using the weighted standard deviation of the B-index across countries within a firm. The weights are based on the number of patents in each firm–country pair in year t-1, capturing the firm's pre-existing geographic exposure to tax policy differences while mitigating concerns that contemporaneous relocation of R&D activities mechanically drives the measure. The B-index represents the after-tax cost of one dollar of R&D expenditure after accounting for corporate income tax provisions and R&D tax credits; lower values indicate more generous R&D tax treatment (Warda, 2001; Thomson, 2013).[8]

We define *R&D tax-induced competition* as a dummy equal to one if the weighted standard deviation of the B-index across countries within a firm increases by at least 50% from year t–1 to year t, and zero otherwise. A value of one indicates increased internal competition arising from

[8] For example, a B-index value of 0.8 means that a firm needs only \$0.80 of pre-tax income to finance \$1 of R&D spending, corresponding to a 20% tax subsidy.

greater dispersion in R&D tax incentives across divisions. Because the B-index is available only for Organisation for Economic Co-operation and Development (OECD) countries, we restrict the sample to firms whose R&D activities are conducted exclusively in OECD countries and exclude firms with R&D laboratories located in non-OECD countries.

Consistent with the internal competition channel, column 4 of Table 5 shows that knowledge spillovers become more limited when internal competition increases due to cross-country differences in R&D tax incentives.

Overall, the evidence in columns 3 and 4 of Table 5 supports the internal competition channel, under which divisional managers are incentivized to differentiate their innovation strategies.

3.4.3. Local Customization of R&D

In addition, we examine whether foreign segments' adaptation of innovation to local needs drives the main result. Foreign R&D labs may adapt a firm's products to local market needs (local customization), resulting in R&D efforts that diverge from those undertaken at headquarters. Such local R&D customization typically co-locates with local manufacturing facilities to facilitate the rapid and efficient implementation of new technologies in local production (Fan, 2025). Colocation enhances communication and enables timely revisions of technologies when necessary. Accordingly, we examine the local customization channel by assessing whether R&D labs are co-located with manufacturing or production facilities.

We use two measures of local manufacturing: one based on country-level data and the other based on firm-level data. First, we use World Bank data on manufacturing value added as a percentage of a country's GDP and classify a country as having high manufacturing activity if its

manufacturing value added is above the annual median. We then define *Local manufacturing1* as a dummy equal to 1 if the pair of R&D laboratories includes a foreign country with high manufacturing activity in a given year, and 0 otherwise.

Alternatively, for a subsample of U.S. firms, we employ textual analysis to identify local manufacturing locations at the firm level. For each year, we search a firm's 10-K statement for manufacturing-related keywords. Specifically, we identify instances in which a country name and any manufacturing-related keyword—such as 'manufactur', 'production', 'factory', 'plant', 'assembly', or 'fabricat'—appear within the same sentence or within a ±1-sentence window. Based on this textual analysis, we define *Local manufacturing2* as a dummy equal to one if either non-headquarters country in the country pair is identified as a manufacturing or production location in the firm's most recent 10-K, and zero otherwise.

Columns 1 and 2 of Table 6 report the estimation results. Contrary to the local customization channel, we do not find evidence that dissimilarity in innovation is driven by local manufacturing activity. If anything, the coefficient on the interaction between *Foreign-Home* and the local manufacturing variables is significantly positive in column 1. This suggests that a firm's headquarters and its foreign R&D labs engage in more similar innovation when there is a stronger need to customize products for local markets. This pattern is consistent with technology spillovers documented in the foreign direct investment literature, but inconsistent with this channel as an explanation for the low patent similarity we document (e.g., Bloom et al., 2012).

3.4.4. Cultural Distance

Another potential explanation for the limited spillover is that foreign inventor-employees have different cultural backgrounds than those in the headquarters, which disrupts communication and consequently the sharing of knowledge.

We consider two measures of cultural similarity. First, we use the presence of a common official language between countries, given that language reflects cultural values to a large extent. We define *Common language* as a dummy equal to one if the two countries share an official language and zero otherwise.

Second, we use the World Values Survey (WVS) to quantify a country's cultural attributes. The WVS covers approximately 100 countries, with surveys conducted among randomly selected individuals who are representative across age, gender, occupation, and geographic regions. Specifically, following Ahern et al. (2015), we use the following three questions from the WVS:

- (Trust *versus* distrust) "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"

- (Hierarchy *versus* egalitarianism) "People have different ideas about following instructions at work. Some say that one should follow one's superior's instructions even when one does not fully agree with them. Other's say that one should follow one's superior's instructions only when one is convinced that they are right. With which of these two opinions do you agree?"

- (Individualism *versus* collectivism) "How would you place your views on this scale? 1 means you completely agree with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between."

We calculate a country's WVS cultural index score by summing the scores from the above three questions, each rescaled to range from 0 to 1. We define *Similar culture* as a dummy equal to one if the absolute difference between the two countries' cultural index scores from the WVS belongs to the bottom quartile of the sample, and zero otherwise.

Columns 3 and 4 of Table 6 report the estimation results. The coefficient on the interaction term between *Foreign-Home* and the cultural similarity measure is positive but statistically insignificant. Although the positive sign is consistent with the prediction that cultural distance between inventors in the home and foreign countries may limit knowledge spillovers, the lack of statistical significance suggests that this mechanism does not drive our results.

3.4.5. Foreign Technologies Unavailable at Home

We consider the possibility that firms establish R&D activities in foreign countries to access technologies that are not available in their home country. Such motives could mechanically generate low patent similarity between foreign laboratories and headquarters. Our baseline regressions already address this concern by controlling for country-pair patent class similarity (*Country-level patent similarity*). Nevertheless, as an additional test, we exclude country-pairs that fall within the bottom 25% of country-level patent similarity in a given year and reestimate our main specification. The subsample characterized by low country-level technological similarity captures situations in which many foreign technologies are likely unavailable in the home country. Column 5 of Table 6 reports the results. The estimates remain similar to our baseline findings, suggesting that our results are not driven by firms pursuing technologies that are unavailable in their home countries.

3.5. Robustness Tests

Some may argue that SBERT embeddings are not well suited for patent texts, given that the underlying models are trained primarily on general-domain corpora—such as Wikipedia, news articles, and web text—and therefore may not fully capture or differentiate specialized technological concepts. To address this concern, we employ an alternative deep-learning–based model, PatentSBERTa, proposed by Bekamiri et al. (2024). PatentSBERTa is an SBERT model further fine-tuned on patent documents, making it more suitable for extracting semantic information from technical descriptions. We recalculate *Tech similarity* using PatentSBERTa embeddings and report the results in Table 7. Our findings remain unchanged.

In addition, we apply the SBERT embedding model to *full patent descriptions* rather than patent claims when constructing our patent similarity measure. The results remain qualitatively similar to our main findings (Table 8).

4. Concluding Remarks

This study examines how knowledge generated in multinational firms diffuses across geographically dispersed R&D laboratories. Using a semantic measure of technological proximity based on deep-learning embeddings of patent claims, we find that headquarters tend to limit the sharing of core technologies with foreign R&D labs. We then investigate the mechanisms underlying these limited internal spillovers. Our evidence indicates that concerns over intellectual property risk and internal R&D competition among divisional managers play important roles in shaping how knowledge spreads across locations within the firm. These findings highlight how both external institutional risks and internal organizational incentives influence the global flow of ideas inside multinational enterprises.

References

Ahern, Kenneth R., Daniele Daminelli, and Cesare Fracassi, 2015, Lost in translation? The effect of cultural values on mergers around the world, *Journal of Financial Economics* 117: 165-189.

Bekamiri, Hamid, Daniel S. Hain, and Roman Jurowetzki, 2024, PatentSBERTa: A deep NLP based hybrid model for patent distance and classification using augmented SBERT, *Technological Forecasting and Social Change* 206: 123536.

Bena, Jan, Miguel A. Ferreira, Pedro Matos, and Pedro Pires, 2017, Are foreign investors locusts? The long-term effects of foreign institutional ownership, *Journal of Financial Economics* 126(1): 122-146.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, 2012, Americans do IT better: US multinationals and the productivity miracle, *American Economic Review* 102(1): 167-201.

Branstetter, Lee G., Raymond Fisman, and C. Fritz Foley, 2006, Do stronger intellectual property rights increase international technology transfer? Empirical evidence from US firm-level panel data, *Quarterly Journal of Economics* 121(1): 321-349.

Carlino, Gerald, and William R. Kerr, 2015, Agglomeration and innovation, *Handbook of Regional and Urban Economics* 5: 349-404.

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

Denis, David J., Diane K. Denis, and Keven Yost, 2002, Global diversification, industrial diversification, and firm value, *Journal of Finance* 57(5): 1951-1979.

Fan, Jingting, 2025, Talent, geography, and offshore R&D, *Review of Economic Studies* 92(2): 1022-1060.

Gartzke, Erik, 1998, Kant we all just get along? Opportunity, willingness, and the origins, of the democratic peace, *American Journal of Political Science* 42(1): 1-27.

Griliches, Zvi, 1979, Issues in assessing the contribution of research and development to productivity growth, *Bell Journal of Economics*: 92-116.

He, Jie, and Xuan Tian, 2020, Institutions and innovation, *Annual Review of Financial Economics* 12(1): 377-398.

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

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 108(3): 577-598.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty, 2000, Knowledge spillovers and patent citations: Evidence from a survey of inventors, *American Economic Review* 90(2): 215-218.

Jiang, Cheng, Kose John, Kyeong Hun Lee, and Emma Qianying Xu, 2025, Local labor markets and corporate innovation, *Journal of Financial and Quantitative Analysis* 61(1): 441-479.

Lang, Larry H.P., and Rene M. Stulz, 1994, Tobin's q, corporate diversification, and firm performance, *Journal of Political Economy* 102(6): 1248-1280.

Lerner, Josh, 2002, 150 years of patent protection, *American Economic Review* 92(2): 221-225.

Lerner, Josh, and Amit Seru, 2022, The use and misuse of patent data: Issues for finance and beyond, *Review of Financial Studies* 35(6): 2667-2704.

Li, Kai, and Jin Wang, 2023, Inter-firm inventor collaboration and path-breaking innovation: Evidence from inventor teams post-merger, *Journal of Financial and Quantitative Analysis* 58(3): 1144-1171.

Pairolero, Nicholas A., Alexander V. Giczy, Gerard Torres, Tisa Islam Erana, Mark A. Finlayson, and Andrew A. Toole, 2025, The artificial intelligence patent dataset (AIPD) 2023 update, *Journal of Technology Transfer*: 1-24.

Reimers, Nils, and Iryna Gurevych, 2019, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, arXiv preprint arXiv:1908.10084.
Seru, Amit, 2014, Firm boundaries matter: Evidence from conglomerates and R&D activity, *Journal of Financial Economics* 111(2): 381-405.
Stiebale, Joel, 2016, Cross-border M&As and innovative activity of acquiring and target firms, *Journal of International Economics* 99: 1-15.
Teece, David J., 1977, Technology transfer by multinational firms: The resource cost of transferring technological know-how, *Economic Journal* 87(346): 242-261.
Thomson, Russell, 2013, Measures of R&D tax incentives for OECD countries, *Review of Economics and Institutions* 4(3): 35.
Warda, Jacek, 2001, Measuring the value of R&D tax treatment in OECD countries, *Science and Technology Review* 27: 185-211.
Whalen, Ryan, Alina Lungeanu, Leslie DeChurch, and Noshir Contractor, 2020, Patent similarity data and innovation metrics, *Journal of Empirical Legal Studies* 17(3): 615-639.

Table 1. Sample Statistics

This table presents descriptive statistics for our main sample, which consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat database. The sample covers the period from 1990 to 2014. *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on SBERT embeddings derived from patent claims. *Foreign-Home* is an indicator variable that equals one if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and zero otherwise. *Log inventors* is defined as the natural logarithm of the number of unique inventors who produced patents for the firm over the past three years. *Log assets* is the natural logarithm of a firm's total assets. *Sales growth* is a firm's annual sales growth from year *t*-1 to year *t*. *R&D* is the ratio of a firm's R&D expense to total assets. *CAPEX* is the ratio of a firm's capital expenditure to total assets. *PPE* is the ratio of a firm's net plant, property, and equipment to total assets. *Leverage* is defined as long-term debt plus short-term debt minus cash and cash equivalents, all scaled by total assets. *Industry concentration* is the firm sales-based Herfindahl index for each country and two-digit SIC industry to which the firm belongs. *Tech concentration* is the Herfindahl index of a firm's patents across technological classes, based on patents produced over the past three years. *Country-level patent similarity* is the cosine similarity between the patent class distributions of two countries, where each element of the vector corresponds to the share of patents in a particular class among all patents filed in that country during the previous three years. |*ΔGDP growth*| is the absolute difference in GDP growth rates between countries. |*ΔCorporate tax*| is the absolute difference in corporate income tax rates between countries. All variables are winsorized at the 1st and 99th tails.

Variable	Obs.	Mean	Std. Dev.	25th Pctl.	Median	75th Pctl.
Tech similarity	496,496	4.425	2.009	3.174	3.874	4.954
Foreign-Home	496,496	0.154				
Log inventors	496,496	1.412	1.698	0.000	0.693	2.485
Log assets	496,496	8.900	2.277	7.693	9.309	10.526
Sales growth	496,496	1.341	6.414	-0.283	0.000	0.559
R&D	496,496	0.075	0.079	0.024	0.055	0.100
Capex	496,496	0.050	0.039	0.022	0.040	0.068
PPE	496,496	0.222	0.149	0.102	0.196	0.309
Leverage	496,496	0.197	0.160	0.074	0.174	0.281
Industry concentration	496,496	0.273	0.279	0.062	0.185	0.386
Tech concentration	496,496	0.662	0.228	0.485	0.683	0.852
Country-level patent similarity	496,496	0.778	0.211	0.651	0.811	1.000
|*ΔGDP growth*|	496,496	0.019	0.024	0.000	0.011	0.029
|*ΔCorporate tax*|	496,496	5.652	6.332	0.000	4.000	9.250

Table 2. Knowledge Spillover between a Firm's Headquarters and Foreign R&D Labs

This table presents the results of OLS regressions to examine differences in innovation between a firm's headquarters and foreign R&D labs. The sample consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat North America and Compustat Global datasets during the period from 1990 to 2014. The dependent variable *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on SBERT embeddings derived from patent claims. *Foreign-Home* is an indicator variable that equals one if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and zero otherwise. All variables are winsorized at the 1st and 99th tails. Reported in parentheses are *t*-statistics with standard errors clustered by country-pair. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	Dependent variable: *Tech similarity*		
	1	2	3
Foreign-Home	-0.692***	-0.724***	-0.723***
	(-6.93)	(-7.02)	(-7.01)
Log inventors	-0.215***	-0.213***	-0.214***
	(-5.50)	(-5.74)	(-5.76)
Log assets	-0.054***	-0.046***	-0.050***
	(-4.69)	(-3.63)	(-3.75)
Sales growth	0.003***	0.003***	0.003***
	(10.37)	(9.06)	(8.39)
R&D	-0.150	-0.145	-0.155
	(-1.08)	(-0.95)	(-1.05)
CAPEX	-0.164	-0.237	-0.132
	(-0.43)	(-0.64)	(-0.32)
PPE	0.031	0.014	0.075
	(0.21)	(0.11)	(0.55)
Leverage	0.072	0.095**	0.068
	(1.62)	(2.01)	(1.39)
Industry concentration	0.177*	0.163	0.169
	(1.69)	(1.57)	(1.60)
Tech concentration	0.227***	0.429***	0.463***
	(5.62)	(4.03)	(4.70)
Country-level patent similarity	2.612***	2.623***	2.632***
	(13.09)	(13.25)	(13.29)
$\lvert\Delta$*GDP growth*$\rvert$	-8.298***	-8.320***	-8.381***
	(-9.34)	(-9.40)	(-9.43)
$\lvert\Delta$*Corporate tax*$\rvert$	-0.050***	-0.050***	-0.050***
	(-10.18)	(-10.23)	(-10.19)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	–
Country-pair FE	–	Yes	Yes
Industry × year FE	–	–	Yes
Adj. R^2	0.45	0.45	0.45
Obs.	496,496	496,496	496,496

Table 3. Knowledge Spillover Limitations: U.S. versus Non-U.S. Firms

This table presents the results of OLS regressions examining differences in innovation between a firm's headquarters and its foreign R&D laboratories, separately for subsamples of U.S. and non-U.S. firms. The sample consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat North America and Compustat Global datasets during the period from 1990 to 2014. Column 1 excludes firms headquartered outside the US, and column 2 excludes firms headquartered in the US. The dependent variable *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on SBERT embeddings derived from patent claims. *Foreign-Home* is an indicator variable that equals one if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and zero otherwise. All variables are winsorized at the 1st and 99th tails. Reported in parentheses are *t*-statistics with standard errors clustered by country-pair. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	US firms	Non-US firms
	1	2
Foreign-Home	-0.916***	-0.453***
	(-5.92)	(-5.31)
Log inventors	-0.248***	-0.200***
	(-6.54)	(-5.92)
Log assets	-0.045**	-0.064***
	(-2.30)	(-2.70)
Sales growth	0.003***	0.002
	(9.34)	(1.44)
R&D	-0.194	-0.080
	(-1.28)	(-0.40)
CAPEX	-0.861***	0.492*
	(-5.01)	(1.82)
PPE	0.275**	-0.200
	(2.16)	(-1.12)
Leverage	-0.023	0.236***
	(-0.34)	(3.12)
Industry concentration	0.240	0.039
	(1.60)	(0.40)
Tech concentration	0.440***	0.529***
	(5.04)	(3.60)
Country-level patent similarity	2.740***	2.580***
	(11.73)	(13.13)
$\lvert\Delta$*GDP growth*$\rvert$	-8.586***	-8.320***
	(-7.64)	(-9.80)
$\lvert\Delta$*Corporate tax*$\rvert$	-0.047***	-0.052***
	(-8.63)	(-10.96)
$\beta_{US} - \beta_{NonUS}$ (unequal-variance *t*-statistic)	-0.463*** (-2.61)	
Firm FE	Yes	Yes
Country-pair FE	Yes	Yes
Industry × year FE	Yes	Yes
Adj. R^2	0.48	0.39
Obs.	305,784	190,712

Table 4. AI Patents and Limited Knowledge Spillover

This table reports OLS regression results comparing AI-related patents and non-AI-related patents in terms of patent dissimilarity between a firm's headquarters and foreign R&D labs. The sample consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat database during the period from 1990 to 2014. The dependent variable *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on SBERT embeddings derived from patent claims. *Foreign-Home* is an indicator variable that equals one if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and zero otherwise. *AI patents 50* (*86*) [*93*] is a dummy equal to one if at least 20% of the headquarters' patents over the previous 3 years are AI-related, and zero otherwise. Patent-level AI classification is based on USPTO AIPD probability scores using thresholds of 50% (86%) [93%]. All variables are winsorized at the 1st and 99th tails. Reported in parentheses are *t*-statistics with standard errors clustered by country-pair. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	1	2	3
Foreign-Home	-0.717***	-0.719***	-0.720***
	(-6.97)	(-7.00)	(-7.02)
Foreign-Home × AI patents 50	-0.540***		
	(-7.48)		
Foreign-Home × AI patents 86		-0.601***	
		(-6.35)	
Foreign-Home × AI patents 93			-0.590***
			(-6.02)
Controls	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Country-pair FE	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes
Adj. R2	0.44	0.44	0.44
Obs.	496,496	496,496	496,496

Table 5. IP Risk and Internal R&D Competition as Channels Limiting Knowledge Spillover

This table reports OLS regression results examining the IP risk and internal competition channels underlying patent dissimilarity between a firm's headquarters and foreign R&D labs. The main sample consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat database during the period from 1990 to 2014. The dependent variable *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on SBERT embeddings derived from patent claims. *Foreign-Home* is an indicator variable that equals 1 if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and 0 otherwise. *BIT* is a dummy equal to 1 if a bilateral investment treaty was signed between the pair of countries, and 0 otherwise. *Low alignment* is a dummy equal to 1 if two countries' political alignment is in the bottom 25th percentile in a given year, and 0 otherwise. Political alignment is measured using the similarity of voting patterns in the United Nations General Assembly. *Internal competition* is defined as the inverse of the segment-share–based Herfindahl index within a firm. Segment share is calculated as the number of patents produced by a segment over the past 3 years divided by the number of unique inventors during the same period. *R&D tax-induced competition* is a dummy equal to 1 if the weighted standard deviation of the B-index across countries within a firm increases by at least 50% from year *t*–1 to year *t*, and 0 otherwise. The weighted standard deviation is computed using the number of patents of each firm-country pair in year *t*–1 as weights. All variables are winsorized at the 1st and 99th tails. Reported in parentheses are *t*-statistics with standard errors clustered by country-pair. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	IP Risk		Competition among Divisional Managers	
	BIT	Political Alignment	Innovation Output Dispersion	R&D Tax Incentive Dispersion
	1	2	3	4
Foreign-Home	-0.733***	-0.555***	-0.591***	-0.630***
	(-7.04)	(-6.56)	(-6.67)	(-6.95)
Foreign-Home × *BIT*	0.533***			
	(5.26)			
Foreign-Home × *Low alignment*		-0.271***		
		(-2.68)		
Foreign-Home × *Internal competition*			-0.298***	
			(-5.31)	
Foreign-Home × *R&D tax-induced competit.*				-0.220***
				(-3.45)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Country-pair FE	Yes	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes	Yes
Adj. R^2	0.45	0.45	0.45	0.38
Obs.	496,496	496,496	496,496	328,170

Table 6. Alternative Channels: Local Customization, Cultural Distance, and Technological Distance

This table reports OLS regression results examining the potential channels underlying patent dissimilarity between a firm's headquarters and foreign R&D labs. Columns 1-2 examine the local customization channel, while columns 3-4 examine the cultural distance channel. Column 5 tests the channel of accessing foreign technologies unavailable at home by excluding country pairs in the bottom 25% of country-level patent similarity. The main sample consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat database during the period from 1990 to 2014. The dependent variable *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on SBERT embeddings derived from claims. *Foreign-Home* is an indicator variable that equals one if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and zero otherwise. *Local manufacturing1* is a country-level indicator variable that equals one if a pair of R&D laboratories includes a foreign country with high manufacturing activity in a given year. A country is classified as having high manufacturing activity if its manufacturing value added as a percentage of GDP is above the annual median. *Local manufacturing2* is an indicator variable, based on firm-level textual analysis, that equals 1 if either non-headquarters country in the country pair is mentioned as a manufacturing or production location in the firm's most recent 10-K filing, and 0 otherwise. *Common language* is a dummy equal to 1 if the two countries share an official language and 0 otherwise. *Similar culture* is a dummy equal to 1 if the absolute difference between two countries' cultural index scores from the WVS falls within the bottom quartile of the sample, and 0 otherwise. All variables are winsorized at the 1st and 99th tails. Reported in parentheses are *t*-statistics with standard errors clustered by country-pair. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	Full Sample	US Firms	Full Sample	WVS Sample	Excl. Tech-distant Countries
	1	2	3	4	5
Foreign-Home	-0.635***	-0.939***	-0.719***	-0.652***	-0.602***
	(-5.35)	(-5.94)	(-8.24)	(-6.12)	(-24.36)
Foreign-Home × *Local manufacturing1*	0.571***				
	(3.55)				
Foreign-Home × *Local manufacturing2*		0.019			
		(0.23)			
Foreign-Home × *Common language*			0.248		
			(1.53)		
Foreign-Home × *Similar culture*				0.144	
				(0.92)	
Controls	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Country-pair FE	Yes	Yes	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.46	0.48	0.46	0.49	0.49
Obs.	496,496	305,784	496,496	317,043	372,817

Table 7. Alternative Measure of Knowledge Spillover

This table replicates our main results using PatentSBERTa embeddings of patent claims. The sample consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat database during the period from 1990 to 2014. The dependent variable *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on PatentSBERTa embeddings derived from claims. *Foreign-Home* is an indicator variable that equals one if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and zero otherwise. *BIT* is a dummy equal to one if a bilateral investment treaty was signed between the pair of countries, and zero otherwise. *Low alignment* is a dummy equal to 1 if two countries' political alignment is in the bottom 25th percentile in a given year, and 0 otherwise. Political alignment is measured using the similarity of voting patterns in the United Nations General Assembly. *Internal competition* is defined as the inverse of the segment-share–based Herfindahl index within a firm. Segment share is calculated as the number of patents produced by a segment over the past 3 years divided by the number of unique inventors during the same period. *R&D tax-induced competition* is a dummy equal to 1 if the weighted standard deviation of the B-index across countries within a firm increases by at least 50% from year *t*–1 to year *t*, and 0 otherwise. *Local manufacturing1* is a country-level indicator variable that equals one if a pair of R&D laboratories includes a foreign country with high manufacturing activity in a given year. A country is classified as having high manufacturing activity if its manufacturing value added as a percentage of GDP is above the annual median. *Common language* is a dummy equal to 1 if the two countries share an official language and 0 otherwise. All variables are winsorized at the 1st and 99th tails. Reported in parentheses are *t*-statistics with standard errors clustered by country-pair. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	Full Sample	Full Sample	Full Sample	Full Sample	OECD	Full Sample	Full Sample	Excl. tech-distant country
	1	2	3	4	5	6	7	8
Foreign-Home	-0.126***	-0.127***	-0.099***	-0.102***	-0.109***	-0.110***	-0.122***	-0.097***
	(-7.07)	(-7.09)	(-6.65)	(-6.64)	(-6.97)	(-5.10)	(-7.55)	(-9.51)
Foreign-Home × *BIT*		0.093***						
		(4.82)						
Foreign-Home × *Low alignment*			-0.043**					
			(-2.42)					
Foreign-Home × *Internal competition*				-0.053***				
				(-5.55)				
Foreign-Home × *R&D tax-ind. competit.*					-0.037***			
					(-3.51)			
Foreign-Home × *Local manufacturing1*						0.098***		
						(3.43)		
Foreign-Home × *Common language*							0.044	
							(1.51)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.42	0.42	0.42	0.42	0.41	0.42	0.43	0.47
Obs.	496,496	496,496	496,496	496,496	328,170	496,496	496,496	372,817

Table 8. Full Patent Description Embeddings

This table replicates our main results using SBERT embeddings based on full patent descriptions. The sample consists of all unique dyads of geographic R&D segments within a firm whose parent company is included in the Compustat database during the period from 1990 to 2014. The dependent variable *Tech similarity* is defined as the average cosine similarity between patents produced by a pair of R&D units, based on SBERT embeddings derived from descriptions. *Foreign-Home* is an indicator variable that equals one if the pair consists of an R&D segment located in the firm's headquarters country and another located in a foreign country, and zero otherwise. *BIT* is a dummy equal to one if a bilateral investment treaty was signed between the pair of countries, and zero otherwise. *Low alignment* is a dummy equal to 1 if two countries' political alignment is in the bottom 25th percentile in a given year, and 0 otherwise. Political alignment is measured using the similarity of voting patterns in the United Nations General Assembly. *Internal competition* is defined as the inverse of the segment-share–based Herfindahl index within a firm. Segment share is calculated as the number of patents produced by a segment over the past 3 years divided by the number of unique inventors during the same period. *R&D tax-induced competition* is a dummy equal to 1 if the weighted standard deviation of the B-index across countries within a firm increases by at least 50% from year *t*–1 to year *t*, and 0 otherwise. *Local manufacturing1* is a country-level indicator variable that equals one if a pair of R&D laboratories includes a foreign country with high manufacturing activity in a given year. A country is classified as having high manufacturing activity if its manufacturing value added as a percentage of GDP is above the annual median. *Common language* is a dummy equal to 1 if the two countries share an official language and 0 otherwise. All variables are winsorized at the 1st and 99th tails. Reported in parentheses are *t*-statistics with standard errors clustered by country-pair. ***, **, and * indicate statistical significance at 1%, 5%, and 10% levels, respectively.

	Full Sample	Full Sample	Full Sample	Full Sample	OECD	Full Sample	Full Sample	Excl. Tech-distant country
	1	2	3	4	5	6	7	8
Foreign-Home	-0.709***	-0.719***	-0.533***	-0.558***	-0.629***	-0.606***	-0.675***	-0.541***
	(-6.67)	(-6.72)	(-5.59)	(-5.98)	(-6.61)	(-4.90)	(-6.73)	(-9.15)
Foreign-Home × *BIT*		0.601***						
		(5.12)						
Foreign-Home × *Low alignment*			-0.288***					
			(-2.67)					
Foreign-Home × *Internal competition*				-0.342***				
				(-5.63)				
Foreign-Home × *R&D tax-ind. competit.*					-0.237***			
					(-3.86)			
Foreign-Home × *Local manufacturing1*						0.638***		
						(3.69)		
Foreign-Home × *Common language*							0.234	
							(1.35)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.43	0.43	0.43	0.43	0.37	0.44	0.44	0.49
Obs.	496,496	496,496	496,496	496,496	328,170	496,496	496,496	372,817

Appendix A. Variable Definition

- *Tech similarity*: The average cosine similarity between patents produced by a pair of R&D units over a 3-year period (t+1 to t+3) multiplied by 100, based on SBERT embeddings derived from patent claims (*Source*: PatentsView).
- *Log inventors*: The natural logarithm of the number of unique inventors who produced patents for the firm during the 3-year period spanning *t*-2 through *t* (*Source*: PatentsView).
- *Log assets*: The natural logarithm of a firm's total assets (*Source*: Compustat).
- *Sales growth*: A firm's annual sales growth from year *t*-1 to year *t* (*Source*: Compustat).
- *R&D*: The ratio of a firm's R&D expense to total assets (*Source*: Compustat).
- *CAPEX*: The ratio of a firm's capital expenditure to total assets (*Source*: Compustat).
- *PPE*: The ratio of a firm's net plant, property, and equipment to total assets (*Source*: Compustat).
- *Leverage*: The sum of long-term and short-term debt, net of cash and cash equivalents, scaled by total assets (*Source*: Compustat).
- *Industry concentration*: The firm sales-based Herfindahl index for each country and two-digit SIC industry (*Source*: Compustat).
- *Tech concentration*: The Herfindahl index of a firm's patents across technological classes, based on patents produced over the past three years. Technological classes are defined using IPC patent classes and subclasses (*Source*: PatentsView).
- *Country-level patent similarity*: The cosine similarity between the patent class distributions of two countries, where each element of the vector corresponds to the share of patents in a particular class among all patents filed in that country during the previous three years (*Source*: PatentsView).
- |*ΔGDP growth*|: Absolute difference in GDP growth rates between countries (*Source*: Penn World Table).
- |*ΔCorporate tax*|: Absolute difference in corporate income tax rates between countries (*Source*: Tax Foundation).
- *AI patents 50* (*86*) [*93*]: A dummy equal to one if at least 20% of the headquarters' patents over the previous three years are AI-related, and zero otherwise. To classify a patent as AI-related, we use the USPTO Artificial Intelligence Patent Dataset (AIPD) probability scores and apply the 50% (86%) [93%] thresholds proposed by Pairolero et al. (2025) (*Source*: USPTO AIPD).
- *BIT*: A dummy equal to 1 if a bilateral investment treaty (BIT) was signed between the pair of countries, and 0 otherwise (*Source*: UNCTAD).
- *Low alignment*: A dummy equal to 1 if two countries' political alignment is in the bottom 25th percentile in a given year. Political alignment is measured using the similarity of voting patterns in the United

Nations General Assembly using the formula: $Alignment_{i,j,t} = 1 - \frac{\sum_{k=1}^{N_t} |v_{i,k,t} - v_{j,k,t}|}{2N_t}$, where i and j denote countries, t denotes year, k indicates a United Nations General Assembly vote, N_t indicates the total number of votes in year t, and $v_{i,k,t}$ indicates the vote of country i on vote k. Voting is coded as 1 = Yes, 0 = Abstain, -1 = No (*Source*: Gartzke (1998)).

- *Internal competition*: The inverse of the segment-share–based HHI within a firm. Segment share is calculated as the number of patents produced by a segment during the 3-year period spanning *t*-2 through *t*, divided by the number of unique inventors during the same period (*Source*: PatentsView).
- *R&D tax-induced competition*: A dummy variable equal to one if the weighted standard deviation of the B-index across countries within a firm increases by at least 50% from year *t*–1 to year t, indicating a widening dispersion in R&D tax incentives across the firm's divisions. The weighted standard deviation is computed using the number of patents of each firm-country pair in year *t*–1 as weights. The country-level B-index measures the after-tax cost of one dollar of R&D expenditure; lower values indicate more generous R&D tax incentives (*Source*: Thomson (2013); the OECD R&D Tax Incentives Database).
- *Local manufacturing1*: A country-level indicator variable that equals one if a pair of R&D laboratories includes a foreign country with high manufacturing activity in a given year. A country is classified as having high manufacturing activity if its manufacturing value added as a percentage of GDP is above the annual median (*Source*: World Bank).
- *Local manufacturing2*: An indicator variable derived from firm-level textual analysis. It equals 1 if either non-headquarters country in the country pair is identified as a manufacturing or production location in the firm's most recent 10-K, and 0 otherwise (*Source*: SEC EDGAR).
- *Common language*: A dummy equal to 1 if the two countries share an official language and 0 otherwise (*Source*: CIA World Factbook).
- *Similar culture*: A dummy equal to 1 if the absolute difference between the two countries' cultural index scores from the World Values Survey (WVS) falls below the 25th percentile of the sample, and 0 otherwise. A country's WVS cultural index score is calculated by summing the scores from the following three questions, each rescaled to range from 0 to 1:
    1. (Trust *versus* distrust) "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"
    2. (Hierarchy *versus* egalitarianism) "People have different ideas about following instructions at work. Some say that one should follow one's superior's instructions even when one does not fully agree with them. Other's say that one should follow one's superior's instructions only when one is convinced that they are right. With which of these two opinions do you agree?"

3. (Individualism *versus* collectivism) “How would you place your views on this scale? 1 means you completely agree with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between.”

Appendix B. Distribution of Patent-pair-level Cosine Similarity

This table reports the distribution of patent-pair-level cosine similarity between SBERT-embedded representations of patent claims from two R&D segments within the same firm. The sample consists of patent pairs constructed from patents applied for during 1991–1993, corresponding to firm-year 1990. Panel A reports summary statistics, and Panel B presents a histogram of the similarity distribution using a bin width of 0.01.

Panel A. Sample Statistics at the Patent-pair-level

	Mean	Std. Dev.	25th Pctl.	Median	75th Pctl.	Obs.
Patent-pair-level Cosine Similarity	0.276	0.129	0.184	0.272	0.364	1,112,292,195

Panel B. Histogram

