Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Corporate Innovation Along the Supply Chain

Yongqiang Chu, Xuan Tian, Wenyu Wang

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 65, No. 6, June 2019, pp. 2445–2466
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Corporate Innovation Along the Supply Chain

Yongqiang Chu,[a] Xuan Tian,[b] Wenyu Wang[c]

[a] Darla Moore School of Business, University of South Carolina, Columbia, South Carolina 29208; [b] PBC School of Finance, Tsinghua University, Beijing 100083, China; [c] Kelley School of Business, Indiana University, Bloomington, Indiana 47405

Contact: yongqiang.chu@moore.sc.edu, http://orcid.org/0000-0003-1739-7422 (YC); tianx@pbcsf.tsinghua.edu.cn, http://orcid.org/0000-0002-2746-576X (XT); wenywang@indiana.edu, http://orcid.org/0000-0003-0587-3655 (WW)

Received: August 9, 2016
Revised: May 14, 2017
Accepted: August 7, 2017
Published Online in Articles in Advance: March 9, 2018

https://doi.org/10.1287/mnsc.2017.2924

Abstract. In this paper, we document a positive effect of supplier–customer geographic proximity on supplier innovation. To establish causality, we explore plausibly exogenous variation in proximity caused by customer relocations. The positive effect of supplier–customer proximity on supplier innovation is stronger when customers are more innovative themselves, when suppliers and customers are closer in technological space, and when customers' demand accounts for a larger fraction of suppliers' total sales. These findings suggest that the feedback channel and the demand channel are likely underlying mechanisms through which supplier–customer proximity affects supplier innovation. Overall, our paper sheds new light on the real effect of supplier–customer relationship on corporate innovation.

History: Accepted by Gustavo Manso, finance.
Funding: X. Tian acknowledges financial support from the National Natural Science Foundation of China [Grant 71790591] and the Tsinghua University Research Grant [Grant 20151080451].

Keywords: geographic proximity • innovation • supply chain

1. Introduction

A growing literature has examined various effects of supplier–customer relationships on corporate decisions.[1] While most existing studies highlight the importance of the interactions between suppliers and customers along the supply chain in corporate finance, these studies mainly focus on how supplier–customer relationships affect financial decisions. The existing literature has largely overlooked an important impact of supplier–customer relationships: their real effects on corporate investment decisions. In this paper, we focus on a special type of corporate investment—technological innovation, which is critical for a firm's long-term competitive advantages and sustainable growth (Porter 1992). Specifically, we study how geographic proximity between a supplier and its customer (i.e., supplier–customer proximity) affects the supplier's innovation outputs.

Supplier–customer proximity could affect the supplier's innovation through a few plausible channels. First, in a seminal paper, Manso (2011) constructs a principal–agent model and proposes that timely feedback from the principal to the agent enhances the agent's innovation. In our setting, timely feedback from customers makes possible more prompt adjustment by their suppliers in the intermediate stages of R&D, which is crucial to the ultimate success of the suppliers' innovation. Proximity between the customer and the supplier can capture timely feedback, because a short distance facilitates soft information exchange between the two parties. Despite the rapid development of transportation and communication tools, supplier–customer proximity remains important as feedback from customers often involves soft information production and transmission. Soft information, by definition, is difficult to put down on paper, store electronically, or transfer to others (Petersen and Rajan 2002). Hence, obtaining timely feedback from customers requires the supplier's frequent, on-site, and, in many cases, face-to-face interactions with customers. In recent decades, the contribution of customer feedback becomes even more pivotal, because more and more firms engage their customers in the innovation process (Chesbrough 2006, Prahalad and Ramaswamy 2013, Von Hippel 2005). We term this channel as the *feedback* channel.

The second plausible channel through which supplier–customer proximity affects supplier innovation is suggested by a few theoretical models on innovation (e.g., d'Aspremont and Jacquemin 1988, Kamien et al. 1992, Leahy and Neary 1997). In these models, innovation contributes to the production process by reducing the marginal cost, and thus a supplier's incentive to innovate for its customer is closely linked to the quantity of the products (services) it produces (provides) for the customer. A short distance between a supplier and its customer reduces transportation costs and could increase the demand from the customer. Therefore, these theories predict that supplier–customer proximity motivates supplier innovation through increased demand from the customer. We term this channel as the *demand* channel.

The third plausible channel is related to the agglomeration effect documented in the literature. Suppliers and customers, when they locate close to each other, may share important factors in the production process, such as intermediate inputs, talent pools, and natural resources (e.g., Orlando 2004). Numerous studies have shown that the agglomeration effect has important implications for industrial organization and can, under certain circumstances, generate equilibrium growth paths for the economy as a whole. The agglomeration effect could enhance suppliers' innovation, and it diminishes with distance, which may make supplier–customer proximity an important determinant of supplier innovation. We term this channel as the *agglomeration* channel.

There are also other channels at play. For example, a recent study by Dasgupta et al. (2015) documents that a tight social connection between managers in the supplier and customer firms helps mitigate hold-up problems and improves supplier innovation. Since managers located in proximate regions have more opportunities to communicate, a short distance between a supplier and its customer can create a tight social connection, which in turn enhances supplier innovation. We term this channel as the *connection* channel.

Identifying a causal effect of supplier–customer proximity on supplier innovation is challenging. The location choices of suppliers and customers are likely endogenous and are affected by unobservable firm characteristics and economic conditions. Hence, a correlation between supplier–customer proximity and supplier innovation may tell us little about causality. We overcome this identification challenge by exploiting plausibly exogenous shocks caused by customer relocations. Specifically, we notice that, in the Compustat segment customer database, customers are much larger than their suppliers (often more than 100 times larger). Hence, it is reasonable to believe that large customers are unlikely to change their locations in response to factors related to their suppliers that are much smaller. This feature allows us to use customer firm relocations as plausibly exogenous shocks to the geographic proximity between suppliers and customers.

Using a generalized difference-in-differences method, we find that the geographic proximity between the supplier and its major customer has a positive effect on the quantity, quality, and efficiency of supplier innovation, measured by patent counts, the number of citations per patent, and the ratio of patent counts to the R&D investment accumulated (and depreciated) over the last five years, respectively. We verify that our baseline results are robust to using alternative proximity measures, such as the inventor-location–based distance (Lychagin et al. 2016), the shortest point-to-point traveling time (Catalini et al. 2016, Giroud 2013), and a binary distance measure designed to capture the possible nonlinear effect of proximity on innovation (e.g., Alam et al. 2014, Knyazeva et al. 2013, Malloy 2005). We also confirm that about 70% of firm inventors in our sample work in corporate headquarters that are affected by relocations, and therefore the exogenous shocks we employ are likely to result in substantial relocations of these key innovation employees as well.

To further establish causality, we address various concerns of our baseline identification strategy. First, while customers are much larger than their suppliers and hence customers are unlikely to relocate simply for reasons related to the innovation of suppliers, we cannot completely rule out this possibility if we do not exactly observe customer relocation reasons. To address this concern, we manually collect exact reasons of customer relocations. We exclude customer relocations caused by reasons that are potentially related to suppliers. Our main results remain similar for the sample of exogenous relocations.

Second, one potential concern is that customer relocations could be correlated with local economic or social conditions that affect supplier innovation, which may not be stated in customers' public announcements and hence cannot be captured by our tests above. To address this possibility, we add state-year fixed effects in our baseline regressions. The inclusion of state-year fixed effects controls for any time-varying, state-level confounding factors. The baseline results remain robust to the inclusion of state-year fixed effects.

Third, customers that change their locations may also experience structural changes, which can potentially affect their business with suppliers. If such structural changes correlate with changes in proximity caused by relocations, they may bias our results. We calculate partial correlations between distance and lagged, contemporaneous, and lead customer characteristics. We find that the partial correlations are all economically small and statistically insignificant, suggesting that the structural changes, if ever they exist, are unlikely to be correlated with the changes in proximity and bias our results.

We do further identification attempts through three falsification tests. In the first two falsification tests, for each pair of supplier and customer in our sample, we create a fictitious customer (supplier) by finding a matched noncustomer (nonsupplier) firm. We find that the effect of proximity between a supplier (customer) and the fictitious customer (supplier) on innovation is mixed and statistically insignificant. We conduct a third falsification test based on the timing of exogenous shocks. Specifically, we obtain the empirical distribution of customer relocation time in our sample and then randomly assign the relocation time (without replacement) to each customer that relocates in our sample period. This approach maintains the distribution of relocation time but disrupts the proper assignment of customer relocation time. We find that these falsely assumed customer relocations have no effect on supplier innovation.

All of these findings suggest that there exists a positive, causal effect of supplier–customer proximity on supplier innovation. Next, we examine plausible underlying channels through which customer–supplier proximity affects the supplier innovation we proposed above. We first examine the feedback channel. Since customer feedback is difficult to measure directly from the data, we perform three tests that are closely linked to the feedback channel. In particular, if customer feedback is truly one of the underlying driving forces, we expect geographic proximity to have a more pronounced effect on supplier innovation when the customer itself is more innovative and when the customer and its supplier are closer in technological space. This is because customer feedback is more relevant and informative in such circumstances. Consistent with our conjecture, we find that the effect of geographic proximity on supplier innovation is stronger when customers have larger R&D expenditures and innovation output. We also find that the effect of proximity on supplier innovation is stronger when the supplier and the customer are closer in technological space. Meanwhile, if customer feedback enhances supplier innovation, it could also lead to more frequent citations of customer patents by the supplier. Indeed, we find that when a customer relocates closer to (away from) a supplier, the supplier's citations to this customer's patents increase (drop) while its citations to other patents remain unaffected. These findings provide evidence that is consistent with the feedback channel.

We next examine the demand channel. We create a measure, *Customer Share*, that is defined as a supplier's sales to a given customer divided by the supplier's total sales in the same year. A higher *Customer Share* therefore implies a stronger demand from the customer. We include *Customer Share* and its interaction with the proximity measure in the baseline regression. We find that the loadings on the interaction term are highly significant and economically large, suggesting that our results are much stronger for customers whose demand constitutes a larger fraction of the supplier's total sales. This evidence is consistent with our conjecture that supplier–customer proximity facilitates innovation through enhanced demand from the customer.

Then, we test the agglomeration channel by exploring a unique feature of the agglomeration effect. Specifically, agglomeration is a local effect, and it mainly affects firms in the same area. As a result, if an agglomeration channel is driving our findings, then our results should become much weaker if we drop the customer relocations in which the customer is either moving to or moving away from the metropolitan areas where the supplier locates. In contrast, we find that our results remain unchanged after these observations are excluded from the baseline regression. This evidence suggests that the positive effect of geographic proximity on supplier innovation is unlikely driven by the agglomeration channel.

Finally, we control for the social-connection measure constructed by Dasgupta et al. (2015) in our baseline regression and our results remain robust to the inclusion of the measure that captures social connections between supplier and customer managers, suggesting that the connection channel is not driving our findings.

This paper contributes to three strands of literature. First, it is related to the literature that examines the effect of geographic proximity on productivity and innovation. Eaton and Kortum (1996) construct and estimate a model of how innovation affects economic growth. In their model, geographic proximity affects the diffusion of ideas. Adams and Jaffe (1996), Lychagin et al. (2016), and Keller (2002) investigate how R&D spillovers affect productivity at the plant, firm, and country level, respectively. Orlando (2004) provides evidence that geographic localization of R&D spillovers may be partially driven by other agglomerative forces. These studies focus on how innovation spurs productivity through the channel of R&D spillover and the geographic proximity facilitates the spillover. Our paper contributes to this literature by documenting how geographic proximity enhances firm innovation. To this end, our paper is closely related to that of Catalini et al. (2016), in which it is document that a reduction in travel cost significantly enhances scientific collaboration and innovation. Our paper, however, focuses more on how the effects of geographic proximity on innovation take place through the supply chain.

Second, our paper contributes to the emerging literature on finance and innovation. The theoretical work by Ferreira et al. (2014) builds on Manso (2011) and shows that private ownership structure is more tolerant of failures and thus facilitates innovation. Recent empirical research testing the implications of Manso (2011) includes Azoulay et al. (2011), who explore key differences across funding streams within the academic life science; Ederer and Manso (2013), who conduct a controlled laboratory experiment; Tian and Wang (2014), who show that IPO firms financed by more failure tolerant venture capital investors are more innovative after they go public; and Chemmanur et al. (2014), who find that corporate venture capital firms (CVCs) are more failure tolerant than independent venture capital firms (IVCs). Other firm and market factors that may or may not tolerate failure are also shown to affect innovation—for example, bankruptcy laws (Acharya and Subramanian 2009), labor laws and unions (Acharya et al. 2013, 2014; Bradley et al. 2017), stock liquidity (Fang et al. 2014), banking competition (Cornaggia et al. 2015), firm boundaries (Bena and Li 2014, Seru 2014), management compensation (Baranchuk et al. 2014), financial development (Hsu et al. 2014), governance (Chemmanur and Tian 2018), nonexecutive compensation (Chang et al. 2015, Brav et al. 2018), and analyst coverage (He and Tian 2013). However, the existing literature has ignored the important role played by timely feedback in motivating innovation, as discussed in Manso (2011). Our paper provides the first empirical analysis of the feedback mechanism and highlights the importance of timely feedback and intensive interactions and collaborations between customers and suppliers in enhancing innovation.

Third, our paper contributes to a broader literature on the role of supply-chain relationships in corporate finance. One group of this literature examines how corporate financing and investment decisions affect supply-chain relationships, such as antitakeover measures (Cen et al. 2015b), mergers and acquisitions (Fee and Thomas 2004, Shahrur 2005), cross-ownership (Fee et al. 2006), and financial distress (Hertzel et al. 2008). The other group of literature examines how supplier–customer relationships may affect corporate financing decisions—for example, capital structure (Banerjee et al. 2008, Kale and Shahrur 2007, Chu 2012) and the cost of debt (Cen et al. 2015a). However, the impacts of supplier–customer relationships on corporate real decisions are not well explored, with the only exception of Kale et al. (2011), who investigate how CEO risk-taking incentives affect the motives of customers and suppliers to engage in relationship-specific investments. Our paper tries to fill the gap by examining how supplier–customer proximity affects corporate innovation, an important real decision a firm has to make to keep its competitive advantages.

The rest of the paper is organized as follows. Section 2 describes the data and sample construction. Section 3 presents the baseline results with various tests to address the endogeneity problems. Section 4 examines possible underlying mechanisms, and Section 5 concludes.

2. Data and Sample Construction

2.1. The Sample

Our sample consists of all supplier–customer pairs that can be identified in Compustat between 1976 and 2009. We exclude utility firms (SIC codes from 4900 to 4999) and financial firms (SIC codes from 6000 to 6999) from our sample because these two industries are highly regulated. We also exclude noninnovative firms that file zero patents throughout our sample period. According to the FASB 14 (1976) and 131 (1997), public firms are required to disclose customers who account for at least 10% of total sales, which allows us to identify major customers for a given firm.

A practical difficulty is that, while these disclosures are available in the Compustat segment files, the primary customers are only reported with abbreviated names without any other identifiers. To address this problem, we use a method similar to that of Fee and Thomas (2004) to match the reported customer names to Compustat firms. From the Compustat segment data file, we first exclude all of the customers that are reported as governments, regions, or militaries. We then run a text-matching program to find the potential matches of the reported customer name with the Compustat firm names. The program requires all of the letters in the reported customer name to be sequentially presented in the potential match. To ensure matching accuracy, we manually identify customers from the matched pairs from the text-matching program. If there are multiple potential matches and we cannot choose the unique match by screening the available public information (firm web sites, annual reports, and Google), we conservatively exclude all of these possible firm–customer pairs. Finally, we drop all pairs in which the reported customer is in the retail industry (SIC codes 5200 to 5999), because retail customers are less likely to demand specific products and therefore are less likely to give valuable feedback that can help the suppliers improve their innovation.

Our sample-selection procedure results in a total of 8,645 firm–customer pairs and 35,153 supplier–customer pair years. From the 35,153 pair year observations, we delete any observations for which the total assets or sales are either zero or negative, and firm-year observations with missing data.

While the existing literature typically uses a firm's headquarters reported in Compustat to identify a firm's physical location, the Compustat location data only provide a snapshot of state and county information of firms' headquarters locations. This information is not sufficient to obtain accurate data about corporate headquarters relocations, which we need for our analysis in this paper. To correct this deficiency, we use Compact Disclosure, Corporate Library, and Fortune Magazine to identify corporate headquarters relocations of customer firms. Since our empirical tests below use supplier–customer proximity changes caused by customer firm relocations, we exclude from our sample the observations in which suppliers relocate their headquarters. This is because these observations may cause a confounding effect in our tests, given that supplier relocations are more likely endogenous.

We are able to find 254 customer relocation cases, including 193 cases of cross-city relocations (44 of which are cross-state relocations) and 61 cases of within-city relocations. To capture meaningful changes in distance, we focus on those cross-city relocations.[2] The cross-city relocation sample includes 2,933 firm-year observations, and 1,018 supplier–customer pairs with 869 unique suppliers and 120 unique customers. The relocations are not clustered in time. As shown in Table 1, the number of relocations is almost evenly distributed over time and does not appear to exhibit a strong correlation with business cycles or other economic conditions. The relocations are not clustered geographically either, so firms in our sample are not moving into or out of some specific areas.

We use the relocation data constructed above to test the effect of supplier–customer proximity on supplier innovation in our baseline regressions. A common concern of this identification strategy is that customers' relocations may be endogenous and be possibly related to their suppliers. Therefore, it is important to understand the exact reasons for corporate relocations. To this end, we search news from Factiva, LexisNexis, and corporate websites for the exact reasons of customer relocations.

Among all of the relocation cases, we are able to find relocation reasons for 45 cases. We summarize these relocation reasons into nine main categories in Table 1: (1) move close to customers, (2) move close to suppliers, (3) retain or attract top executives, (4) low cost, (5) low real estate and living cost, (6) internal restructuring, (7) merger and acquisition related, (8) local government incentives, and (9) reduce travel cost. Among these categories, three categories—moving close to suppliers, local government incentives, and reducing travel cost—are potentially related to supplier unobservable characteristics. To address the potential concern of endogenous relocations, we exclude from our baseline regressions the relocation cases that fall into these three categories and the cases for which we cannot clearly identify the underlying relocation reasons.

Table 1. The Distribution of Customer Relocations

Years	Number of relocations	Moving reason	Number of relocations
1976–1979	5	Close to customer	2
1980–1984	28	Close to supplier	1
1985–1989	32	Retain or attract top executives	2
1990–1994	31	Low cost	12
1995–1999	53	Low real estate or living cost	2
2000–2004	28	Internal restructuring	15
2005–2009	16	M&A related	9
		Local government incentive	1
		Reduce travel cost	1
		Unknown	148

Notes. This table reports the distribution of customer relocations over years and for different reasons. The relocations are identified by searching Compact Disclosure, Corporate Library, and Fortune Magazine. The reasons of relocations are identified by news searching of Factiva, LexisNexis, and corporate websites.

2.2. Variable Measurement

2.2.1. Measuring Innovation. We construct innovation variables using the National Bureau of Economic Research (NBER) patent citation database initially created by Hall et al. (2001). This database provides detailed information on more than three million patents granted by the United States Patent and Trademark Office (USPTO) from 1976 to 2006. The patent database provides information on patent assignee names, three-digit patent technology classes, and the number of future citations received by each patent. We then augment the NBER database with the Harvard Business School (HBS) Patent Network Dataverse to extend the coverage to 2010.

Based on the augmented patent database, we construct two measures for innovation output. The first measure is the number of patent applications filed in a year that are eventually granted. This measure captures the quantity of innovation output. To capture the quality of innovation output, we construct a second measure by counting the total number of future citations a patent receives in subsequent years.

Following the existing literature, we adjust the output measures for two types of truncation problems. The first truncation problem arises as patents appear in the database only after they are granted, and it may take several years for the USPTO to approve a patent. For example, if one firm files a patent application in 2009 and it is approved in 2011, the patent will not be included in our measure of patent output in 2009. To adjust this truncation bias, we follow Hall et al. (2001) to use the "weight factors" computed from the application-grant empirical distribution to adjust patent counts. The second truncation problem arises as patents keep receiving citations over a long period, but we only observe the citations received up to 2010. We follow Hall et al. (2001) to adjust the truncation bias in citation counts by using the citation-lag distribution.

In addition to the two innovation output measures described above, we construct an innovation efficiency measure, which captures innovation output per unit of input, in which the innovation input is measured by R&D capital accumulated over the previous five years. Specifically, we follow Hirshleifer et al. (2013) to define accumulated R&D capital as the sum of R&D investment that is depreciated by an annual rate of 20% in the previous five years.

Finally, as previous literature shows, patent counts and citation counts are right skewed. We therefore use the natural logarithm of one plus patent counts

(*LnPatents*), one plus citation counts (*LnCites*), and one plus innovation efficiency (*LnIE*) as the innovation measures in our analysis.

2.2.2. Measuring Distance and Control Variables. We construct the distance variable as the geographic proximity between the headquarters of the supplier and the headquarters of the customer. We collect information on historical headquarters addresses from Compact Disclosure and Fortune Magazine to augment the current headquarters address information in Compustat (Pirinsky and Wang 2006). For each supplier and customer, we obtain the pair of latitude and longitude coordinates for the addresses of their headquarters. Because of the earth's near-spherical shape (technically an oblate spheroid), calculating an accurate distance between two points requires the use of spherical geometry and trigonometric math functions.[3] Because the distribution of distance is right skewed, we compute the natural logarithm of the distance (*LnDistance*) and use it as the main variable of interest.

Though it is common in the literature to use corporate headquarters as the main location of firm production and operation (e.g., Barrot and Sauvagnat 2016), one potential concern is that not all firm activities are concentrated at the headquarters locations. To address this concern, we check whether innovation activities are concentrated at the firm's headquarters. We collect individual inventor data, especially the inventor location information, from the HBS patent and inventor database, and then calculate the distance from an inventor's location to the firm headquarters location. We find that: (1) the median inventor-to-headquarters distance for supplier firms is about 22 miles, and about 70% of supplier inventors live within 120 miles (about two hours' drive) to supplier headquarters; (2) the median inventor-to-headquarters distance for customer firms is about 30 miles, and about 60% of customer inventors live within 120 miles of customer headquarters. These results suggest that most innovation activities do concentrate at firm headquarters, and therefore it is reasonable to use the distance between the headquarters of supplier firms and customer firms to measure the ease of timely feedback from customers to suppliers.

As additional robustness checks, we construct three alternative measures of geographic proximity. The first alternative proximity measure follows Lychagin et al. (2016), which explicitly takes into account the distribution of inventor locations. We retrieve inventor address information from the HBS patent database and use those addresses to compute a weighted average distance among inventors in the supplier firm and those in the customer firm. The advantage of this measure is that it captures the effect of proximity at the inventor level instead of the corporate headquarters level. The second alternative measure follows Catalini et al. (2016) and Giroud (2013), who consider the ease of traveling. For example, because of the availability of direct flights among large cities, traveling between large cities takes less time than traveling between small cities, even if the geographic distance between large cities may be longer than that between small cities. Using data from the Bureau of Transportation Statistics T-100 form, we define a minimum point-to-point traveling time as the shortest duration of traveling between the headquarters of the supplier and that of the customer among all possible routes, including local travel time from/to the airport and flight duration. This measure allows us to gauge the actual cost of distance more accurately. The third alternative proximity measure we use acknowledges that the effect of proximity might be nonlinear. Hence, following Alam et al. (2014), among others, we define a dummy variable that equals one if the supplier is more than 200 miles away from the customer, and zero otherwise. We use these alternative proximity measures in the baseline regressions to check the robustness of our main findings.

We follow the existing literature and control for a vector of firm characteristics that may affect a firm's innovation output. The control variables include *R&D* (R&D expenditure divided by total assets), *LnAssets* (natural logarithm of total assets), *ROA* (operating income divided by total assets), *Q* (market value of assets divided by book value of total assets), *Leverage* (total debt divided by market value of assets), *Sales Growth* (growth rate of sales), *Cash* (cash holdings divided by total assets), *Tangibility* (total property, plant, and equipment divided by total assets), *Cap Ex* (capital expenditures divided by total assets), and *LnAge* (natural logarithm of years listed in Compustat). In some specifications, we also include customer characteristics, which are similarly defined as the supplier variables. To gauge the strength of the supplier–customer relationship, we also control for a pair-specific variable, *Customer Share* (the customer's demand of the supplier's output divided by the supplier's total sales). We report all variable definitions in Table 2.

2.3. Summary Statistics

Table 3 provides summary statistics of the variables used in this study. An average supplier produces about 14 patents a year, and each patent receives nine future citations. These numbers are higher than those typically reported in previous innovation studies using Compustat firms for two possible reasons. First, we focus only on innovative suppliers—i.e., suppliers produced at least one patent over the sample period. Second, by construction, suppliers in our sample have large customers and are more likely to make relationship-specific investment (Kale and Shahrur 2007, Banerjee et al. 2008, Chu 2012), which results in a higher level of innovation output.

Table 2. Variable Definitions

Variable	Definition
LnPatents	Natural logarithm of one plus the number of patents filed (and eventually granted) of the supplier
LnCites	Natural logarithm of one plus the number of citations received on the supplier's patents filed (and eventually granted)
LnIE	Natural logarithm of one plus the ratio of number of patents to accumulated R&D expense (xrd + 0.8 xrd $(t-1)$ + 0.6 xrd $(t-2)$ + 0.4 xrd $(t-3)$ + 0.2 xrd $(t-4)$)
LnDistance	Natural logarithm of the geographic distance between the headquarters of the supplier and its customer
Technology Proximity	$(S'\Omega C)^2/(S'S)(C'C)$, where S and C are vectors of the ratios of patents awarded in patent classes to total patents
R&D	R&D expense divided by total assets
Q	Market value of total assets to book value of total assets
ROA	Net income divided by total assets
Leverage	Book value of total debt divided by market value of total assets
LnAssets	Natural logarithm of total assets
Sale Growth	The growth rate of sales
Cash	Cash holding divided by total assets
Tangibility	Total property, plant, and equipment divided by total assets
Cap Ex	Capital expenditure divided by total assets
LnAge	Natural logarithm of the number of years in Compustat
Asset Turnover	Sales divided by total assets
Profit Margins	Net income divided by total sales
Customer Share	The customer's demand of the supplier's output divided by the supplier's total sales
LnConnection	Natural logarithm of one plus the social-connection measure

The average distance between a supplier and its customer is 912 miles with a standard deviation of 897 miles. Customer relocations lead to significant changes in the proximity between customers and their suppliers: in the subsample of relocations in which customers move closer to their suppliers, average distance reduces by 460 miles, and in the subsample of relocations in which customers move away from their suppliers, average distance increases by 250 miles. Both of them are sizable compared with the average distance between suppliers and customers in our sample.

All firm characteristics are comparable to those reported in existing studies. Comparing the summary statistics of variables of suppliers with those of customers, one observation stands out—customer firms are much larger than supplier firms, and in fact they are about 120 times larger than supplier firms on average. This feature of the data is critical for our identification strategy because these large customers are unlikely to change headquarters locations simply because of factors related to their much smaller suppliers.

3. Empirical Results

In this section, we first discuss our baseline specifications and present the baseline results in a general difference-in-differences framework. We then address some potential concerns regarding our identification strategy. We also conduct a variety of falsification tests, lending strong support to our baseline results.

3.1. Baseline Specifications and Results

Identifying the causal effect of supplier–customer proximity on supplier innovation is challenging, because geographic concentration and economic outcomes are often jointly determined. Specifically, in our setting, location choices of suppliers or customers and innovation activities could be simultaneously determined by some unobservables, leading to biased inferences from the standard ordinary least squares (OLS) regressions in which innovation measures are regressed on proximity measures.

To help establish causality, our baseline identification uses a difference-in-differences approach that explores customer headquarters relocations as a plausibly exogenous shock to the geographic proximity between suppliers and their major customers. Our identification strategy relies on one critical feature of the U.S. supplier–customer relationships observed in the Compustat segment customer database—that is, customers in our sample are much larger than their suppliers (more than 100 times larger on average). Arguably, headquarters relocation decisions made by those large customers are unlikely to be driven by their suppliers that are much smaller in size.

Specifically, we estimate the following model:

$$Innovation_{i\tau} = \beta\, LnDistance_{ijt} + \gamma' X_{ijt} + Year_t + Pair_{ij} + \varepsilon_{ijt}, \qquad (1)$$

where i indexes supplier firm, j indexes customer firm, and t indexes time. The dependent variable in this model is our measure of the supplier's innovation quantity (*LnPatents*), quality (*LnCites*), or efficiency (*LnIE*), measured at $\tau = t+1$, $t+2$, or $t+3$. X_{ijt} is a vector of supplier and customer characteristics. We include both the year fixed effects, $Year_t$, and supplier–customer pair fixed effects, $Pair_{ij}$, in our regressions. This specification is a generalized difference-in-differences specification because the variation in $LnDistance_{ijt}$ only comes from the supplier–customer pairs in which customer headquarters relocations occur. For supplier–customer pairs in which

Table 3. Summary Statistics

Variable	Obs.	Mean	Std. dev.	P25	Median	P75
Supplier						
Patent	6,254	17.919	59.581	0.000	1.000	4.095
Cite	6,254	9.687	21.157	0.000	0.000	12.023
Innovation Efficiency	5,516	0.271	0.636	0.000	0.023	0.148
Q	6,254	1.954	2.056	0.772	1.212	2.297
R&D	6,254	0.101	0.140	0.011	0.054	0.135
ROA	6,254	0.038	0.248	−0.001	0.104	0.166
Leverage	6,254	0.199	0.232	0.006	0.107	0.332
LnAssets	6,254	4.873	1.909	3.553	4.745	6.108
Sales Growth	6,254	0.295	1.080	−0.045	0.105	0.311
Cash	6,254	0.242	0.245	0.036	0.153	0.389
Tangibility	6,254	0.243	0.176	0.096	0.209	0.361
Cap Ex	6,254	0.062	0.058	0.024	0.046	0.080
LnAge	6,254	2.310	0.676	1.792	2.303	2.833
Customer						
Patent	6,254	381.446	631.831	14.000	166.000	462.000
Q	5,426	1.521	1.491	0.638	0.925	1.810
R&D	5,788	0.055	0.043	0.024	0.049	0.074
ROA	6,246	0.141	0.082	0.087	0.127	0.185
Leverage	5,426	0.295	0.279	0.075	0.198	0.418
LnAssets	6,254	9.842	1.819	8.707	10.100	11.064
Sales Growth	6,225	0.113	0.202	0.016	0.086	0.168
Cash	6,253	0.121	0.117	0.047	0.079	0.153
Tangibility	6,254	0.245	0.154	0.121	0.233	0.348
Cap Ex	6,223	0.066	0.050	0.029	0.054	0.088
LnAge	6,254	2.310	0.676	1.792	2.303	2.833
Supplier–customer pair						
Distance	6,254	911.878	896.833	150.802	536.208	1619.296
Customer Share	6,254	0.247	0.211	0.130	0.180	0.293
Technology Proximity (Mal)	6,674	0.046	0.072	0.010	0.019	0.046

Notes. This table reports the summary statistics for variables used in this paper. *Patent* is the number of patents filed (and eventually granted); *Cite* is the number of citations received on the patents filed; *Innovation Efficiency* is the ratio of number of patents to accumulated R&D expense $(\text{xrd} + 0.8\ \text{xrd}(t-1) + 0.6\ \text{xrd}(t-2) + 0.4\ \text{xrd}(t-3) + 0.2\ \text{xrd}(t-4))$; *Q* is market value of total assets to book value of total assets; *R&D* is R&D expense divided by total assets; *ROA* is the operating income divided by total assets; *Leverage* is the book value of total debt divided by market value of total assets; *Sales Growth* is the growth rate of sales; *Cash* is the cash holding divided by total assets; *Tangibility* is total property, plant, and equipment divided by total assets; *Cap Ex* is the capital expenditure divided by total assets; *LnAge* is the natural logarithm of the number of years in Compustat; *Distance* is the geographic distance (in miles) between the headquarters of the supplier and its customer; and *Technology Proximity* is the Mahalanobis measure computed as $((S'\Omega C)^2)/((S'S)(C'C))$, where S and C are vectors of the ratios of patents awarded in patent classes to total patents for suppliers and customers. *Customer Share* is the customer's demand of the supplier's products divided by the supplier's total sales.

customers' headquarters locations remain unchanged in our sample period, $LnDistance_{ijt}$ is time invariant.

We report the regression results estimating Equation (1) in Table 4. Columns (1)–(3) show the regression results for innovation quantity, *LnPatents*, in years $t+1$ to $t+3$. The coefficient estimates on *LnDistance* are all negative and statistically significant, suggesting a negative effect of the geographic distance between the supplier and its major customers on the supplier's future innovation patent counts. The economic effect is sizeable: a one-standard-deviation increase in distance from its mean leads to a 7% decrease in the number of patents filed in the next year. The results in columns (2) and (3) suggest that the effects extend to patent filings in the next two years.

Columns (4)–(6) show the results for innovation quality measured by patent citations (*LnCites*). Since the dependent variable is only well defined if the supplier produces at least one patent in the corresponding year, we therefore exclude all firm-year observations in which the supplier does not produce any patent. The coefficient estimates on *LnDistance* are again negative and statistically significant in all three columns, suggesting that a long distance between a supplier and its major customer negatively affects the quality of its patents generated in the subsequent years. The effect is also economically large: a one-standard-deviation increase in the distance from its mean leads to a 12.5% decrease in the number of citations per patent in the following year.

Lastly, columns (7)–(9) report the results for innovation efficiency, which is measured by innovation output (patents) per unit of innovation input (R&D stock). We exclude all firm-year observations in which the supplier has zero total R&D expenses over the last five years because the accumulated R&D expenses appears

Table 4. Baseline Regression Results

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.072**	−0.052*	−0.038*	−0.131***	−0.045*	−0.206***	−0.055	−0.070**	−0.119**
	(0.030)	(0.027)	(0.020)	(0.040)	(0.027)	(0.026)	(0.050)	(0.033)	(0.048)
Customer Share	0.017	0.028	−0.054	0.028	−0.014	0.143	−0.065	0.004	−0.141
	(0.082)	(0.089)	(0.163)	(0.060)	(0.069)	(0.327)	(0.081)	(0.066)	(0.263)
Q	0.019	0.019	0.025*	−0.002	0.046**	0.001	0.016	0.002	−0.000
	(0.015)	(0.014)	(0.014)	(0.022)	(0.019)	(0.026)	(0.022)	(0.023)	(0.025)
R&D	0.601**	0.572*	−0.260	0.213	0.557	0.017			
	(0.273)	(0.294)	(0.292)	(0.532)	(0.537)	(0.682)			
ROA	−0.068	0.099	0.035	0.031	0.275	0.292	0.067	0.024	0.089
	(0.120)	(0.141)	(0.128)	(0.296)	(0.323)	(0.342)	(0.227)	(0.310)	(0.296)
Leverage	−0.216	−0.300*	−0.302*	−0.300	0.364	0.202	−0.578**	−0.341	−0.076
	(0.155)	(0.159)	(0.168)	(0.220)	(0.259)	(0.258)	(0.294)	(0.313)	(0.341)
LnAssets	0.309***	0.236***	0.164***	−0.041	0.022	−0.029	−0.299***	−0.192*	−0.120
	(0.049)	(0.058)	(0.062)	(0.086)	(0.077)	(0.087)	(0.106)	(0.103)	(0.108)
Sales Growth	−0.003	−0.022	−0.017	−0.030	−0.019	−0.014	0.047	0.017	0.022
	(0.016)	(0.017)	(0.026)	(0.025)	(0.022)	(0.051)	(0.030)	(0.041)	(0.039)
Cash	0.089	0.009	−0.088	−0.008	−0.105	0.137	0.238	−0.151	−0.179
	(0.183)	(0.202)	(0.219)	(0.257)	(0.347)	(0.307)	(0.285)	(0.345)	(0.349)
Tangibility	0.214	0.087	−0.014	−0.585	−0.846	−0.361	0.197	0.085	−0.574
	(0.353)	(0.355)	(0.360)	(0.642)	(0.648)	(0.619)	(0.837)	(0.936)	(0.905)
Cap Ex	−0.430	−0.545	−0.389	0.280	−0.011	0.052	−0.302	−1.214	−0.184
	(0.383)	(0.377)	(0.384)	(0.659)	(0.727)	(0.768)	(0.685)	(0.814)	(0.919)
LnAge	0.288**	0.227	0.147	−0.357*	−0.213	−0.231	0.237	0.161	0.092
	(0.143)	(0.170)	(0.168)	(0.190)	(0.230)	(0.194)	(0.323)	(0.353)	(0.279)
Customer R&D	0.188	0.028	0.244	−0.029	−0.124	−0.914	0.300	−0.680	0.510
	(0.281)	(0.438)	(0.547)	(0.751)	(0.675)	(1.249)	(0.705)	(0.684)	(0.987)
Customer LnAssets	−0.072	−0.043	−0.074	0.023	−0.004	0.000	−0.028	−0.120	−0.239**
	(0.061)	(0.070)	(0.076)	(0.098)	(0.094)	(0.104)	(0.125)	(0.118)	(0.114)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.845	0.830	0.831	0.798	0.793	0.793	0.879	0.878	0.888

Notes. This table reports the baseline regression results of the model $Innovation_{i\tau} = \alpha_{ij} + \alpha_t + \beta\, LnDistance_{ijt} + \gamma' X_{it} + Year_t + Pair_{ij} + \varepsilon_{ijt}$. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Definitions of variables are listed in Table 2. Year fixed effects and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

on the denominator of the innovation efficiency measure. The coefficient estimates on *LnDistance* are negative in all three columns and are statistically significant in columns (8) and (9). The evidence suggests that a supplier's distance from its major customers negatively affects the supplier's innovation efficiency, especially in two and three years.

Overall, our baseline results show that the distance between a supplier and its major customer has a significant effect on the supplier's innovation output. Suppliers' innovation quantity, quality, and efficiency all rise significantly after their major customers relocate closer to them. The effect is persistent in the next three years following the relocation, suggesting that the effect of changes in supplier–customer proximity on supplier innovation activities and output is long-lasting.

3.2. Alternative Proximity Measures

In our baseline results, we use the geographic distance between a supplier and its major customer as the main variable of interest. In this section, we construct three alternative proximity measures and examine whether our findings are robust to these alternative measures.

3.2.1. Inventor-Distribution–Based Proximity Measure. We construct the first alternative measure following the idea proposed in Lychagin et al. (2016). The intuition is that some plausible channels we discuss before (e.g., feedback channel and agglomeration channel) rely on the proximity between inventors rather than corporate headquarters. Hence, it is important to consider inventor locations. We retrieve inventors' address information from the HBS patent database and create the measure as follows. Suppose that there are K geographic regions, $k = 1, 2, \ldots, K$. Let F_{ik}^G be the fraction of firm i's inventors that are located in region k. We define the inventor-distribution–based distance between firm i and firm j as

$$LnInvDist = \sum_k \sum_l w_{ijkl} \cdot d_{kl}, \quad i \neq j,$$

where d_{kl} is the Euclidean distance between regions k and l, and the weight w_{ijkl} is a function depending on the F_{ik}^G's:

$$w_{ijkl} = F_{ik}^G \cdot F_{jl}^G.$$

This measure computes a weighted average of geographic distance based on the distribution of inventors in the supplier and customer firms.

To show that the inventor-distribution–based distance measure in fact affects supplier innovation, we run regressions of the innovation measures on *LnInvDist*. The results are presented in panel A of Table 5.[4] The coefficient estimates on *LnInvDist* are all negative and statistically significant, which is consistent with our baseline results. This test suggests that our main results are robust to an alternative inventor-distribution–based distance measure.

3.2.2. Minimum Point-to-Point Traveling Time. Given that the average distance between a supplier and its customer in our sample is about 900 miles, one may argue that the minimum point-to-point traveling time is a better measure of the cost or friction induced by geographic distance (see e.g., Catalini et al. 2016, Giroud 2013). We therefore compute the minimum point-to-point traveling duration between the supplier and the customer and use it as an alternative measure of geographic distance.

We construct this alternative measure using data from the U.S. Department of Transportation's Bureau of Transportation Statistics T-100 form. We define minimum point-to-point traveling time as the shortest duration of traveling between the headquarters of the supplier and that of the customer among all possible

Table 5. Alternative Proximity Measures

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Inventor-distribution–based measure									
LnInvDist	−0.085***	−0.094***	−0.092***	−0.103**	−0.101**	−0.095**	−0.105***	−0.115***	−0.118***
	(0.032)	(0.033)	(0.035)	(0.041)	(0.043)	(0.043)	(0.032)	(0.037)	(0.028)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,052	5,804	5,498	3,038	2,843	2,666	2,911	2,730	2,558
R^2	0.905	0.893	0.916	0.805	0.792	0.823	0.846	0.857	0.879
Panel B: Minimum point-to-point traveling time									
LnMPtP	−0.082***	−0.064***	−0.045*	−0.121***	−0.115***	−0.116***	−0.073*	−0.078***	−0.115***
	(0.027)	(0.024)	(0.025)	(0.030)	(0.031)	(0.023)	(0.042)	(0.029)	(0.037)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.824	0.835	0.844	0.756	0.777	0.796	0.842	0.853	0.847
Panel C: Dummy distance measure									
DummyDist	−0.154***	−0.179***	−0.162**	−0.546***	−0.467***	−0.489***	−0.428**	−0.439**	−0.478**
	−0.062	(0.63)	(0.076)	(0.139)	(0.145)	(0.164)	(0.204)	(0.213)	(0.224)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.845	0.830	0.831	0.798	0.793	0.793	0.879	0.878	0.888

Notes. This table reports the baseline regression results using alternative proximity measures. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Definitions of variables are listed in Table 2. The independent variable of interest includes the alternative proximity measures. Specifically, *LnInvDist* (panel A) is the logarithm of the inventor-distribution–based proximity measure, *LnMPtP* (panel B) is the logarithm of minimum point-to-point traveling time, and *DummyDist* (panel C) is the dummy variable that equals one if supplier–customer distance is larger than 200 miles, and zero otherwise. Common controls, year fixed effects, and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

routes, including local travel time from/to the airport and flight duration.

Not surprisingly, the minimum point-to-point traveling duration is highly correlated with geographic distance, with a correlation of 0.93 in our sample. The average minimum point-to-point traveling duration is 182 minutes (i.e., about three hours) with a standard deviation of 96 minutes. The median is slightly lower at 153 minutes.

We then rerun our baseline regression, replacing the geographic distance with the minimum point-to-point traveling duration. We report the results in panel B of Table 5. The main variables of interest, the coefficients on the logarithm of minimum point-to-point traveling time, are all negative and statistically significant. This finding confirms that our baseline results are robust to the distance measure constructed based on the minimum point-to-point traveling time.

3.2.3. Distance Dummy Measure. It is also possible that the effect of distance on supplier innovation is nonlinear. Previous studies (e.g., Tian 2011, Alam et al. 2014, Knyazeva et al. 2013, Malloy 2005) use a distance dummy variable with cutoff points varying in different studies, ranging from 50 miles to 200 miles. To test whether our findings are robust to the discretized measure of distance, we define a dummy variable to indicate whether the distance between a supplier and its customer allows for easy face-to-face soft information production. We define the dummy variable to be one if the distance is larger than 200 miles and zero otherwise. Panel C of Table 5 presents the results. Consistent with our main findings, when customers relocate into (out of) the soft information production zone of their suppliers, suppliers' innovation activities and output increase (decrease).

Overall, the above tests confirm that our baseline results are robust to alternative proximity measures that meant to capture the distribution of inventors, the actual cost of traveling, and the nonlinearity effect of proximity, respectively.

3.3. Additional Identification Attempts

In this subsection, we undertake additional analyses to address several potential concerns regarding the identification strategy adopted in our baseline regressions.

We first show that our baseline results continue to hold when we restrict our analysis to a subsample in which the reasons of customer relocations can be clearly identified as exogenous. We then control for local economic conditions that can possibly create spurious correlations between the supplier–customer distance and suppliers' innovation. We also show that the results are unlikely to be driven by structural changes of the customers accompanying headquarters relocations.

Next, we conduct three falsification tests to demonstrate that the positive effect of supplier–customer proximity on supplier innovation identified in the baseline analysis is absent when we artificially assign supplier–customer pairs or artificially assign customer relocation years.

Finally, we adopt an independent identification strategy to verify our main results. We follow Catalini et al. (2016) and Giroud (2013), and use the addition of airline routes between the supplier and customer headquarters as plausibly exogenous shocks to supplier–customer proximity. This identification strategy delivers consistent findings to our baseline results.

3.3.1. Addressing Endogenous Customer Relocations. The key identification assumption in our baseline tests is that customers' relocation decisions are uncorrelated with factors that may potentially affect a supplier's innovation activities. Though the large discrepancy in size between the customers and their suppliers helps mitigate this concern, we cannot completely rule out this possibility without knowing the exact reasons of customer relocations. We thus search through different sources such as Compact Disclosure, Corporate Library, and Fortune Magazine to manually collect the reasons of corporate headquarters relocations of customer firms. As discussed in Section 2.1, we are able to find relocation reasons for 45 cases, and we summarize the relocation reasons into nine main categories in Table 1. Among these categories, three categories—moving close to suppliers, local government incentives, and reducing travel cost—are potentially related to supplier unobservable characteristics. We exclude the relocation cases falling into these three categories and the relocation cases for which we cannot clearly identify their moving reasons.

We then reestimate Equation (1) in this restricted sample and report the results in panel A of Table 6. Similar to Table 4, we report results for innovation quantity (*LnPatents*) in columns (1)–(3), innovation quality (*LnCites*) in columns (4)–(6), and innovation efficiency (*LnIE*) in columns (7)–(9). The coefficient estimates on *LnDistance* are negative and significant at the 5% or 1% level in all regressions, and their magnitudes remain similar and economically sounded. This finding suggests that our baseline results are unlikely to be driven by customer relocation decisions that are correlated with supplier innovation activities.

One remaining concern is that even if we exclude customer relocations for stated reasons that are likely to be correlated with supplier innovation activities, customers may still move for reasons that are not publicly stated but are related to supplier innovation. Local economic conditions, for example, could be such an unstated relocation reason.

To address this concern, we add *State* × *Year* fixed effects in our baseline regressions. Including *State* × *Year* fixed effects controls for any time-varying, con-

Table 6. Addressing the Potential Endogeneity of Customer Relocation Decisions

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Excluding relocations related to supplier and for unknown reasons									
LnDistance	−0.059***	−0.035***	−0.034***	−0.114***	−0.037**	−0.208***	−0.028***	−0.064***	−0.140***
	(0.011)	(0.009)	(0.011)	(0.011)	(0.015)	(0.021)	(0.011)	(0.015)	(0.021)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,696	5,461	5,193	2,953	2,672	2,423	2,701	2,441	2,192
R^2	0.845	0.832	0.833	0.802	0.795	0.793	0.884	0.882	0.892
Panel B: Results with state-year fixed effects									
LnDistance	−0.075*	−0.051*	−0.032	−0.108***	−0.038	−0.254***	−0.063	−0.076*	−0.202***
	(0.045)	(0.029)	(0.029)	(0.033)	(0.050)	(0.036)	(0.039)	(0.044)	(0.041)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.845	0.830	0.831	0.798	0.793	0.793	0.879	0.878	0.888

Notes. This table reports four sets of tests that aim at addressing the potential bias caused by the endogeneity of customer relocation decisions. Panel A reports the regression results of the model in Equation (1) excluding customer relocations that are categorized as being related to the suppliers. Panel B reports the regression results with state/year fixed effects in controls. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Control variables are the same as in Table 4 but are omitted for brevity. Relevant control variables, year fixed effects, and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

founding state-level factors that can affect supplier innovation but are otherwise unobservable. The results with *State* × *Year* fixed effects controlled are presented in panel B of Table 6. The coefficient estimates on *LnDistance* are very similar to those in Table 4, although we lose statistical significance in three out of nine specifications.

3.3.2. Customer Structural Changes Accompanying Headquarters Relocations. Although we argue that customer headquarters relocation decisions are unlikely to be directly related to their suppliers, it is possible that customer headquarters relocations are accompanied by structural changes in customer characteristics. These structural changes can potentially affect supplier innovation through changes in the customers' demands for suppliers' output. If these structural changes are meanwhile correlated with customer headquarters relocations, they may raise a concern on our identification strategy. We therefore first examine whether customer headquarters relocations are possibly accompanied with customer firm structural changes. We then examine whether the structural changes, if they exist, are likely to be correlated with changes in distance.

We first compare key customer firm characteristics one year before and one year after headquarters relocations to examine whether headquarters relocations are possibly accompanied with structural changes. We present the results in panel A of Table 7. Except for return on assets (*ROA*) and capital expenditures (*Cap Ex*), other customer characteristics do not change significantly after headquarters relocations. The results appear to suggest that most relocations are not accompanied by firm structural changes. Though the drops in *ROA* and investment could be temporary (i.e., caused by the interruption to operation during the process of headquarters relocation), they can still pose a challenge to our identification if decreasing operating performance and capital expenditure or other unobservable changes are correlated with supplier–customer proximity.

To examine whether potential structural changes accompanying customer relocations are likely to be correlated with the distance between suppliers and customers, we calculate the partial correlation of the distance with lagged, contemporaneous, and lead customer characteristics using a similar regression framework as in our baseline analyses. Specifically, we run regressions as follows:

$$LnDistance_{ijt} = \delta_0 + \delta' Y_{j\tau} + Year_t + Pair_{ij} + \varepsilon_{ijt}, \qquad (2)$$

where $Y_{j\tau}$ is a vector of customer characteristics measured at τ, and τ takes the value of $t-1$, t, $t+1$, $t+2$, or $t+3$. The specification also includes the year fixed effects and pair fixed effects, which ensures that

Table 7. Customer Relocation and Structural Changes

Panel A: Customer characteristics before and after headquarters relocations				
	Before	After	Difference	*t*-statistic
Patent	122.93	129.29	6.365	0.151
Cite	16.978	18.149	1.171	0.268
Q	2.471	2.523	0.052	0.076
R&D	0.045	0.050	0.005	0.446
ROA	0.159	0.115	−0.044	−1.947*
Leverage	0.216	0.216	0.000	0.000
LnAssets	8.406	8.388	−0.018	−0.063
Sales Growth	0.230	0.113	−0.116	−1.484
Cash	0.129	0.161	0.031	1.193
Tangibility	0.266	0.228	−0.037	−0.977
Cap Ex	0.067	0.048	−0.019	−2.170**

Panel B: Partial correlations between customer characteristics and the distance					
	(1)	(2)	(3)	(4)	(5)
	$t-1$	t	$t+1$	$t+2$	$t+3$
LnAssets	0.025	0.040	0.025	0.003	0.000
	(0.035)	(0.041)	(0.031)	(0.005)	(0.006)
Tobin's Q	−0.011	−0.002	−0.011	−0.001	0.000
	(0.015)	(0.005)	(0.014)	(0.002)	(0.000)
Leverage	−0.186	−0.224	−0.115	0.001	0.011
	(0.252)	(0.271)	(0.154)	(0.047)	(0.012)
ROA	−0.218	−0.071	−0.225	−0.015	−0.056
	(0.289)	(0.122)	(0.307)	(0.018)	(0.058)
Tangibility	0.227	0.388	0.147	0.185	0.132
	(0.195)	(0.543)	(0.215)	(0.175)	(0.171)
R&D	−0.188	−0.197	−0.231	0.033	−0.131
	(0.284)	(0.328)	(0.330)	(0.042)	(0.115)
Patent	−0.000	0.000	−0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cap Ex	−0.010	0.160	−0.011	−0.162	−0.183
	(0.195)	(0.293)	(0.244)	(0.171)	(0.296)
Sales Growth	0.001	−0.001	0.028	0.000	0.008
	(0.001)	(0.001)	(0.046)	(0.001)	(0.014)
Constant	5.798***	5.552***	5.749***	5.717***	5.861***
	(0.267)	(0.276)	(0.216)	(0.117)	(0.072)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	8,206	7,773	8,021	7,932	7,624
Adjusted R^2	0.995	0.994	0.996	0.999	0.999

Panel C: Regressions with additional customer control variables									
	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.074*	−0.064*	−0.055**	−0.042	−0.140***	−0.215***	−0.066*	−0.076**	−0.125***
	(0.044)	(0.038)	(0.023)	(0.042)	(0.042)	(0.024)	(0.037)	(0.037)	(0.038)
Supplier control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Customer control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,479	6,224	5,945	3,397	3,106	2,877	3,096	2,828	2,588
R^2	0.853	0.845	0.842	0.793	0.796	0.803	0.876	0.877	0.885

Notes. This table reports results aimed at addressing the potential problem that customer headquarters relocations are accompanied with customer firm structural changes, which in turn affect supplier innovation. Panel A reports results comparing customer characteristics one year before and one year after headquarters locations. Panel B reports the partial correlations of the natural logarithm of the distance between a supplier and its customer and lagged, contemporaneous, and lead customer characteristics. Panel C report regression results with added customer controls. We run the regressions $LnDistance_{ijt} = \delta_0 + \delta' Y_{j\tau} + Year_t + Pair_{ij} + \varepsilon_{ijt}$, where $Y_{j\tau}$ is a vector of customer characteristics measured at τ, and τ takes the value of −1, 0, 1, 2, or 3. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

we are calculating the partial correlation between customer characteristics and relocation-induced distance changes. Notice that we are only looking for partial correlation but not causality. Therefore, we are not concerned about look-ahead bias when we put lead variables on the right-hand side.

We present the results in panel B of Table 7. None of the coefficient estimates are statistically significant. The results suggest that even if customer headquarters relocations are accompanied with some structural changes, these structural changes are unlikely to be correlated with changes in distance. Therefore, our baseline estimates are not biased by these structural changes.[5]

Finally, to further address the potential effects of structural changes, we add a broad set of customer characteristics to our baseline regressions. If the structural changes are captured by these customer characteristics, controlling for these characteristics should mitigate the confounding effect of structural changes. The results are presented in panel C of Table 7. With additional customer characteristics, the coefficients on *LnDistance* remain negative and statistically significant, which further suggests that the baseline results are unlikely to be driven by customer structural changes accompanying their headquarters relocations.

3.3.3. Falsification Tests. In this subsection, we conduct three sets of falsification tests to provide further evidence to support our main findings.

First, in our regressions above, we exclude all within-city relocation cases and only keep cross-city relocation cases. Our argument is that only cross-city relocations create meaningful changes in the distance between suppliers and customers, and therefore affect supplier innovation. This rationale constructs the first set of falsification test for our baseline results. That is, our results should *not* hold for the subsample of within-city relocations because these relocation cases do not change the supplier–customer distance significantly. We rerun our baseline regressions using within-city relocation cases and report the results in panel A of Table 8. As expected, none of the coefficients are significant, suggesting that it is the change in distance rather than the relocations per se that affects supplier innovation.

Second, if supplier–customer proximity truly affects the suppliers' innovation, this effect has to take place through the real supplier–customer pairs. In other words, we shall not expect to observe any correlation between a firm's innovation output and its distance from another firm that is not its customer. We conduct the second set of falsification tests with two exercises to verify this conjecture. In the first exercise, for each supplier–customer pair observed in data, we take the customer as given and artificially assign a supplier for it. We select the fictitious supplier from the firms that are in the same state, in the same three-digit SIC industry, and have the closest total assets as the real supplier. The match is performed at the time when the real supplier and its customer first report their supplier–customer relationship. We then follow the fictitious supplier–customer pair for the same number of years of the real supplier–customer relationship. We reestimate Equation (1) for the sample of fictitious supplier–customer pairs. Because the fictitious supplier is in the same state as the real supplier, if our main results are driven by local economic conditions, we should still observe the effects on this falsification test. We report the results in panel B of Table 8. In all columns, the coefficient estimates on *LnDistance* have mixed signs and almost all of them are insignificant.

In the second exercise, we take the supplier as given and artificially assign a customer for it. The fictitious customer matches the real customer observed in data in the same industry and have the closest total assets. We rerun the regressions estimating Equation (1) using the geographic distance between the supplier and the fictitious customer firm. We report the results in panel C of Table 8. In all columns, the coefficient estimates on *LnDistance* have mixed signs, and none of them is statistically significant. These two exercises suggest that our baseline results are absent in artificially assigned supplier–customer pairs, supporting our argument that it is the timely feedback provided by customers to their suppliers that drives our baseline results.

Finally, there still exists a potential concern that an omitted variable coinciding with customer relocations could be the true underlying cause of changes in supplier innovation. If this is the case, then the changes in supplier innovation we attribute to customer headquarters relocations reflect merely an association rather than a causal effect. Our baseline identification strategy employs shocks (customer relocations) that affect different firms at different times. Hence, it is unlikely that an omitted variable unrelated to customer relocations would fluctuate every time (or even most of the times) customer relocation occurs. Therefore, our strategy of using multiple shocks due to customer relocations over time mitigates this concern. To further rule out this possibility, we conduct the third set of falsification tests.

Specifically, we begin by obtaining an empirical distribution of the relocation timing of customers in our sample. Next, we randomly assign the customer relocation timing (without replacement) to the customers that actually relocate their headquarters during our sample period. This approach maintains the distribution of customer relocation years from our baseline specification but disrupts the proper assignment of customer relocation years. Therefore, if an unobservable shock occurs at approximately the same

Table 8. Falsification Tests

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Falsification tests with, within-city relocations									
LnDistance	−0.019	−0.017	−0.014	0.009	0.005	0.006	−0.008	−0.015	−0.012
	(0.019)	(0.019)	(0.018)	(0.008)	(0.008)	(0.010)	(0.013)	(0.013)	(0.016)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,338	5,111	4,857	2,802	2,534	2,292	2,567	2,322	2,084
R^2	0.845	0.834	0.835	0.806	0.798	0.795	0.883	0.883	0.889
Panel B: Falsification tests with fictitiously assigned matched suppliers									
LnDistance	−0.015	0.013	0.012	−0.011	0.006	0.009	0.006	0.002	−0.004
	(0.035)	(0.037)	(0.039)	(0.032)	(0.045)	(0.046)	(0.067)	(0.056)	(0.078)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,569	5,214	4,987	3,205	2,945	2,703	2,956	2,665	2,421
R^2	0.736	0.708	0.678	0.785	0.776	0.765	0.768	0.745	0.721
Panel C: Falsification tests with fictitiously assigned matched customers									
LnDistance	0.006	0.009	0.005	−0.005	−0.005	−0.006	−0.009	−0.008	−0.006
	(0.017)	(0.021)	(0.026)	(0.026)	(0.040)	(0.045)	(0.019)	(0.026)	(0.036)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,091	5,825	5,532	3,193	2,887	2,624	2,938	2,648	2,382
R^2	0.848	0.837	0.836	0.801	0.798	0.798	0.882	0.882	0.890
Panel D: Falsification tests with randomized relocation timing									
LnDistance	0.021	0.0054	−0.011	0.027	0.021	0.003	0.011	0.004	0.005
	(0.034)	(0.038)	(0.043)	(0.024)	(0.040)	(0.045)	(0.037)	(0.028)	(0.045)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.845	0.830	0.831	0.798	0.793	0.793	0.879	0.878	0.888

Notes. This table reports four falsification tests. Panel A reports the falsification test results of the model $Innovation_{i\tau} = \beta\, LnDistance_{ijt} + \gamma' X_{it} + Year_t + Pair_{ij} + \varepsilon_{ijt}$ when only within-city relocations are included. Panel B reports the falsification test results of the model with fictitiously assigned suppliers, and panel C reports the falsification test results with fictitiously assigned customers. The fictitious supplier or customer is in the same three-digit industry as the true supplier or customer and is closest in firm size. Panel D reports the falsification test results of the model $Innovation_{i\tau} = \alpha + \beta\, LnDistance_{ijt} + \gamma' X_{it} + Year_t + Pair_{ij} + \varepsilon_{ijt}$ with randomized relocation timing. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Definitions of variables are listed in Table 2. Relevant control variables, year fixed effects, and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

time as the customer relocation years, it should still reside in the testing framework and, thus, have an opportunity to drive the results. However, if no such shock exists, then our incorrect assignments of customer relocation years should weaken our results when we reestimate the baseline tests because, intuitively, the changes in supplier innovation well before or well after the year of customer relocations should not be systematically correlated with the changes in the distance that occurred at the year of relocations.

We report the results in panel D of Table 8. Almost all of the coefficient estimates on *LnDistance* are statistically insignificant, and the magnitudes of coefficient estimates are also small. These nonresults corroborate the notion that our paper's main results are not driven by the omitted-variable problem.

In addition to the falsification tests above, our results remain robust if we control for additional supplier and customer characteristics in the regressions. In fact, the magnitudes of the coefficients on *LnDistance* do not change much when we use different sets of control variables. However, standard errors do change when we increase or decrease the number of control variables, which further suggests that customer relocation decisions are likely exogenous (Roberts and Whited 2013).

Table 9. The Effect of the Introduction of New Airline Routes

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.024***	0.027***	0.023**	0.021**	0.016**	0.016**	0.031***	0.032***	0.033***
	(0.007)	(0.008)	(0.011)	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)	(0.010)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.903	0.885	0.823	0.793	0.758	0.749	0.864	0.853	0.801

Notes. This table reports the baseline regression results of the model $Innovation_{i\tau} = \alpha_{ij} + \alpha_t + \beta\,Treatment_{ijt} + \gamma' X_{it} + Year_t + Pair_{ij} + \varepsilon_{ijt}$. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Definitions of variables are listed in Table 2. The independent variable *Treatment* is equal to one if a new airline route that reduces the travel time between the supplier and the customer headquarters has been introduced. Relevant control variables, year fixed effects, and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates. Significance levels at 5% and 1% levels are denoted by ** and ***, respectively.

3.3.4. Additional Identification Attempt. Our main identification strategy relies on the exogenous shocks to supplier–customer proximity caused by customer relocations. Customer relocations change the geographic distance between a supplier and its customer. In this subsection, we explore another identification strategy following Catalini et al. (2016) and Giroud (2013). Specifically, we examine how the addition of a new airline route that reduces the traveling time between the supplier and its customer affects the supplier's innovation. In this setting, the geographic distance between a supplier and its customer does not change, but the effective cost and friction caused by distance changes when traveling by air becomes more convenient after a new airline route between the supplier and its customer is added.

We collect airline route data from the T-100 Domestic Segment Database and the ER-586 Service Segment Data. We then identify the introduction of new airline routes that reduces traveling time between headquarters of the supplier–customer pairs. We are able to identify 104 cases that reduce the travel time between the supplier–customer pairs significantly.

We then undertake a difference-in-differences regression analysis:

$$Innovation_{i\tau} = \alpha_{ij} + \alpha_t + \beta\,Treatment_{ijt} + \gamma' X_{it} + Year_t + Pair_{ij} + \varepsilon_{ijt}, \tag{3}$$

where the independent variable *Treatment* equals one if a new airline route that reduces the traveling time between the supplier and the customer is added, and zero otherwise.

Table 9 reports the results estimating Equation (3). Consistent with our baseline results, the difference-in-differences estimates (i.e., the coefficients on *Treatment*) are all positive and statistically significant, suggesting that reducing air travel time between suppliers and customers enhances supplier innovation.

4. Possible Mechanisms

In this section, we explore possible economic mechanisms through which the geographic distance between a supplier and its major customers affects supplier innovation. We examine four plausible channels—namely, the feedback channel, the demand channel, the agglomeration channel, and the social-connection channel.

Investigating economic channels is challenging in our setting, because not all channels are easily observable and measurable in the data. Hence, our goal in this section is to provide suggestive evidence that could help advance our understandings of these channels.

4.1. The Feedback Channel

Feedback from customers is unobservable in the data, so we cannot create a direct measure to capture feedback and investigate its effect in our setting. In this section, we explore the feedback channel by performing a few tests that are closely related to the quality and relevance of customer feedback, trying to provide some supporting evidence of this channel.

If it is the feedback from customers that drives our findings, we should expect to observe significant cross-sectional heterogeneity in the results when the importance of customers' feedback varies across firms. In particular, we expect the results to be stronger if

(1) The customers are more innovative by themselves; or

(2) The customers and suppliers employ closely related technologies.

Conjecture (1) is intuitive as illustrated in a simple example: though both general retailers and auto producers could be big customers of tire producers, feedback provided by auto producers will be more valuable in improving the tire producers' innovation than that provided by general retailers. This is because auto producers, who are presumably more innovative than general retailers, know much better what improvement in tires will enhance the performance of autos given their own experiences in producing and improving autos.

The importance of conjecture (2) is motivated by Jaffe (1986), who shows that the effect of knowledge spillovers is stronger between firms that are close in technological space. In our context, if the distance affects supplier innovation through the customers' timely feedback (i.e., more knowledge spillovers from customers to suppliers), the effect should be stronger if the supplier and the customer are close in technological space.

To test the first conjecture, we add two interaction terms in our baseline regressions: the interaction between *LnDistance* and customer R&D expenditures and the interaction between *LnDistance* and the number of patents the customer has. We use customers' R&D expenditure and their patent counts to capture their own innovation intensity.

We present the results in panel A of Table 10. The coefficient estimates on the interaction terms are negative in all columns and statistically significant mainly in regressions in which innovation efficiency is examined. These results suggest that the effect of *LnDistance* on supplier innovation efficiency is stronger when the customers themselves spend more on R&D or produce more innovation output. The evidence is consistent with the argument that timely feedback from the customer affects supplier innovation more when timely feedback is more important.

To test the second conjecture, we construct a measure of technological proximity following Bloom et al. (2013). They propose a Mahalanobis measure to capture the technological proximity, which is defined as follows:

$$\mathit{Technology\ Proximity} = \frac{(S'\Omega C)^2}{(S'S)(C'C)}, \tag{4}$$

where S is a column vector, and each element of S is the ratio of the number of supplier's patents granted in the last three year in a patent class to the total number of supplier's patents granted in the last three years. The column vector C is similarly defined for customer's patents. Ω is a weighting matrix (see the online appendix of Bloom et al. 2013 for technical details on the definition of Ω). Intuitively, the element $\Omega(i, j)$ of the matrix measures the closeness of patents in patent classes i and j. $\Omega(i, j)$ is close to one if patents in classes i and j often appear in the same firm, and is close to zero if patents in classes i and j hardly appear in the same firm. The measure *Technology Proximity* is bounded between zero and one.

The Mahalanobis measure improves the Jaffe's measure (Jaffe 1986) by allowing for spillovers across technology fields. We then add the interaction term between *LnDistance* and *Technology Proximity* to our baseline regressions, and present the results in panel B of Table 10. The coefficient estimates on the interaction term are negative and are statistically significant in most specifications. Theses results suggest that the effect of *LnDistance* on supplier innovation is more pronounced if the supplier and the customer are closer in technological space. Together with the notion that technological proximity facilitates knowledge spillovers (Jaffe 1986), this evidence is consistent with the argument that timely feedback from customers to suppliers has a larger effect on supplier innovation when such feedback is more relevant and of high quality.

The last test we conduct is related to the patent citation. Intuitively, customer feedback is likely to be related to the customers' own patents. This is especially true when customers guide their suppliers in the innovation process and rely on the suppliers to innovate and create new intermediate inputs for their production needs. For example, when Boeing's suppliers follow Boeing's feedback to conduct R&D and innovate, they have to abide by Boeing's standards, most of which are patented. Therefore, as more customer feedback gets incorporated into the supplier's innovation process, we expect to see the supplier cite more frequently the customer's patents in its own patents. Feedback from one customer, however, should not affect the supplier's citation to other patents that are not generated by the customer.

To test this prediction, for each supplier–customer pair, we classify the supplier's patents into two categories: citing patents that include the supplier's patents that cite its customer's patents, and nonciting patents that include the supplier's patents that do not cite its customer's patents. The first type of patent, as we discussed above, is influenced by the customer's feedback to a large extent, while the second type of patent is less sensitive to the feedback channel, ceteris paribus. We use the two types of patents as the dependent variable and rerun the baseline regressions. The results are reported in panel D of Table 10. Consistent with the feedback channel, we find a negative, significant effect of distance on the supplier's number of citing patents, but not the number of nonciting patents. The difference between the two groups of patent citations is economically large and statistically significant.

Overall, these tests provide suggestive evidence that is consistent with the feedback channel. Though we cannot draw a conclusive inference from these tests, they lend support to the feedback effect as a plausible underlying economic channel through which proximity affects supplier innovation.

4.2. The Demand Channel

Models of innovation and strategic competition (e.g., d'Aspremont and Jacquemin 1988, Kamien et al. 1992,

Table 10. The Feedback Channel

Panel A: The effects of customer R&D expense and patents

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.010	0.014	0.061	−0.070	−0.047	−0.172*	0.051	0.079	0.118
	(0.067)	(0.078)	(0.085)	(0.076)	(0.077)	(0.093)	(0.101)	(0.116)	(0.144)
LnDistance × *Ln Customer Patent*	−0.011*	−0.011*	−0.019**	−0.012	−0.001	−0.008	−0.020**	−0.024**	−0.039***
	(0.006)	(0.006)	(0.008)	(0.013)	(0.015)	(0.015)	(0.010)	(0.011)	(0.013)
LnDistance × *Customer R&D*	0.043	−0.184	0.142	0.023	0.085	−0.360	−0.610	−0.595*	−0.828**
	(0.179)	(0.202)	(0.229)	(0.337)	(0.335)	(0.433)	(0.371)	(0.359)	(0.412)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.845	0.831	0.831	0.799	0.794	0.794	0.880	0.879	0.890

Panel B: The effect of technological proximity

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.027	−0.02	−0.019	−0.029	−0.011	−0.02	−0.038	−0.026	0.014
	(0.178)	(0.191)	(0.161)	(0.385)	(0.331)	(0.162)	(0.404)	(0.449)	(0.487)
LnDistance × *Technology Proximity*	−1.343***	−1.045**	−0.295	−1.153**	−0.441	−1.105**	−2.170***	−1.186	−1.853**
	(0.552)	(0.524)	(0.417)	(0.465)	(0.425)	(0.532)	(0.608)	(0.884)	(0.917)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,946	5,694	5,421	3,083	2,785	2,547	2,864	2,585	2,327
R^2	0.847	0.833	0.836	0.800	0.795	0.794	0.871	0.868	0.883

Panel C: Patents citing and not citing customer patents

	Ln Citing Patents			*Ln Nonciting Patents*			*Differences*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	(1) − (4)	(2) − (5)	(3) − (6)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.087***	−0.065***	−0.049***	0.015	0.013	0.011	−0.102***	−0.078***	−0.060***
	(0.023)	(0.024)	(0.018)	(0.044)	(0.039)	(0.025)	(0.026)	(0.028)	(0.021)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes			
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes			
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes			
Observations	6,254	5,988	5,689	3,237	2,926	2,652			
R^2	0.780	0.754	0.745	0.833	0.796	0.778			

Notes. This table reports regression results of testing the feedback channel. Panel A reports the regression results of the model $Innovation_{i\tau} = \beta_1 LnDistance_{ijt} + \beta_2 \times LnDistance \times Ln\ Customer\ Patent + \beta_3 \times LnDistance \times Customer\ R\&D + \gamma_1 X_{it} + \gamma_2 Y_{jt} + Year_t + Pair_{ij} + \varepsilon_{ijt}$. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Two interaction terms between *LnDistance* and *Ln Customer Patent*, *Customer R&D* are included in the regressions. Definitions of variables are listed in Table 2. Panels B and C report the regression results of the model $Innovation_{i\tau} = \alpha + \beta_1 LnDistance_{ijt} + \beta_2 \times LnDistance \times Technology\ Proximity + \gamma_1 X_{it} + \gamma_2 Y_{jt} + Year_t + Pair_{ij} + \varepsilon_{ijt}$. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). The interaction term between *LnDistance* and *Technology Proximity* is included in the regressions. Panel B uses the technological proximity measure developed in Jaffe (1986), and panel C uses the Mahalanobis measure of technological proximity developed in Bloom et al. (2013). Panel D uses the citing and nonciting patents as the dependent variables and reruns the baseline regression. Citing patents are the supplier's patents in which the customer's patents are cited, and nonciting patents are those in which the customer's patents are not cited. Relevant control variables, year fixed effects, and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Leahy and Neary 1997) posit that a firm's incentive to innovate is proportional to its production quantity. More specifically, a supplier's incentive to innovate for a given customer may be proportional to its sales to the customer (i.e., the demand channel). Proximity reduces transportation costs and for a given price plausibly increases the demand from the customer. This, in turn, motivates supplier innovation. To test this demand channel, we construct a measure, *Customer Share*, to measure a supplier i's sales to its major customer j:

$$customer\ share = \frac{Sales_{ijt}}{Sales_{it}}, \tag{5}$$

where $Sales_{ijt}$ is supplier i's dollar value of sales to its customer firm j at year t, and $Sales_{it}$ is the total sales of supplier i at year t. *Customer Share* therefore measures the fraction of the supplier's sales to customer j. For a supplier–customer pair, a larger *Customer Share* represents a customer of greater importance.

To test the demand channel, we include *Customer Share* and its interaction with the main independent variable *LnDistance* in the baseline model. If the demand channel plays an important role in explaining the negative effect of distance on supplier innovation, it would subsume the prediction power of the distance variable, *LnDistance*.

The results are reported in Table 11. After controlling for the demand channel, the coefficient estimates on *LnDistance* remain negative in all columns but become statistically insignificant. The coefficient estimates on the interaction term, however, are negative and significant in all columns. These findings suggest that the positive effect of supplier–customer proximity on supplier innovation is much more pronounced when the customer demand accounts for a significant fraction of the supplier's total sales and, therefore, the customer is of great importance to the supplier. The demand channel, therefore, indeed has an important impact on supplier innovation.

4.3. The Agglomeration Channel

A few previous studies document the effect of geographic proximity on productivity and innovation (e.g., Adams and Jaffe 1996, Lychagin et al. 2016, Keller 2002). Suppliers and customers, when they locate close to each other, may share important factors in the production process, such as intermediate input, talent pool, and natural resources (e.g., Orlando 2004), which could enhance suppliers' innovation. Hence, it is important to examine whether our findings in this paper are driven by the agglomeration effect.

It is difficult to directly measure agglomeration because we cannot observe whether the supplier and customer share the same intermediate inputs, talent pool, natural resources, etc. However, for the agglomeration channel to play a role that explains our main results, the distance between the supplier and the customer cannot be too long. This is because, apparently, two firms locating far away, such as in different states, cannot benefit from shared inputs or skilled labor forces. As a result, agglomeration, if it is an important underlying channel, should play a much less important role in these observations. However, if proximity still plays an important role in determining supplier innovation output in these observations, agglomeration may not be an important channel driving our results.

Following this intuition, we exclude customer headquarters relocations in which the customer moves either away from or to the same state as the supplier. Dropping these observations from our sample allows us to repeat our main tests only for the pairs of

Table 11. The Demand Channel

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.073	−0.072	−0.038	−0.048	−0.099	−0.051	−0.127	−0.097	−0.082
	(0.148)	(0.182)	(0.125)	(0.378)	(0.311)	(0.171)	(0.373)	(0.387)	(0.464)
LnDistance × *Customer Share*	−0.153***	−0.147***	−0.118**	−0.330***	−0.235**	−0.320**	−0.242*	−0.349**	−0.265**
	(0.059)	(0.060)	(0.056)	(0.100)	(0.118)	(0.124)	(0.128)	(0.120)	(0.127)
Customer Share	0.536	0.515	0.452	0.554	0.579	0.698	0.792	0.773	0.722
	(0.554)	(0.572)	(0.341)	(0.620)	(0.743)	(0.768)	(0.805)	(0.742)	(0.788)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,254	5,988	5,689	3,237	2,926	2,652	2,971	2,681	2,406
R^2	0.845	0.831	0.831	0.798	0.793	0.794	0.879	0.878	0.888

Notes. This table reports the results when the main effect of *Customer Share* and the interaction between *LnDistance* and *Customer Share* are included simultaneously in the baseline regression. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Definitions of variables are listed in Table 2. Year fixed effects and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 12. The Agglomeration Channel

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.059**	−0.055**	−0.037*	−0.126***	−0.049*	−0.178***	−0.015**	−0.044**	−0.127***
	(0.029)	(0.026)	(0.019)	(0.043)	(0.025)	(0.031)	(0.005)	(0.021)	(0.042)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,984	5,728	5,442	3,108	2,810	2,544	2,853	2,574	2,308
R^2	0.844	0.830	0.830	0.800	0.791	0.791	0.880	0.879	0.889

Notes. This table reports the regression results of the model in Equation (1) by excluding customer relocations in which the customer is either moving to the same state as the supplier or moving away from the same state as the supplier. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Control variables are the same as in Table 4, year fixed effects and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

suppliers and customers that locate in different states and, therefore, are less subject to the agglomeration effect. We repeat our analysis by estimating Equation (1) in this restricted sample and report the results in Table 12. The coefficient estimates on *LnDistance* are negative and statistically significant in all columns for innovation quantity, quality, and efficiency. In an untabulated analysis, we repeat the analysis in a sample in which we exclude customer headquarters relocations in which the customer moves either away from or to the same city as the supplier. The results continue to hold. These findings suggest that the negative effect of geographic distance on supplier innovation is unlikely to be driven by the agglomeration channel.

4.4. The Social Connection Channel

A contemporaneous paper by Dasgupta et al. (2015) examines how supplier–customer relationships affect corporate innovation through the social connections between managers or board members in the customer and supplier firms. They find that a tight social connection helps mitigate hold-up problems and, thus, encourages supplier innovation.

The distance measure we use in this paper might be correlated with the social-connection measure used in Dasgupta et al. (2015), because a short distance between a supplier and its customer facilitates more interactions between the managers of the supplier and the customer, which, in turn, leads to tight social connections.

To test whether social connections between managers of suppliers and customers play an important role in explaining our main findings, we explicitly control for the social-connection measures used in Dasgupta et al. (2015) in our baseline regression.[6] Table 13 reports the results. Our baseline results remain similar after

Table 13. The Social Connection Channel

	LnPatents			*LnCites*			*LnIE*		
	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$	$t+1$	$t+2$	$t+3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LnDistance	−0.034**	−0.043*	−0.021	−0.069**	−0.067**	−0.058*	−0.075*	−0.091**	−0.060
	(0.016)	(0.022)	(0.035)	(0.033)	(0.034)	(0.033)	(0.042)	(0.041)	(0.042)
LnDistance × *LnConnection*	−0.011	−0.012	−0.025	−0.013	−0.007	−0.003	−0.004	−0.003	−0.029
	(0.012)	(0.018)	(0.024)	(0.015)	(0.022)	(0.042)	(0.015)	(0.020)	(0.028)
LnConnection	0.023	−0.032	0.063	0.003	−0.152	−0.013	0.090	0.002	0.287
	(0.096)	(0.105)	(0.143)	(0.121)	(0.168)	(0.340)	(0.109)	(0.147)	(0.238)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,064	2,617	2,198	1,138	788	487	1,104	765	470
R^2	0.882	0.824	0.819	0.925	0.956	0.958	0.913	0.937	0.963

Notes. This table reports the regression results after controlling for the social-connection measures used in Dasgupta et al. (2015). The regression model is $Innovation_{i\tau} = +\beta\, LnDistance_{ijt} + \gamma' X_{it} + Year_t + Pair_{ij} + \varepsilon_{ijt}$, where the social-connection measure *LnConnection* and its interaction with *LnDistance* are included in the controls. The dependent variables are *LnPatents* in columns (1)–(3), *LnCites* in columns (4)–(6), and *LnIE* in columns (7)–(9). Definitions of variables are listed in Table 2. Other relevant control variables, year fixed effects, and supplier–customer pair fixed effects are included in all regressions. Robust standard errors are reported in parentheses below the coefficient estimates.

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

controlling for the social-connection measures. Hence, the social-connection channel does not appear to drive the positive effect of supplier–customer proximity on supplier innovation documented in this paper.

5. Conclusion

In this paper, we examine the effect of supplier–customer geographic proximity on supplier innovation. To establish causality, we explore plausibly exogenous variation in proximity caused by customer headquarters relocations. In a difference-in-differences framework, we show that geographic proximity between a supplier and its major customer has a positive, causal effect on supplier innovation. The effect is stronger when the customer is more innovative itself, when the supplier and customer are closer in technology space, and when the customer's demand accounts for a larger fraction of the supplier's total sales. Our findings are consistent with the feedback effect proposed in Manso (2011) and the demand channel proposed in a few innovation models. Our paper sheds new light on the real effect of supplier–customer relationship on corporate innovation.

Acknowledgments

For helpful comments, the authors thank Matt Billett, Andrew Ellul, Edward Fee, Janet Gao, Paul Gao, Charles Hadlock, Dongmei Li, Francisco Perez-Gonzales, Sophie Shive, Sergey Tsyplakov, Jun Yang, and Donghang Zhang, as well as participants at the 2015 American Finance Association annual meeting, the 2015 China International Conference in Finance, the 2014 State of Indiana Conference, Indiana University, and the University of South Carolina. They also thank Sudipto Dasgupta, Kuo Zhang, and Chenqi Zhu for providing the social-connection measures used in Dasgupta et al. (2015).

Appendix

We convert latitude or longitude from decimal degrees to radians by dividing the latitude and longitude values by $180/\pi$, or approximately 57.296. Because the radius of the earth is assumed to be 6,378.8 kilometers, or 3,963 miles, we use the great circle distance formula to calculate mileage between two pairs of latitudes and longitudes:

$$3{,}963 \times \arccos[\sin(Lat_1) \times \sin(Lat_2) + \cos(Lat_1) \times \cos(Lat_2) \times \cos(Long_2 - Long_1)],$$

where Lat_1 and Lat_2 ($Long_1$ and $Long_2$) represent the latitudes (longitudes) of two points, respectively.

Endnotes

[1] These effects include, for example, financing cost (Cen et al. 2015a), capital structure decisions (Kale and Shahrur 2007, Banerjee et al. 2008, Chu 2012), relationship-specific investments (Kale et al. 2011), cross-ownership (Fee et al. 2006), mergers and acquisitions (Fee and Thomas 2004, Shahrur 2005, Ahern and Harford 2014), and financial distress (Hertzel et al. 2008).

[2] Since within-city relocations do not create meaningful changes in distance, we use them as a falsification test reported in panel A of Table 8. As expected, the within-city relocations that do not create significant changes in distance have no effect on supplier innovation.

[3] We describe the details of distance calculation in the appendix.

[4] Note that, to save space, we suppress the coefficient estimates of all control variables starting from Table 5. They are available on request.

[5] The uncontrolled or unobservable structural changes, if exist, will be in the error term. But since they are uncorrelated with the distance, the error term will be uncorrelated with the distance. We therefore still have consistent coefficient estimates in our baseline regressions.

[6] We thank Sudipto Dasgupta, Kuo Zhang, and Chenqi Zhu for providing us the social-connection measures used in Dasgupta et al. (2015).

References

Acharya VV, Subramanian KV (2009) Bankruptcy codes and innovation. *Rev. Financial Stud.* 22(12):4949–4988.

Acharya VV, Baghai RP, Subramanian KV (2013) Labor laws and innovation. *J. Law Econom.* 56(4):997–1037.

Acharya VV, Baghai RP, Subramanian KV (2014) Wrongful discharge laws and innovation. *Rev. Financial Stud.* 27(1):301–346.

Adams JD, Jaffe AB (1996) Bounding the effects of R&D: An investigation using matched establishment-firm data. *RAND J. Econom.* 27(4):700–721.

Ahern K, Harford J (2014) The importance of industry links in merger waves. *J. Finance* 69(2):527–576.

Alam ZS, Chen MA, Ciccotello CS, Ryan HE (2014) Does the location of directors matter? Information acquisition and board decisions. *J. Financial Quant. Anal.* 49(1):131–164.

Azoulay P, Graff Zivin JS, Manso G (2011) Incentives and creativity: Evidence from the academic life sciences. *RAND J. Econom.* 42(3):527–554.

Banerjee S, Dasgupta S, Kim Y (2008) Buyer–supplier relationships and the stakeholder theory of capital structure. *J. Finance* 63(5):2507–2552.

Baranchuk N, Kieschnick R, Moussawi R (2014) Motivating innovation in newly public firms. *J. Financial Econom.* 111(3):578–588.

Barrot JN, Sauvagnat J (2016) Input specificity and the propagation of idiosyncratic shocks in production networks. *Quart. J. Econom.* 131(3):1543–1592.

Bena J, Li K (2014) Corporate innovations and mergers and acquisitions. *J. Finance* 69(5):1923–1960.

Bloom N, Schankerman M, Van Reenen J (2013) Identifying technology spillovers and product market rivalry. *Econometrica* 81(4):1347–1393.

Bradley D, Kim I, Tian X (2017) Do unions affect innovation? *Management Sci.* 63(7):2251–2271.

Brav A, Jiang W, Ma S, Tian X (2018) How does hedge fund activism reshape corporate innovation? *J. Financial Econom.* 130(2): 237–264.

Catalini C, Fons-Rosen C, Gaulé P (2016) Did cheaper flights change the geography of scientific collaboration? MIT Sloan Research Paper 5172-16, MIT Sloan School of Management, Cambridge, MA.

Cen L, Dasgupta S, Sen R (2015b) Discipline or disruption? Stakeholder relationships and the effect of takeover threat. *Management Sci.* 62(10):2820–2841.

Cen L, Dasgupta S, Elkamhi R, Pungaliya RS (2015a) Reputation and loan contract terms: The role of principal customers. *Rev. Finance* 20(2):501–533.

Chang X, Fu K, Low A, Zhang W (2015) Non-executive employee stock options and corporate innovation. *J. Financial Econom.* 115(1):168–188.

Chemmanur TJ, Tian X (2018) Do anti-takeover provisions spur corporate innovation? *J. Financial Quant. Anal.* 53(3):1163–1194.

Chemmanur TJ, Loutskina E, Tian X (2014) Corporate venture capital, value creation, and innovation. *Rev. Financial Stud.* 27(8): 2434–2473.

Chesbrough HW (2006) *Open Innovation: The New Imperative for Creating and Profiting from Technology* (Harvard Business School Press, Boston).

Chu Y (2012) Optimal capital structure, bargaining, and the supplier market structure. *J. Financial Econom.* 106(2):411–426.

Cornaggia J, Mao Y, Tian X, Wolfe B (2015) Does banking competition affect innovation? *J. Financial Econom.* 115(1):189–209.

Dasgupta S, Zhang K, Zhu C (2015) Innovation, social connections, and the boundary of the firm. Working paper, Lancaster University, Lancaster, UK.

d'Aspremont C, Jacquemin A (1988) Cooperative and noncooperative R&D in duopoly with spillovers. *Amer. Econom. Rev.* 78(5): 1133–1137.

Eaton J, Kortum S (1996) Trade in ideas patenting and productivity in the OECD. *J. Internat. Econom.* 40(3):251–278.

Ederer F, Manso G (2013) Is pay for performance detrimental to innovation? *Management Sci.* 59(7):1496–1513.

Fang VW, Tian X, Tice S (2014) Does stock liquidity enhance or impede firm innovation? *J. Finance* 69(5):2085–2125.

Fee CE, Thomas S (2004) Sources of gains in horizontal mergers: Evidence from customer, supplier, and rival firms. *J. Financial Econom.* 74(3):423–460.

Fee CE, Hadlock CJ, Thomas S (2006) Corporate equity ownership and the governance of product market relationships. *J. Finance* 61(3):1217–1251.

Ferreira D, Manso G, Silva AC (2014) Incentives to innovate and the decision to go public or private. *Rev. Financial Stud.* 27(1): 256–300.

Giroud X (2013) Proximity and investment: Evidence from plant-level data. *Quart. J. Econom.* 128(2):861–915.

Hall BH, Jaffe AB, Trajtenberg M (2001) The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper 8498, National Bureau of Economic Research, Cambridge, MA.

He JJ, Tian X (2013) The dark side of analyst coverage: The case of innovation. *J. Financial Econom.* 109(3):856–878.

Hertzel MG, Li Z, Officer MS, Rodgers KJ (2008) Inter-firm linkages and the wealth effects of financial distress along the supply chain. *J. Financial Econom.* 87(2):374–387.

Hirshleifer D, Hsu PH, Li D (2013) Innovative efficiency and stock returns. *J. Financial Econom.* 107(3):632–654.

Hsu PH, Tian X, Xu Y (2014) Financial development and innovation: Cross-country evidence. *J. Financial Econom.* 112(1):116–135.

Jaffe A (1986) Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value. *Amer. Econom. Rev.* 76(5):984–1001.

Kale J, Shahrur H (2007) Corporate capital structure and the characteristics of suppliers and customers. *J. Financial Econom.* 83(2): 321–365.

Kale J, Kedia S, Williams R (2011) The effect of CEO risk-taking incentives on relationship-specific investments by customers and suppliers. Working paper, Northeastern University, Boston.

Kamien MI, Muller E, Zang I (1992) Research joint ventures and R&D cartels. *Amer. Econom. Rev.* 82(5):1293–1306.

Keller W (2002) Geographic localization of international technology diffusion. *Amer. Econom. Rev.* 92(1):120–142.

Knyazeva A, Knyazeva D, Masulis RW (2013) The supply of corporate directors and board independence. *Rev. Financial Stud.* 26(6):1561–1605.

Leahy D, Neary JP (1997) Public policy towards R&D in oligopolistic industries. *Amer. Econom. Rev.* 87(4):642–662.

Lychagin S, Pinkse J, Slade ME, Reenen JV (2016) Spillovers in space: Does geography matter? *J. Indust. Econom.* 64(2):295–335.

Malloy CJ (2005) The geography of equity analysis. *J. Finance* 60(2): 719–755.

Manso G (2011) Motivating innovation. *J. Finance* 66(5):1823–1869.

Orlando MJ (2004) Measuring spillovers from industrial R&D: On the importance of geographic and technological proximity. *RAND J. Econom.* 35(4):777–786.

Petersen MA, Rajan RG (2002) Does distance still matter? The information revolution in small business lending. *J. Finance* 57(6): 2533–2570.

Pirinsky C, Wang Q (2006) Does corporate headquarters location matter for stock returns? *J. Finance* 61(4):1991–2015.

Porter M (1992) Capital disadvantage: America's failing capital investment system. *Harvard Bus. Rev.* 19(4):65–82.

Prahalad CK, Ramaswamy V (2013) *The Future of Competition: Co-Creating Unique Value with Customers* (Harvard Business School Press, Boston).

Roberts MR, Whited TM (2013) Endogeneity in empirical corporate finance. *Handbook Econom. Finance* 2(Part A):493–572.

Seru A (2014) Firm boundaries matter: Evidence from conglomerates and R&D activity. *J. Financial Econom.* 111(2):381–405.

Shahrur H (2005) Industry structure and horizontal takeovers: Analysis of wealth effects on rivals, suppliers, and corporate customers. *J. Financial Econom.* 76(1):61–98.

Tian X (2011) The causes and consequences of venture capital stage financing. *J. Financial Econom.* 101(1):132–159.

Tian X, Wang TY (2014) Tolerance for failure and corporate innovation. *Rev. Financial Stud.* 27(1):211–255.

Von Hippel E (2005) Democratizing innovation: The evolving phenomenon of user innovation. *J. Für Betriebswirtschaft* 55(1):63–78.