Agglomeration, Coordination, and Corporate Investment*

William Grieser[†] Gonzalo Maturana[‡] Santiago Truffa[§]

October 2019

Firm agglomeration is positively correlated with productivity, and it exhibits significant heterogeneity across industries. Yet, the connection between agglomeration and corporate investment remains underexplored. We develop a model of information sharing which predicts that knowledge-intensive industries and industries with more uncertainty benefit the most from agglomeration due to the subsequent improvement in project selection. Using counterfactuals that account for nonrandom location decisions and industry concentration, we find a strong positive relationship between (a) industry uncertainty/knowledge intensity and (b) the proximity of headquarters, patent inventors, and customer–suppliers. In addition, we exploit techniques designed to address inherent difficulties in estimating peer effects, and we find that investment externalities are positively related to proximity and uncertainty/knowledge intensity. This is consistent with the predictions of our model.

*We are grateful to Andres Almazan, Jonathan Cohn, Ernesto Dal Bo, David Dicks, William Fuchs, Charles Hadlock, Zoran Ivkovich, Zack Liu, John Morgan, Andrei Simonov, Coen Teulings, Adam Winegar, and Morad Zekhnini. We are also grateful for the helpful comments of the seminar participants at Michigan State University, Pontificia Universidad Catolica de Chile, Universidad de los Andes Chile, University of California-Berkeley, Tulane University, the 2016 North American Regional Science Association Meetings, and the 2019 Urban Economics Association Meetings. This paper was previously circulated under the title "Clustering to Coordinate: Who Benefits from Knowledge Spillovers?"

[†]Neeley School of Business, Texas Christian University. Email: w.grieser@tcu.edu

[‡]Goizueta Business School, Emory University. Email: gonzalo.maturana@emory.edu

[§]ESE Business School, Universidad de los Andes Chile. Email: struffa.ese@uandes.cl

Agglomeration, Coordination, and Corporate Investment

October 2019

Abstract – Firm agglomeration is positively correlated with productivity, and it exhibits significant heterogeneity across industries. Yet, the connection between agglomeration and corporate investment remains underexplored. We develop a model of information sharing which predicts that knowledge-intensive industries and industries with more uncertainty benefit the most from agglomeration due to the subsequent improvement in project selection. Using counterfactuals that account for nonrandom location decisions and industry concentration, we find a strong positive relationship between (a) industry uncertainty/knowledge intensity and (b) the proximity of headquarters, patent inventors, and customer–suppliers. In addition, we exploit techniques designed to address inherent difficulties in estimating peer effects, and we find that investment externalities are positively related to proximity and uncertainty/knowledge intensity. This is consistent with the predictions of our model.

JEL classification: R1, R3, L1, O3

keywords: Agglomeration, industrial clustering, knowledge spillovers, innovation, R&D, investment, customer-supplier

How do firms benefit from agglomeration, and why do some industries cluster more than others? Urban economists have provided compelling evidence that industrial activity is spatially concentrated and that such *agglomeration* generates gains in firm and worker productivity. Since the work of Marshall (1920), part of this effect is often attributed to the notion that geographic concentration facilitates the spread of knowledge. However, few theories provide microfoundations as to when this mechanism should be more prevalent empirically. Furthermore, while productivity gains materialize through the actions of firms, previous research has only narrowly explored an explicit relationship between corporate investment decisions and the drivers of agglomeration. In this paper, we develop a specific mechanism through which information sharing in the presence of investment externalities improves project selection. We also provide empirical evidence consistent with this mechanism being an important driver of agglomeration.

Corporate investment involves selecting among projects that have uncertain future cash flows. This uncertainty reduces the likelihood that profitable projects are undertaken. If project payouts depend on the actions of related firms, then ambiguity regarding the actions of those firms will exacerbate the effects of uncertainty. One possible way for firms to mitigate this problem is to co-locate with related firms to facilitate communication and the sharing of private information regarding project payouts. These improvements can reduce uncertainty and thus improve project selection. While this is an important mechanism to explore, many competing channels can also explain firms' decisions to co-locate. Therefore, we start from a theoretical framework that isolates the features that are unique to an information sharing channel.

In our setting, firms make investment decisions in an environment that features incomplete information and investment externalities (e.g., technological spillovers, knowledge spillovers). Before investing, firms observe both a private and a public signal that informs their beliefs about project valuations as well as their beliefs about other firms' beliefs about those valuations. As firms recursively incorporate other firms' beliefs into their decision-making process, they overweight the importance of the public signal relative to their own private signal (we label this *strategic miscoordination*).[1] While the combination of incomplete information and investment externalities makes it difficult to solve for equilibrium outcomes, the global games framework developed by Carlsson and

1. Morris and Shin (2003) define higher-order beliefs as "... players' beliefs about other players' beliefs, players' beliefs about other players' beliefs about other players' beliefs, and so on."

Van Damme (1993) and Morris and Shin (2003) offers a useful approach for circumventing these difficulties.

In the context of our model, the physical proximity that dense urban centers provide can be thought of as a technology that allows firms to share private knowledge, thus reducing the informational inefficiencies that distort optimal investment. Specifically, the ability to share private knowledge mitigates the miscoordination problem, which translates into firms choosing projects more precisely. Moreover, we derive that the coordination gains from knowledge sharing are increasing in an industry's level of uncertainty and knowledge intensity.

Testing the predictions of our model is empirically challenging, since even in the absence of agglomeration externalities, firms are unlikely to locate randomly. Specifically, the spatial concentration of an industry should depend on its size and the general concentration of the population at large, which may be determined by a variety of factors that are difficult to observe (e.g., local amenities). To mitigate this concern, we conduct tests of agglomeration (i.e., localization) using counterfactuals that account for nonrandom location patterns and industry concentration following the work of Duranton and Overman (2005).[2]

We develop a measure of industry uncertainty and knowledge intensity by combining two industry-level metrics: (1) stock return volatility and (2) a measure of worker skill (i.e., knowledge intensity).[3] We collect geographic coordinates for corporate headquarters (HQ) ZIP codes for more than 9,000 firms in 24 industries, and we find a strong positive relationship between industry clustering and industry uncertainty/knowledge intensity at close distances (0–20 miles). Moreover, we do not find this relationship at longer distances, which is consistent with the predictions of our model.

We further explore the abnormal regional clustering of capital expenditures (CapEx) and R&D across industries. We find that CapEx and R&D exhibit even stronger clustering than HQ locations in more uncertain and knowledge-intensive industries. This finding is reassuring, especially since R&D has been shown to be particularly reliant on knowledge spillovers (e.g., Audretsch and Feldman 2004). Finally, while HQ location provides a first approximation for the locality of a firm's

2. To control for industry size and population concentration, Duranton and Overman (2005) construct counterfactuals by generating "pseudo-industries" through randomly sampling firms from the full set of possible locations according to industrial organization and general population conditions.

3. The intuition behind this choice is that more knowledge-intensive industries require better-trained and better-educated workers. We use a ranking of occupations from the U.S. Department of Labor.

activity, a firm may also conduct some of its operations elsewhere. To mitigate this concern, we conduct localization tests, as described above, using patent inventor locations (which may differ significantly from headquarter locations), and we find similar results.

In addition to describing firm location decisions, our model predicts that agglomeration generates positive externalities in corporate investment behavior that are increasing in the uncertainty and knowledge-intensity of an industry. As such, we explore the effects of locating within a regional industry cluster on the level and timing of corporate investment. We start by building on the framework of Dougal, Parsons, and Titman (2015), and explore the extent to which the level of investment depends on regional industry peers. We find that investment externalities are significantly stronger among firms in more uncertain and more knowledge-intensive industries. Specifically, the investment comovement among competitors within an industry cluster is more than twice as strong as the comovement among competitors located outside the cluster, on average. Moreover, this relation is driven entirely by firms in more uncertain and more knowledge-intensive industries.

While these results are informative, the estimation of firm interactions poses empirical challenges. For example, it is difficult to separate the effects of peer interactions from the effects of selection and exposure to common shocks. To mitigate these concerns, we implement spatial econometric techniques that are designed specifically to circumvent these issues (Jackson 2010; Grieser et al. 2019). Using a network of regional proximities between firms, we find corroborating evidence that investment externalities are stronger in firm clusters, and that this effect is increasing in the uncertainty/knowledge-intensity of industries. Finally, we examine the clustering of the unexplained component of corporate investment. We show that residuals on CapEx and R&D are significantly more similar for firms in more uncertain and more knowledge-intensive industries–even beyond what can be explained by time fixed effects and selection into industries and regions based on observable criteria. This result is consistent with comovement being induced by firm interaction and not merely by firm selection.

The investment externalities and the coordination predictions of our model distinguishes our channel from several competing channels of agglomeration. Thus, we also study location decisions in a setting in which these features of our model should be particularly prevalent: customer–supplier relationships. We analyze a sample of more than 2,300 customer–supplier pairs in which the customer accounts for at least 10% of the supplier's total sales. We find that suppliers in industries with the highest uncertainty/knowledge intensity are 7.9 percentage points (pp) more likely to locate within 20 miles of customers than suppliers in industries with the lowest uncertainty/knowledge intensity. This 7.9 pp increase is economically large, representing 89.3% of the sample average propensity for suppliers to locate near customers. Moreover, when suppliers sell to customers in industries with the highest uncertainty/knowledge intensity, their headquarters tend to be 172 miles closer, on average, to their customer compared to suppliers that sell to customers in industries with the lowest uncertainty/knowledge intensity.

Our results strongly support the assertion that knowledge spillovers are an important component of firm location and investment decisions, but we acknowledge that multiple mechanisms could be working simultaneously.[4] Nonetheless, the relationship between industry uncertainty/knowledge intensity and investment clustering and comovement is important because most competing channels of agglomeration do not share this feature. We elaborate on this distinction in Section 6.

Our paper contributes to the literature on corporate investment by relating investment behavior to drivers of agglomeration. Corporate investment involves valuing projects that have uncertain cash flows, which can be difficult when information is imprecise (e.g., Dixit and Pindyck 1994). Our model suggests that one plausible channel for firms to reduce uncertainty (and thereby improve project selection) is to co-locate with related firms in order to facilitate the sharing of private information. We find strong evidence that agglomeration drives corporate investment behavior in more uncertain and knowledge-intensive industries. These findings complement three recent papers. First, Dougal, Parsons, and Titman (2015) show that corporate investment depends on regional externalities outside of industry relationships. Next, Almazan et al. (2010) show that agglomerated firms have more acquisition opportunities. Last, Engelberg et al. (2018) show that geography facilitates knowledge spillovers between information intermediaries[5]

Our paper also relates to the literature that provides microfoundations for knowledge diffusion.[6] We use insights from the global games literature to show that the sharing of private information facilitates coordination among firms in the presence of investment externalities. Our findings are consistent with those of Combes and Duranton (2006), Helsley and Strange (2004), and Couture

4. For example, highly skilled workers may benefit more from living in larger cities.

5. More generally, our paper fits into the recent emerging literature on the role of interdependence in corporate decision making (e.g., citealtshue2013,learyroberts2014jf,fracassi15).

6. See Duranton and Puga (2004) for a survey of the theory.

(2015), who show that localization reduces the costs of exchanging ideas, and Glaeser (1999) and Storper and Venables (2004), who motivate human capital externalities through learning. This literature also provides suggestive empirical evidence that knowledge spillovers play an important role in the clustering of certain industries (e.g., Audretsch and Stephan 1996; Arzaghi and Henderson 2008; Buzard et al. 2015; Lychagin 2016). We provide theoretical motivation for industry characteristics that help explain the strong heterogeneity in clustering across industries, and we provide empirical support for our model.

More broadly, our paper relates to the literature on the determinants of agglomeration externalities. Marshall (1920) classifies three main forces of agglomeration: labor pooling, input sharing, and knowledge spillovers. A number of studies has established correlations between measures of agglomeration and industry characteristics in an attempt to uncover the underlying drivers of agglomeration.[7] However, the literature is just beginning to explore the relative influence of the three agglomeration forces (see Rosenthal and Strange 2004; Ellison, Glaeser, and Kerr 2010), and this research has not been integrated into the corporate finance literature, with the exception of Almazan et al. (2007), who examine the location decisions of firms in the presence of the benefits of an active labor market. Our model and our empirical results provide insights into the specific mechanisms through which knowledge is transferred between firms.

Finally, we provide a rationale for the value of coordination in the presence of knowledge spillovers, which relates to the literature that examines why workers are more productive in cities (e.g., Glaeser and Maré 2001; Moretti 2004; De la Roca, Ottaviano, and Puga 2014). More broadly, our work relates to the classical literature that links knowledge externalities to economic development (e.g., Lucas 1988; Romer 1986) and to more recent work on the relation between agglomeration and growth (e.g., Rossi-Hansberg and Wright 2007; Davis, Fisher, and Whited 2014).

The rest of the paper is organized as follows: Section 1 provides the background and motivation for the model and describes the model along with its main predictions. Section 2 describes the data and measures used in our empirical tests. Sections 3 and 4 present the empirical results. Finally, Section 6 provides a discussion of how our findings relate to alternative theories of agglomeration and Section 7 concludes.

7. For example, Carlino and Kerr (2014a) show that innovative activity is more concentrated than industrial activity. Likewise, Ellison, Glaeser, and Kerr (2010) and Faggio, Silva, and Strange (2014) find substantial heterogeneity in patterns of industry agglomeration using establishment data from the U.S. and the U.K..

1. Theoretical Motivation

A variety of microeconomic mechanisms yield observationally equivalent results for aggregate agglomeration patterns. We provide a theoretical framework that yields predictions unique to knowledge diffusion stemming from firm coordination. We build on the notion that, although information flows more freely than ever, tacit knowledge is embedded within workers and is inherently difficult to transfer (Storper and Venables 2004). As a result, distance continues to play a pivotal role in the transfer of knowledge.

1.1. Basic Framework

Consider a continuum of firms that face a binary investment decision: to invest (I) or not invest (NI). Consistent with prior literature (e.g., Murphy et al. 1989), we model the presence of knowledge spillovers by assuming that investments generate positive externalities and thus are complementary among related firms. That is, the profitability of an investment increases with the investment of other firms. It is natural to wonder how these externalities affect corporate investment decisions. We propose that in the presence of investment externalities, coordination among firms can be an important factor in determining their investment behavior and determining the value of co-locating. Moreover, we propose that gains from coordination (and therefore the value of co-locating) are heterogeneous and depend on fundamental industry characteristics.

Our assumption of positive externalities aims to capture the notion that project payouts are often interdependent and that firms can learn from each other.[8] Thus, we model the payoff of the investment decision as given by

$$U = \begin{cases} \theta + l - 1 & if \quad I \\ 0 & if \quad NI \end{cases}, \tag{1}$$

where θ is a random variable that represents the profitability of an uncertain investment and l is the proportion of other firms that also decide to invest.[9] Without a loss of generality, we normalize

8. For instance, AMD produces processors, which are much more valuable if Dell also produces computers that use processors, and vice versa. Moreover, this value can be even larger in the presence of technological spillovers.

9. We assume that θ is randomly drawn from the real line, with each realization equally likely to arise from a uniform distribution.

the profitability of not investing to zero.

Firms anticipate the profitability of investment decisions using two sources of information. First, firms observe a noisy public signal that is available to all firms (e.g., aggregate economic performance). Although this public signal speaks to the true future state of the world, it is potentially very noisy and may not provide enough information for firms to make precise assessments about future profitability. Second, each firm observes a private signal obtained from their own research and proprietary information. We assume that firms' assessments are unbiased. However, because firms are imperfect, they can generate only noisy private signals.

To formalize this idea, we denote the noisy and unbiased public signal as $y_i = \theta + \varepsilon_i$, where $\varepsilon_i \backsim N(0, \tau^2)$ and τ captures the noise of the public signal. An industry with a noisier public signal will exhibit higher idiosyncratic uncertainty regarding investment profits. Thus, we can conceptualize this noise as industry-level uncertainty. Additionally, each firm observes an independent private signal $x_i = \theta + \nu_i$, where $\nu_i \backsim N(0, \sigma^2)$ and σ relates to the precision of private assessments.[10] An industry with a noisier private signal will be inherently more complex. Industries that produce complex goods and services often require more highly skilled human capital. Thus, we can conceptualize this noise as industry-level knowledge intensity.

The symmetric equilibrium of a global game is fully characterized by a switching strategy in which firms invest whenever the expected profitability θ is higher than some threshold κ, and choose to not invest otherwise. Since payoffs depend not only on a firm's own actions, but also on the actions of related firms, an optimal investment strategy should include an estimate of the proportion of other firms that will also invest. Because the realization of signals is independent, the expected proportion of players who observe a signal lower than κ will be the same as the players' estimated probabilities that their respective opponents will also observe a signal lower than κ.[11]

Given the observed signals y and x, and given the properties of the normal distribution, each firm's expectation of θ is

$$\bar{\theta} = \frac{\sigma^2 y + \tau^2 x}{\sigma^2 + \tau^2}, \tag{2}$$

10. By construction, all noise terms are independently distributed and therefore uncorrelated.

11. Since payoffs are linear and all signals are independent, the strategies that arise from a 2-player game are equivalent to the strategies that arise from studying our more general continuous version of the game.

which is the average of both signals, weighted by their signal-to-noise ratio. Similarly, the standard deviation of θ is

$$\hat{\sigma} = \sqrt{\frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}}. \tag{3}$$

As mentioned above, each firm will follow a switching strategy $s()$, which is a function of its posterior:

$$s(\overline{\theta}) = \begin{cases} I & if \quad \bar{\theta} > \kappa \\ NI & if \quad \overline{\theta} \leq \kappa \end{cases}. \tag{4}$$

In the symmetric equilibrium of this game, a firm's strategy depends on its beliefs about the other firms' strategies. Therefore, each firm must anticipate each of the other firms' private signals. The equilibrium of the game is given by the equation:

$$\kappa - \Phi\left\{\gamma(\kappa - y)\right\} = 0, \tag{5}$$

where Φ is the normal distribution operator.[12]

Firms know that public signals are commonly observed, so a public signal will affect a firm's individual information about the future as well as its beliefs about other firms' information about the future. Since this is true for all firms, rational firms will recursively incorporate this reasoning into their strategic decision making, thus overweighting the importance of the public signal due to strategic considerations. As we show below, this can generate situations in which firms make inefficient investment decisions due to this strategic miscoordination. If firms use information efficiently, then their actions and beliefs should adjust to changes in public signals on a one-for-one basis. Any differences between actions and beliefs arise from the strategic effects induced by higher-order beliefs about the public signal.

12. This is a unique equilibrium when $\gamma \equiv \frac{(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} < 2\pi$ (see Morris and Shin 2003 for the proof).

1.2. The Value of Proximity

In our model, proximity allows firms to share private signals. This sharing provides firms with a sufficient statistic to guide their investment decisions. As this unbiased communication takes place, the sufficient statistic will follow a sampling distribution of private signals. As firms share their private assessments with n other discrete unbiased firms, the sample of private signals $\bar{x}$ will be given by

$$\bar{x} \sim N(\theta, \sigma/\sqrt{n}).^{13} \tag{6}$$

Thus, high-density regions facilitate the sharing of private signals. In particular, the signal will become more precise as the number of firms in a cluster increases. In the limit,[14]

$$\hat{\sigma} = \sigma/\sqrt{n} \to_{n\to\infty} 0. \tag{7}$$

Proposition 1. The sharing of private information (i.e., knowledge spillovers) generates gains from coordination that are greater for more uncertain and more knowledge-intensive industries.

Proof. See Appendix A. ■

Gains from agglomeration are larger for firms in industries with noisier public and private signals. Proposition 1 has implications for empirical analysis. As firms in more uncertain and more knowledge-intensive industries benefit to a greater extent from sharing information, we should observe a higher degree of spatial concentration for such industries. The following corollary formalizes this result.

Proposition 2. For any fixed cost of co-location C, there exists $(\bar{\tau}, \bar{\sigma})$ such that $\forall \tau > \bar{\tau}$ and $\forall \sigma > \bar{\sigma}$ only the firms in industries with an uncertainty/knowledge intensity level above this cutoff would co-locate.

Proof. See Appendix A. ■

13. Note that since we have a continuum of players, each discrete and independent signal that gets added to the sample, does not change the overall beliefs about aggregate distributions.

14. Note that as $\sigma \to 0$, $\gamma \to 0$. This limit is consistent with $\gamma < 2\pi$.

Proposition 2 states that in the presence of a fixed cost (e.g., real estate rental rates), we should observe a higher degree of firm clustering in more uncertain and knowledge-intensive industries, since the benefits of proximity are higher for these types of firms. This implication is directly testable in the cross-section.[15]

The assumption of a fixed cost of co-location can be justified by cities without space constraints. On the contrary, if costs are endogenous to the co-location decisions of firms, and if we also assume that (a) agents hold accurate beliefs about other agents' location decisions and (b) each firm acts as a price taker in the housing market, then the location game and the investment game can be divided into two independent games. In this case, the qualitative results that stem from the investment game are unaffected by the location game.

The gains from agglomeration stem from an investment channel, which in turn leads to implications for the investment behavior of firms that have already decided to co-locate.

PROPOSITION 3. Investment intensity is greater for agglomerated firms than for non-agglomerated firms, and this effect is increasing in more uncertain and knowledge-intensive industries.

Proof. See Appendix A. ■

Proposition 3 has important implications for empirical analysis. In particular, for more uncertain/knowledge-intensive industries, there is a larger difference between the investment levels of firms inside an industry cluster and firms outside the cluster. As firms in more uncertain industries can coordinate and make better assessments regarding the future profitability of their investment opportunities, these firms should exhibit a greater propensity to invest, which has implications for total investment.[16]

2. Data and Measures

2.1. Firms and Locations

To test the predictions of our model, we obtain information on firm headquarters (HQ) locations, industry characteristics, and financials from Compustat. Extant research has indicated the

15. Note that although sharing information is valuable even in the absence of knowledge spillovers, the latter is necessary for industry uncertainty/knowledge intensity to matter. We show this in the Internet Appendix.

16. Note that in the model investment is a binary decision for each firm. We show that, conditional on co-locating, firms in more uncertain industries are more likely to investment. Thus, aggregating at the industry level, we should observe more greater total investment in more uncertain industries.

significance of HQ location, which provides a useful first approximation for the location of a firm's activities (e.g., Dougal, Parsons, and Titman 2015; Pirinsky and Wang 2006). Indeed, Landier, Nair, and Wulf (2007) show that even geographically disperse firms choose to keep resources close to their corporate HQs. We collect geographic coordinates based on the ZIP code for each firm's HQ from 2000 to 2012.[17] We restrict the sample of firms to those firms headquartered in the contiguous United States. The sample includes a total of 9,167 unique firms. Descriptive statistics for these firms are presented in the Internet Appendix.

While research suggests that HQ location proxies for the locality of a firm's activity, a firm may also conduct some of its operations outside its HQ location. For instance, Honeywell is headquartered in New Jersey, but a significant portion of its patents are produced in Boston and in the San Francisco economic areas. We obtain information on inventors from the Harvard Patent Network Dataverse, which contains the locations of inventors associated with over 151,000 U.S. patent applications from 2006 to 2009.[18] We are concerned with the patents that can be assigned to a firm in the Compustat universe at the time of the patent application.

2.2. Industry Uncertainty

We conduct our analysis at the Fama and French 48 industry classification level. We exclude the finance and utilities industries, as well as any industry for which there are fewer than 100 firms in our sample.[19] Thus, 24 industries remain, which comprise 91.1% of the firms in our initial sample. In our model, uncertainty pertains to aggregate economic performance. Thus, we use stock price volatility as a proxy for industry-level uncertainty. To construct this measure, we use data from the Center for Research in Security Prices' (CRSP) Monthly Stock File. For each industry classification, we construct a series of value-weighted monthly returns from 2000 to 2012. Then, we compute industry volatility as the standard deviation of each series of returns.

17. We start in 2000 and not earlier because the productive structure of the U.S. economy has recently undergone important shifts (Herrendorf et al. 2014). This "Structural Transformation" can be understood as the reallocation of economic activity from agriculture to manufacturing and, recently, to knowledge services. Thus, in recent years, the forces in our model should be more salient.

18. We restrict the sample to a 4-year period for computational reasons. Also, ending the sample in 2009 provides enough time for a patent application to be granted, which mitigates truncation problems. Our sample consists of 5,670 patents from 12,769 inventor locations, operating in 17 different industries.

19. This cutoff at 100 firms per industry greatly increases the accuracy of constructing the Duranton and Overman counterfactual.

2.3. Industry Knowledge Intensity

We use a measure of worker skill as a proxy for industry-level knowledge intensity. The intuition behind this choice is that industries that focus on more knowledge-intensive products and markets require workers that are more highly trained and educated than the workers of firms that serve less knowledge-intensive markets. We use data from the Occupational Information Network (O*NET), a website that contains detailed information provided by the U.S. Department of Labor in a survey of randomly sampled U.S. workers for each occupation. O*NET classifies each occupation into one of five skill categories according to the degree of preparation needed. The skill level of occupations range from *little or no preparation needed* (Job Zone 1) to *extensive preparation* (Job Zone 5).[20]

To aggregate the O*NET skill measures to the industry level, we create a wage-weighted average skill for each 4-digit NAICS code, using the job zone assigned to each occupation according to the O*NET database. Wage estimates come from the Bureau of Labor and Statistics (BLS) Occupational Employment Statistics (OES) database. We calculate the total industry cost of input (wage) for each occupation by multiplying its annual mean wage by the number of people employed in an industry at that occupation according to the OES. Finally, we aggregate the average skill level across all 4-digit NAICS contained in each Fama–French 48 industry classification.

2.4. Industry Uncertainty/Knowledge Intensity Index

A fundamental feature of our model is that firms producing *knowledge-intensive* goods or services in *uncertain* environments benefit more from agglomeration externalities. Therefore, we construct an index to capture both uncertainty and knowledge intensity based on the two metrics described above. More specifically, the uncertainty and knowledge intensity metrics are standardized so that both measures, which initially differ in levels, become comparable. Then, the resulting values are averaged and normalized so that the index ranges from 0 to 1. While this transformation alters the the distribution of the underlying measures, it facilitates interpretation of the results. In the Internet Appendix, we show qualitatively similar results when using the raw measures of

20. Job Zone 1 includes occupations that may require a high school diploma or GED, little or no previous work-related skill required, and a few days to a few months of on-the-job training. Job Zone 5 includes occupations that typically require a master's degree, Ph.D., M.D., or J.D.; in other words, extensive skill, knowledge, and experience. Examples of occupations in Job Zone 1 include taxi drivers, amusement and recreation attendants, and non-farm animal caretakers, while examples from Job Zone 5 include lawyers, sports medicine physicians, surgeons, treasurers, and controllers.

uncertainty and knowledge intensity separately.

Table 1 lists the 24 industries, along with their annualized volatility, their required worker skill level, and their uncertainty/knowledge-intensity (UKI) index (i.e., the combination of industry volatility and skill). The industries are ordered according to the UKI index. Consistent with general intuition, the industries with the highest uncertainty/knowledge intensity are "Electronic equipment," "Measuring and control equipment," and "Computers," while the industries with the lowest uncertainty/knowledge intensity are "Meals, restaurants, and hotels," "Food," and "Retail."

[INSERT TABLE 1 HERE]

3. Measuring Localization

3.1. Kernel Density Estimations

In this section, we construct a measure of agglomeration at the industry level. We use the methodology in Duranton and Overman (2005) (hereafter DO), who develop a test of localization based on kernel density estimations of bilateral distances between firms in an industry. More specifically, they estimate the following function for each industry A:

$$\hat{K}_A(d) = \frac{1}{n_A(n_A-1)h} \sum_{i=1}^{n_A-1} \sum_{j=1}^{n_A} f\left(\frac{d-d_{i,j}}{h}\right), \tag{8}$$

where $d_{i,j}$ is the Euclidean distance between the locations of establishments i and j in industry A. The number of establishments in an industry is denoted by n_A. The function f is a Gaussian kernel density with bandwidth h. Note that Equation (8) generates a density distribution for all potential bilateral distances. Industries with a high degree of agglomeration will have high values of $\hat{K}_A(d)$ at shorter distances that dissipate at longer distances.

Panel A of Figure 1 plots the kernel density estimations (Equation (8)) for industries with the highest and lowest UKI index (i.e., "Electronic equipment" and "Meals, restaurants, and hotels," respectively). Consistent with the predictions of our model, the probability of two firms being located within 20 miles of each other is about four times larger in the most uncertain/knowledge-intensive industry than in the least uncertain/knowledge-intensive industry. In fact, most of the differences between the two densities are driven by distances of less than 40 miles. For longer distances, the densities are quite similar across industries.

[INSERT FIGURE 1 HERE]

Although the kernel density provides useful information about the distribution of the different localities in an industry, it does not provide the full picture. Even if a value of $\hat{K}_A(d)$ at a given distance in a given industry appears to be high, it cannot be concluded that the value is *abnormally* high without comparing it to the appropriate counterfactual. In particular, the comparison with other industries may not be informative, as spatial concentration depends on both the size and the concentration of the industries as well as the general population density. To address this issue, DO construct counterfactuals by generating 1,000 pseudo-industries of equivalent size as the industry of interest by randomly sampling from the full set of possible locations. From these simulations, DO construct confidence intervals for each industry and distance. In particular, let $\bar{K}_A(d)$ be the upper limit for the 95% confidence interval. DO define the following index of localization:

$$\gamma_A(d) \equiv max\left(\hat{K}_A(d) - \bar{K}_A(d), 0\right). \tag{9}$$

A positive value of $\gamma_A(d)$ (i.e., when the kernel density exceeds the upper bound of the 95% confidence interval) indicates a departure from randomness, subject to stylized industry concentration and overall population characteristics. Therefore, a positive value of $\gamma_A(d)$ suggests that industry A exhibits excess localization at distance d, with higher values of $\gamma_A(d)$ suggesting a greater degree of excess localization.

Panel B of Figure 1 contrasts the two density estimates in Panel A of Figure 1 against their respective 95% confidence intervals. The graph on the left-hand side of Panel B shows that the most uncertain/knowledge-intensive industry exhibits significant excess localization ($\gamma_A(d) > 0$) for all distances less than 30 miles. In contrast, the graph on the right-hand side of Panel B shows that this is not the case for the least uncertain/knowledge-intensive industry. The kernel density estimate lies within the confidence interval for most distance values.

3.2. Excess Localization and Uncertainty/Knowledge-Intensity

In this section, we aim to provide more generalizable evidence for the implications of our theoretical model. Recall that Proposition 1 suggests that in environments where knowledge spillovers are more prevalent, information sharing generates gains from coordination that are increasing on the uncertainty/knowledge-intensity of an industry. Therefore, in our empirical setting, if knowledge spillovers are an important determinant of a firm's location decision, then the industry localization index should be positively correlated with the UKI index at close distances (i.e., there should be a higher degree of clustering in higher uncertainty/knowledge-intensive industries).

Panel A of Figure 2 plots the relation between the localization index and the UKI index of the 24 industries in our sample, along with the corresponding quadratic interpolation. The relation is strongly positive for short distances between 0 and 20 miles. This relationship weakens substantially once we increase the distance interval to between 20 and 40 miles (Panel B), and it disappears completely at longer distances greater than 40 miles (Panels C and D). In Figures IA.1 and IA.2 of the Internet Appendix, we repeat this exercise using industry uncertainty and knowledge intensity separately, and we find a similar pattern. Overall, this evidence is consistent with the benefits of coordination in the presence of knowledge spillovers driving the localization decisions of firms.

[INSERT FIGURE 2 HERE]

To facilitate comparison, Panel A of Figure 3 consolidates the quadratic interpolations from the four panels of Figure 2 into a single panel. In Panels B and C, we repeat our analysis using CapEx- and R&D-weighted HQ locations, respectively. Specifically, each pair of HQ locations is weighted by the firm's aggregated CapEx (R&D) expenses when computing the kernel densities. Consequently, the resulting kernel densities indicate the probability of an additional dollar of CapEx (R&D) agglomerating within a certain distance for a given industry. Similar to HQ locations, CapEx and R&D exhibit a high degree of excess localization in more uncertain and more knowledge-intensive industries at close distances, and this relation dissipates at longer distances.

[INSERT FIGURE 3 HERE]

In Panel D of Figure 3, we repeat the localization test using patent inventor locations instead of HQ locations to alleviate concerns that firms do not conduct a significant proportion of their operations at their HQ location.[21] As with the HQ locations, there is a positive relation between the localization index and the UKI index. This relationship is most pronounced for short distances between 0 and 20 miles and disappears for longer distances. Overall, the robustness of our results across different settings strengthens our confidence in the predictions of our model.

In addition, it is reassuring for our proposed channel that R&D and inventor locations have the most pronounced localization patterns. The urban economics literature has argued that R&D and innovation depend on new knowledge more than most forms of investment, and R&D is particularly sensitive to knowledge spillovers (e.g., Audretsch and Feldman 2004). Consistent with this view, empirical work has found that knowledge spillovers for R&D tend to operate at the smallest spatial scales of the agglomeration forces (e.g., Buzard and Smith 2017; Carlino and Kerr 2014b). While these studies document that R&D is more geographically concentrated than other economic activity (e.g., employment), our model generates a theoretically motivated rationale for why R&D can benefit to a greater extent from agglomeration forces.

4. Excess Localization and Firm Behavior

In this section, we implement three tests related to Proposition 3, which highlight the implications of agglomeration for corporate investment decisions. That is, we study corporate investment behavior, conditional on firm locations. First, we study whether the interaction between localization and uncertainty/knowledge intensity influences corporate investment levels. Second, we implement a novel approach that borrows from spatial econometrics, which helps address the inherent difficulties in estimating and interpreting empirical models that feature externalities. Third, we implement an approach that isolates the similarity and timing of investment opportunities in excess of what can be explained by selection into regions or industries based on observable criteria. We conclude this section by exploring implications for firm performance.

21. Our sample includes inventor locations for over 151,000 U.S. patents from 2006 to 2009. For each patent, we observe the address of the inventor's office, rather than the firm's HQ location. We convert inventor addresses into geographic coordinates and re-estimate the kernel densities. We lose seven industries relative to our main tests, since some industries do not have sufficient patenting activity. In particular, we impose the restriction that an industry must have at least 20 inventors during the 4-year sample period.

4.1 Corporate Investment

4.1.1. Traditional Empirical Framework

We start by testing Proposition 3 in a traditional panel data framework. Specifically, we test whether a firm's investment level (i.e., CapEx and R&D) is affected by the investment level of co-located firms by estimating the following equation:

$$\begin{aligned} Investment_{j,t}^{i,a} &= \delta + \beta_1 Investment_{p,t}^{i,a} + \gamma Investment_{p,t}^{i,a} \times UKI_{index} \\ &\quad + \beta_2 Investment_{p,t}^{i,-a} + \beta_3 Controls_t^{i,a} + \epsilon_{j,t}^{i,a}, \end{aligned} \tag{10}$$

where $Investment_{j,t}^{i,a}$ represents the capital expenditures (R&D) of firm j in industry i and area a during year t. The variable $Investment_{p,t}^{i,a}$ is an equally weighted portfolio (p) of firms within firm $j's$ industry (i) and its area (a). Similarly, $Investment_{p,t}^{i,-a}$ is the equally weighted portfolio of firms within firm $j's$ industry (i), but located outside its area (a). Firm j is excluded from belonging to its own peer group. As in Dougal, Parsons, and Titman (2015), the control variables include firm and year fixed effects. We estimate Equation (10) with and without the interaction term, $Investment_{p,t}^{i,a} \times UKI_{index}$.

We report the results from estimating Equation (10) in Table 2. We define area a to be the 20-mile (Panel A) and the 40-mile (Panel B) concentric circle that surround the centroid of firm $j's$ HQ ZIP code.[22] The estimation results in both panels suggest that investment comovement is significantly stronger for industry peers located within a regional cluster than for industry peers located outside the cluster. Importantly for our channel, the interaction between the investment of firms located nearby and the UKI index is positive and significant (Columns 3 and 6). This result suggests that the increased comovement between firms within a cluster is increasing with the uncertainty and knowledge-intensity of an industry, which is consistent with Proposition 3. Indeed, all of the increased comovement within regional industry clusters is driven by firms in industries with higher uncertainty/knowledge-intensity. That is, the level term (*Industry Inv. within 20 mi*) becomes insignificant in the capital expenditure regression, but it changes sign and is statistically significant in the research and development regression.

22. We include the 40-mile concentric circles for robustness, since some firms do not have many peers within a 20-mile radius.

[INSERT TABLE 2 HERE]

These results relate to Dougal et al. (2015), who show that corporate investment intensity comoves with the investments of regional peers. Note that our focus is on the regional impact of firms in the same industries, whereas Dougal et al. focus on the impact of regional peers in adjacent industries. Furthermore, our model predicts that as investments materialize for a given firm, this process will induce more investment from co-located firms. Thus, we use the total dollar value of capital expenditures, and not investment as a percentage of assets, as in Dougal et al. (2015).

4.1.2. Spatial Econometric Framework

While our results on investment comovement are informative, the estimation of peer effects is challenging due to the reflection problem, as identified by Manski (1993). In particular, it is difficult to separate (a) effects due to interaction among peers from (b) effects due to selection and (c) exposure to common shocks. To mitigate these concerns, we implement spatial econometric techniques that are designed to addresses inherent difficulties in estimating and interpreting empirical models of peer effects (Jackson 2010; Grieser et al. 2019).

Spatial methods achieve identification by imposing structure on the data according to knowledge about the nature of interactions between firms (LeSage and Pace 2009; Kelejian and Piras 2017). As discussed earlier, in the context of geography, distance between corporate headquarters provides a reasonable approximation for the regional proximity between firms. Specifically, we estimate:

$$Y = \rho GY + X\beta + GX\delta + \epsilon, \tag{11}$$

where Y represents the outcome variables (CapEx and R&D), X includes the same covariates as in Section 4.1.1 (including time and firm fixed effects) and the matrix $G \equiv [g_{ij}]$ contains information on the regional proximity between firms. Specifically, the element $g_{ij} = d(HQ_j, HQ_i)^{-p}$ is the inverse Haversine (i.e., great circle) distance between the geographic coordinates of firm i and firm j. We truncate the inverse distances to a maximum value of 1, and we assign a value of 0 to all firm pairs that are greater than 100 miles apart. Greater values of p assign greater influence to close distances. We use the standard value of $p = 1/2$. The diagonal elements g_{ii} are set to zero to preclude a firm from being a regional peer to itself. The intransitive nature of the matrix G is important because variation in group sizes and variation in the strength of pairwise relationships yields identification of peer effects in Spatial Durbin Models such as Equation (11) must be transformed to eliminate all instances of the outcome variable on the right-hand side to avoid a simultaneity bias:

$$Y = (I - \rho G)^{-1}[X\beta + GX\delta + \epsilon]. \tag{12}$$

This transformation results in a nonlinear model with parameters that can be consistently estimated via Markov Chain Monte Carlo procedures. We refer the reader to LeSage and Pace (2009) for a more in-depth discussion.

The scalar parameter, ρ, often referred to as the *peer interaction* parameter, summarizes the strength of interactions, and it ranges between 0 and 1. In our setting, a value of 0 implies no interaction in investment decisions, and a value of 1 implies that firms' investment decisions are perfectly jointly determined. Proposition 3 suggests that ρ should be greater for firms inside a regional industry cluster than for firms outside the cluster. Further, the gap between the investment intensity of firms within a cluster and firms outside a cluster should be increasing in the uncertainty and knowledge-intensity of an industry.

To test Proposition 3, we split each one of the 24 industries in our sample into firms with a high (low) number of peer firms within 20 miles, based on median industry values. We then estimate Equation (12) separately for each of the 48 groups (i.e., two groups for each of the 24 industries). The differences between estimates for firms within an industry cluster (ρ_{In}) and for firms in the same industry, but outside the cluster (ρ_{Out}, are reported in Figure 4, with industries sorted according to the UKI index.

[INSERT FIGURE 4 HERE]

The pattern in Figure 4 is consistent with the predictions of Proposition 3. Specifically, the externalities (i.e., the strength of interaction) is stronger within clusters for 18 of the 24 industries. Further, the greater strength of the interaction within clusters is more pronounced for industries with a higher UKI index value. Thus, it appears that the strongest investment externalities are exhibited by firms within industry clusters for the most uncertain/knowledge-intensive industries.

4.1.3. Residual Similarity Framework

To further mitigate concerns described above regarding selection effects, we modify the approach used by Shue (2013) to capture similarity in the timing of investment decisions. These tests capture "excess" comovement beyond what can be explained by observable selection into industries or regions. In particular, we examine whether the CapEx and R&D of firms located nearby (i.e., within 20 miles) exhibit greater "excess" similarity within more uncertain and more knowledge-intensive industries. We follow two steps:

Step 1) For each firm–year, we obtain residuals by estimating the specification

$$Investment_{it} = \alpha + BX_{i,t-1} + \widetilde{r}_{it}, \tag{13}$$

where $X_{i,t-1}$ is a vector of firm characteristics and time dummies. The residual $\widetilde{r}_{it}$ captures the unexplained component of the investment variable (i.e., CapEx or R&D).

We report the results for our estimation of Step 1 in the Internet Appendix, Table IA.2. We estimate three specifications. In the first specification, we account for any static differences across HQ locations by implementing Core-Based Statistical Area (CBSA) fixed effects transformations. In the second specification, we additionally control for heterogeneity in the level of investment across industries by including industry dummies. In the third specification, we exploit within-firm variation by implementing firm fixed effects transformations.[23]

Step 2) For each possible pair of firms in a given industry, we calculate the absolute value of the differences in residuals from Step 1 and estimate the following specification:

$$\begin{aligned} |\widetilde{r}_{it} - \widetilde{r}_{jt}| = \beta_0 + \beta_1 1(d \leq 20miles)_{ij} \times UKI\ index_{ij} + \\ \beta_2 1(d \leq 20miles)_{ij} + \beta_3 UKI\ index_{ij} + \epsilon_{ijt}, \end{aligned} \tag{14}$$

where $1(d \leq 20miles)_{ij}$ is an indicator that takes the value of 1 if the HQ of firm i and firm j are within 20 miles, and 0 otherwise. If $\beta_1 < 0$, then firms headquartered within 20 miles of each other make more similar investment decisions in relatively more uncertain and more knowledge-intensive

23. Note that the firm fixed effects transformations subsume the industry dummies and CBSA-level effects, as these are invariant through time within a given firm.

industries, on average. Alternatively, we estimate similarities between changes in CapEx and R&D by estimating the following specification:

$$
\begin{aligned}
|(\widetilde{r}_{it} - \widetilde{r}_{i,t-1}) - (\widetilde{r}_{jt} - \widetilde{r}_{j,t-1})| = \beta_0 + \beta_1 1(d \leq 20miles)_{ij} \times UKI\ index_{ij} + \\
\beta_2 1(d \leq 20miles)_{ij} + \beta_3 UKI\ index_{ij} + \epsilon_{ijt}.
\end{aligned}
\tag{15}
$$

We report the estimates of Equation (14) in Panel A of Table 3. The statistically negative coefficient on $1(d \leq 20mi) \times UKI\ index$ in all specifications indicates that firms located within 20 miles of each other exhibit a greater degree of similarity in CapEx and R&D expenses in more uncertain and more knowledge-intensive industries.

[INSERT TABLE 3 HERE]

Finally, we report the estimates of Equation (15) in Panel B of Table 3. The coefficient on $1(d \leq 20mi) \times UKI\ index$ is negative and statistically significant for all specifications, which indicates that changes in CapEx and R&D expenses for firms headquartered nearby are more similar in more uncertain and knowledge-intensive industries. While we do not have the ideal case in which assignment is random, as in Shue (2013), we believe these tests improve on a simple regression framework by showing the degree of similarity in firm investment decisions as they relate to uncertainty and knowledge intensity. For selection effects to be driving our results, it would need to be the case that selection effects are much stronger in more uncertain and knowledge-intensive industries for reasons other than those proposed by our model.

The clustering of investment in more uncertain and more knowledge-intensive industries in the cross-section and, more importantly, through time is strongly supportive of our channel. The similarity of investment decisions, and the clustering of investment through time, are inputs specific to our model. Our results do not completely rule out other channels of agglomeration, but they do increase our confidence that coordination in the presence of knowledge spillovers plays a first-order role in firm clustering and investment decisions.

4.2. Firm Performance

Our model suggests that firms agglomerate to facilitate the sharing of private information in order to reduce uncertainty regarding project payouts. This process thereby improves project selection, particularly for firms in more uncertain and knowledge-intensive industries. While there are costs of agglomeration that may offset some of its benefits (e.g., higher rents), it is possible that improvement in project selection manifests in aggregate firm performance. In this section, we explore the relation between return on assets (ROA), proximity to industry peers, and the UKI index.

For each industry, we split our sample into firms with a high (low) number of peer firms within 20 miles, based on median values. We then calculate the average ROA for each of the two groups within each industry. Figure 5 plots the difference in ROA between the highly agglomerated group and the less agglomerated group for each of the 24 industries in our sample, along with a quadratic interpolation. Industries are sorted based on their UKI index. While this analysis is mostly descriptive, the figure indicates a generally positive relationship between the benefits of agglomeration and uncertainty/knowledge intensity.

[INSERT FIGURE 5 HERE]

5. Additional Evidence

In this section, we consider the relation between agglomeration and uncertainty/knowledge-intensity in a regression framework, which complements our analysis in Section 3.2. First, we consider the distance between competitor HQs as a function of the UKI index. Next, we study co-location decisions in a sample in which the complementarity assumption of our model is arguably more prevalent. Specifically, we study the relationship between customer–supplier HQ distance and industry uncertainty/knowledge-intensity.

5.1. Competitor Proximity

We start by considering the relation between competitor proximity and uncertainty/knowledge-intensity in a standard cross-sectional data framework. We calculate the pairwise distances for all competitor pairs, defined according to the 24 industries in our sample. Results from estimating the relation between the UKI index and the distance between competitor HQ locations are presented in Panel A of Table 4. The dependent variable in Columns (1) and (2) is the natural log of the distance between HQ locations (in miles) for all competitor pairs. In Columns (2) and (4)–(8) we control for the size (natural log of sales) of both firms, as well as the number of firms in the same Fama–French 48 industry classification. These results suggest that competitors in industries with the highest uncertainty/knowledge intensity locate 92.36 miles closer to each other, on average.[24]

[INSERT TABLE 4 HERE]

In Columns (3)–(7) of Table 4, we examine the relationship between uncertainty/ knowledge-intensity and the likelihood of competitors locating within 20 miles of each other, 20–40 miles, 40–60 miles, and 60–80 miles, respectively. The estimates suggest that competitors are 5.4 pp more likely to locate within 20 miles of of each other in the most uncertain/knowledge-intensive industry relative to firms in the least uncertain/knowledge-intensive industry. This 5.4 pp increase is economically large, representing 123.9% of the sample average propensity for competitors to locate within 20 miles. Consistent with our prior results, the relationship between the UKI index and proximity dissipates at longer distances. For instance, the coefficient estimate associated with uncertainty/knowledge-intensity is almost 12 times larger for distances within 20 miles than for distances between 20 and 40 miles. Further, the relationship between uncertainty/knowledge-intensity and competitor proximity is not statistically significant for distances between 40 and 60 miles (Column (6)), and it changes sign (and is statistically significant) for distances between 60 and 80 miles (Column (7)).

5.2. Customer–Supplier Proximity

Next, we consider the co-location decisions of customers and suppliers, a setting in which the complementarity assumption of our model is arguably more prevalent. Specifically, firms in bilateral relationships, such as customers and suppliers, are likely to develop relation-specific investments, which cause the firms' investments to be heavily complementary (e.g., Banerjee et al. 2008). Choosing an HQ location near customers can allow suppliers to learn information about the investment opportunities of the customers, which in turn can increase the precision of their expectations regarding their own investment payouts. Similarly, customers can learn from suppliers regarding the quality and timing of the production of intermediate goods.

24. The average log distance between competitors is 6.588 miles. Consequently, the marginal effect at the mean is exp(6.588)–exp(6.588–0.136), which yields 92.36 miles.

Following recent work, we identify suppliers and customers from the Compustat segment files (e.g., Fee and Thomas 2004; Banerjee et al. 2008; Hertzel et al. 2008). To map the information from the customer–supplier file to a firm's financial and HQ location information in Compustat, we implement the name-matching algorithm implemented by Fee and Thomas (2004).[25] Our final sample includes 2,323 customer–supplier pairs from 1997 to 2013.

In our sample, we can identify suppliers that are heavily dependent on customers, but we cannot necessarily identify customers that depend heavily on suppliers. For example, Walmart constitutes at least 10% of sales for six firms in our sample, but none of the referenced suppliers constitute 10% of Walmart's expenditures. The average customer is approximately 14 times larger than the average supplier in our sample, according to total assets. Thus, it is more likely that customer locations influence the location decisions of suppliers rather than supplier locations affecting the location decisions of customers, and we construct our tests accordingly. In particular, we examine the relationship between the uncertainty and knowledge intensity of customer industries and the distance from customers that suppliers choose to locate.[26]

Estimates for the relation between customer–supplier proximity and customer uncertainty/ knowledge-intensity are presented in Panel B of Table 4. The dependent variables correspond to those of Panel A. The coefficient estimates in Columns (1) and (2) suggest that when suppliers sell to customers in the most uncertain/knowledge-intensive industry, their HQs tend to be 172 miles closer to the customer, on average, when compared to suppliers that sell to customers that are in the least uncertain/knowledge-intensive industry.[27] Consistently, suppliers are 7.9 pp more likely to locate within 20 miles of customers in the most uncertain/knowledge-intensive industry relative to customers in the least uncertain/knowledge-intensive industry. This 7.9 pp increase is economically large, representing 89.3% of the sample average propensity for suppliers to locate near customers. Consistent with our prior results, the relationship between customer uncertainty/knowledge intensity and customer–supplier proximity dissipates (and then becomes negative) at longer distances.

25. We thank Ted Fee and Shawn Thomas for providing us with this algorithm, which was recently extended to include firms through 2013.

26. In Table IA.3 of the Internet Appendix, we perform a robustness exercise in which we view the co-location of customers and suppliers as a joint decision in relation to the combined uncertainty of the two adjoining industries, and we find similar effects.

27. The average log distance between customers and suppliers is 6.317 miles. Consequently, the marginal effect at the mean is exp(6.317)–exp(6.317–0.3721), which yields 172.11 miles.

6. Interpretation and Discussion

Our theoretical model predicts that the investment benefits that arise from agglomeration are larger in relatively more uncertain and knowledge-intensive industries. In addition, our model also predicts higher levels of investment for firms co-located in more uncertain industries. This relationship helps to differentiate our mechanism from some prominent competing explanations, such as input sharing (Helsley and Strange 2002) and labor matching (Helsley and Strange 1990), which do not currently explain why the gains from agglomeration should be related to firm uncertainty.

Labor pooling (e.g., Krugman 1991; Almazan, De Motta, and Titman 2007) is one mechanism that, under the right conditions, could yield similar results to ours regarding firm location decisions.[28] However, our predictions regarding investment behavior, and the empirical results in Section 4, diverge from the predictions of current labor-pooling models. In our setting, as investments materialize, they induce more investments from co-located firms. On the contrary, in a labor-pooling framework, the investments of firms do not trigger more investments for regional industry peers, because the productive value of those investments are fully reflected in expected wages. This is true under the assumption of perfect competition (Krugman 1991) as well as under monopsony conditions (Heiko Gerlach and Stahl 2009).[29]

Our model, and the empirical framework in Section 4.1.3, also exploit the idea that coordination plays a major role in explaining the location of firms and their investment behavior. If gains from agglomeration materialize through the coordination of investments, then we should expect the timing of such investments to be more strongly correlated for firms agglomerated in more uncertain industries. While the labor-pooling channel can partially explain the cross-sectional patterns of industry clustering, our model emphasizes the timing and similarity of investment decisions between co-located firms that is induced through knowledge spillovers.

We recognize the likelihood that many channels drive firm agglomeration decisions. However,

28. Almazan, De Motta, and Titman (2007) focus on the development of human capital within knowledge-based industries in an environment characterized by incomplete contracts. In their context, access to competitive labor markets (through co-location) alleviates economic inefficiencies, which also leads to greater benefits to clustering for firms in uncertain industries. In Krugman (1991) firms benefit from economies of scale by sharing a market for skills. The expected profits of firms increase with the volatility of firms' productivity shocks, generating greater incentives for firms in more uncertain industries to co-locate.

29. Heiko Gerlach and Stahl (2009) study labor pooling under monopsony conditions in which firms make strategic R&D investments. Under their setting, R&D investments actually generate negative externalities for other firms through the labor market as productive firms attract more workers.

it is not our aim to disprove these alternative channels, or even to measure the relative importance of each channel. Instead, our goal is to show that coordination (i.e., knowledge sharing) has an increasing effect in more uncertain and knowledge-intensive industries, beyond any effects driven by other channels.

7. Conclusion

We build on the global games literature by developing a theory in which investment decisions can be conceived of as games of incomplete information in which payouts depend on the decisions of related firms. We propose that when firms observe noisy private signals about investment opportunities, co-locating with related firms facilitates the sharing of private information, thus improving precision in project valuations and reducing inefficiencies that arise from strategic miscoordination. In this context, dense urban centers can be conceived of as a technology that facilitates face-to-face interaction and knowledge sharing, thus improving project selection.

Our model shows that the benefits from this process are greater for firms in relatively more uncertain and more knowledge-intensive industries. Consistent with this proposition, we use both firm HQ and patent inventor locations as a proxy for business activity to show that agglomeration patterns are significantly more pronounced for relatively more knowledge-intensive industries in relatively more uncertain environments. Further, we show that this pattern also holds for CapEx and R&D expenses. In addition, customer–supplier proximity is strongly and positively related to the uncertainty and knowledge-intensity of the customer's industry, and the same pattern holds for the proximity of competitors.

We also show substantial evidence consistent with our model in the context of corporate investment decisions. We show that clustering is positively related to investment levels for firms in relatively uncertain and knowledge-intensive industries, after controlling for time and firm fixed effects. Using a spatial econometrics framework, we show that investment externalities are stronger in firm clusters, and we show that this effect is increasing in the uncertainty/knowledge-intensity of industries. Finally, we use standard firm-level determinants of corporate investment to show that the unexplained component of investment behaves more similarly for agglomerated firms in more uncertain/knowledge-intensive industries. Overall, our results link knowledge sharing to intrinsic industry characteristics, and our results contribute to our understanding of the important phenomenon of firm clustering.

References

Almazan, Andres, Adolfo De Motta, and Sheridan Titman, 2007, Firm location and the creation and utilization of human capital, *Review of Economic Studies* 74, 1305–1327.

Almazan, Andres, Adolfo De Motta, Sheridan Titman, and Vahap Uysal, 2010, Financial structure, acquisition opportunities, and firm locations, *Journal of Finance* 65, 529–563.

Arzaghi, Mohammad, and J. Vernon Henderson, 2008, Networking off Madison Avenue, *Review of Economic Studies* 75, 1011–1038.

Audretsch, David B, and Maryann P. Feldman, 2004, Knowledge spillovers and the geography of innovation, *Handbook of Regional and Urban Economics* 4.

Audretsch, David B., and Paula E. Stephan, 1996, Company-scientist locational links: The case of biotechnology, *American Economic Review* 86, 641–652.

Banerjee, Shantanu, Sudipto Dasgupta, and Yungsan Kim, 2008, Buyer–supplier relationships and the stakeholder theory of capital structure, *Journal of Finance* 63, 2507–2552.

Buzard, Carlino Gerald A. Carr Jake K. Hunt Robert M., Kristy, and Tony E. Smith, 2017, The agglomeration of R&D labs, *Journal of Urban Economics* .

Buzard, Kristy, Gerald A. Carlino, Robert M. Hunt, Jake K. Carr, and Tony E. Smith, 2015, Localized knowledge spillovers: Evidence from the agglomeration of american R&D labs and patent data, Working paper.

Carlino, Gerald, and William R. Kerr, 2014a, Agglomeration and innovation, Working paper.

Carlino, Gerald A., and William R Kerr, 2014b, Agglomeration and innovation, *Handbook of Regional and Urban Economics* 5A.

Carlsson, Hans, and Eric Van Damme, 1993, Global games and equilibrium selection, *Econometrica* 61, 989–1018.

Combes, Pierre-Philippe, and Gilles Duranton, 2006, Labour pooling, labour poaching, and spatial clustering, *Regional Science and Urban Economics* 36, 1–28.

Couture, Victor, 2015, Knowledge spillovers in cities: An auction approach, *Journal of Economic Theory* 157, 668–698.

Davis, Morris A., Jonas Fisher, and Toni M. Whited, 2014, Macroeconomic implications of agglomeration, *Econometrica* 82, 731–764.

De la Roca, Jorge, Gianmarco Ottaviano, and Diego Puga, 2014, City of dreams, Working paper.

Dixit, Avinash K., and Robert S. Pindyck, 1994, *Investment Under Uncertainty* (Princeton University Press, Princeton, NJ).

Dougal, Casey, Christopher A. Parsons, and Sheridan Titman, 2015, Urban vibrancy and corporate growth, *Journal of Finance* 70, 163–210.

Duranton, Gilles, and Henry G. Overman, 2005, Testing for localization using micro-geographic data, *Review of Economic Studies* 72, 1077–1106.

Duranton, Gilles, and Diego Puga, 2004, Micro-foundations of urban agglomeration economies, *Handbook of Regional and Urban Economics* 4, 2063–2117.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr, 2010, What causes industry agglomeration? Evidence from coagglomeration patterns, *American Economic Review* 100, 1195–1213.

Engelberg, J., A. Ozoguz, and Sean Wang, 2018, Know thy neighbor: Industry clusters, information spillovers and market efficiency, *Journal of Financial and Quantitative Analysis* 53, 1937–1961.

Faggio, Giulia, Olmo Silva, and William C. Strange, 2014, Heterogeneous agglomeration Working paper.

Fee, C. Edward, and Shawn Thomas, 2004, Sources of gains in horizontal mergers: Evidence from customer, supplier, and rival firms, *Journal of Financial Economics* 74, 423–460.

Glaeser, Edward, and David Maré, 2001, Cities and skills, *Journal of Labor Economics* 19, 316–42.

Glaeser, Edward L., 1999, Learning in cities, *Journal of Urban Economics* 46, 254–277.

Grieser, William, James LeSage, and Morad Zekhnini, 2019, What explains regional externalities in firm behavior? .

Heiko Gerlach, Thomas Ronde, and Konrad Stahl, 2009, Labor pooling in R&D intensive industries, *Journal of Urban Economics* 65, 99–111.

Helsley, Robert, and William Strange, 1990, Matching and agglomeration economies in a system of cities, *Regional Science and Urban Economics* 20, 189–212.

Helsley, Robert, and William Strange, 2002, Innovation and input sharing, *Journal of Urban Economics* 51, 25–45.

Helsley, Robert W., and William C. Strange, 2004, Knowledge barter in cities, *Journal of Urban Economics* 56, 327–345.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi, 2014, Growth and structural transformation, in Philippe Aghion, and Steven N. Durlauf, eds., *Handbook of Economic Growth*, volume 2, 855–941.

Hertzel, Michael G., Zhi Li, Micah S. Officer, and Kimberly J. Rodgers, 2008, Inter-firm linkages and the wealth effects of financial distress along the supply chain, *Journal of Financial Economics* 87, 374–387.

Jackson, Matthew O, 2010, *Social and economic networks* (Princeton university press).

Kelejian, H., and G. Piras, 2017, *Spatial Econometrics* (Academic Press).

Krugman, Paul, 1991, Geography and trade, *Cambridge, MA: The MIT Press* .

Landier, Augustin, Vinay B. Nair, and Julie Wulf, 2007, Trade-offs in staying close: Corporate decision making and geographic dispersion, *Review of Financial Studies* 22, 1119–1148.

LeSage, J. P., and R. K. Pace, 2009, *Introduction to Spatial Econometrics* (Florida CRC Press).

Lucas, Robert E., 1988, On the mechanics of economic development, *Journal of Monetary Economics* 22, 3–42.

Lychagin, Sergey, 2016, Spillovers, absorptive capacity and agglomeration, *Journal of Urban Economics* 96, 17–35.

Manski, Charles F., 1993, Identification of endogenous social effects: The reflection problem, *The review of economic studies* 60, 531–542.

Marshall, Alfred, 1920, *Principles of Economics* (Macmillan, London).

Moretti, Enrico, 2004, Human capital externalities in cities, *Handbook of Regional and Urban Economics* 4, 2243–2291.

Morris, Stephen, and Hyun S. Shin, 2003, Global games: Theory and applications, in Mathias Dewatripont, Lars P. Hansen, and Stephen J. Turnovsky, eds., *Advances in Economics and Econometrics*, 56–114.

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny, 1989, Industrialization and the big push, *Journal of political economy* 97, 1003–1026.

Pirinsky, Christo, and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, *Journal of Finance* 61, 1991–2015.

Romer, Paul M., 1986, Increasing returns and long-run growth, *Journal of Political Economy* 1002–1037.

Rosenthal, Stuart S., and William C. Strange, 2004, Evidence on the nature and sources of agglomeration economies, *Handbook of Regional and Urban Economics* 4, 2119–2171.

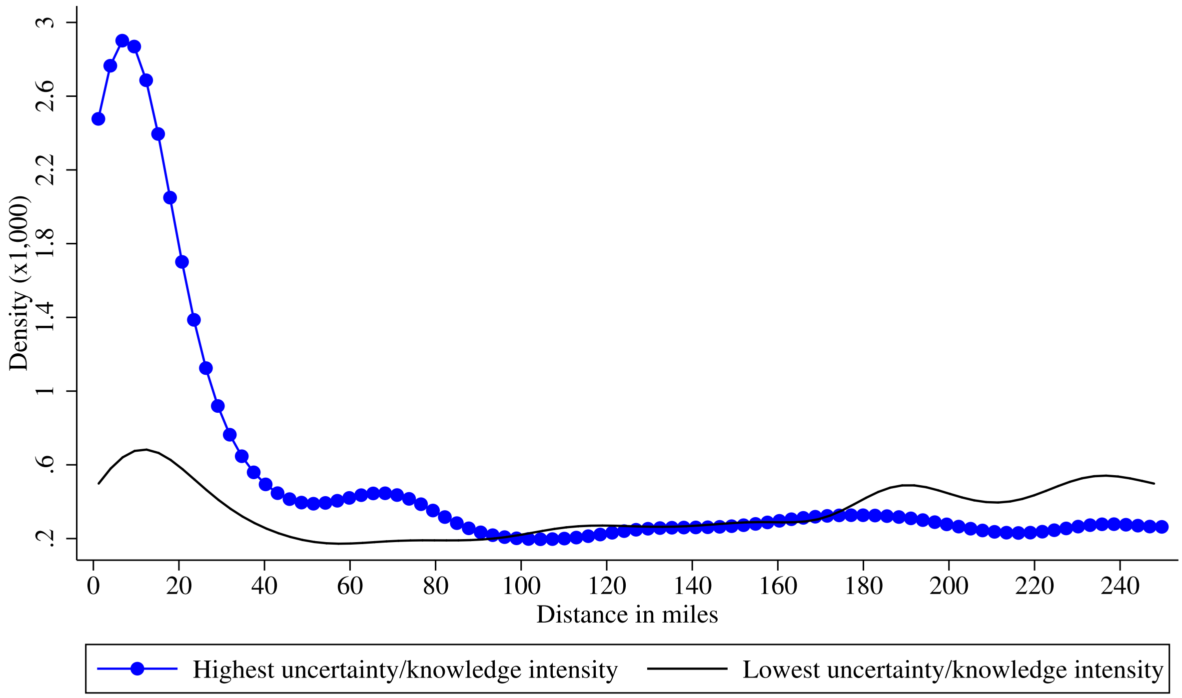
Rossi-Hansberg, Esteban, and Mark Wright, 2007, Urban structure and growth, *Review of Economic Studies* 74, 597–624.

Shue, Kelly, 2013, Executive networks and firm policies: Evidence from the random assignment of MBA peers, *Review of Financial Studies* 26, 1401–1442.

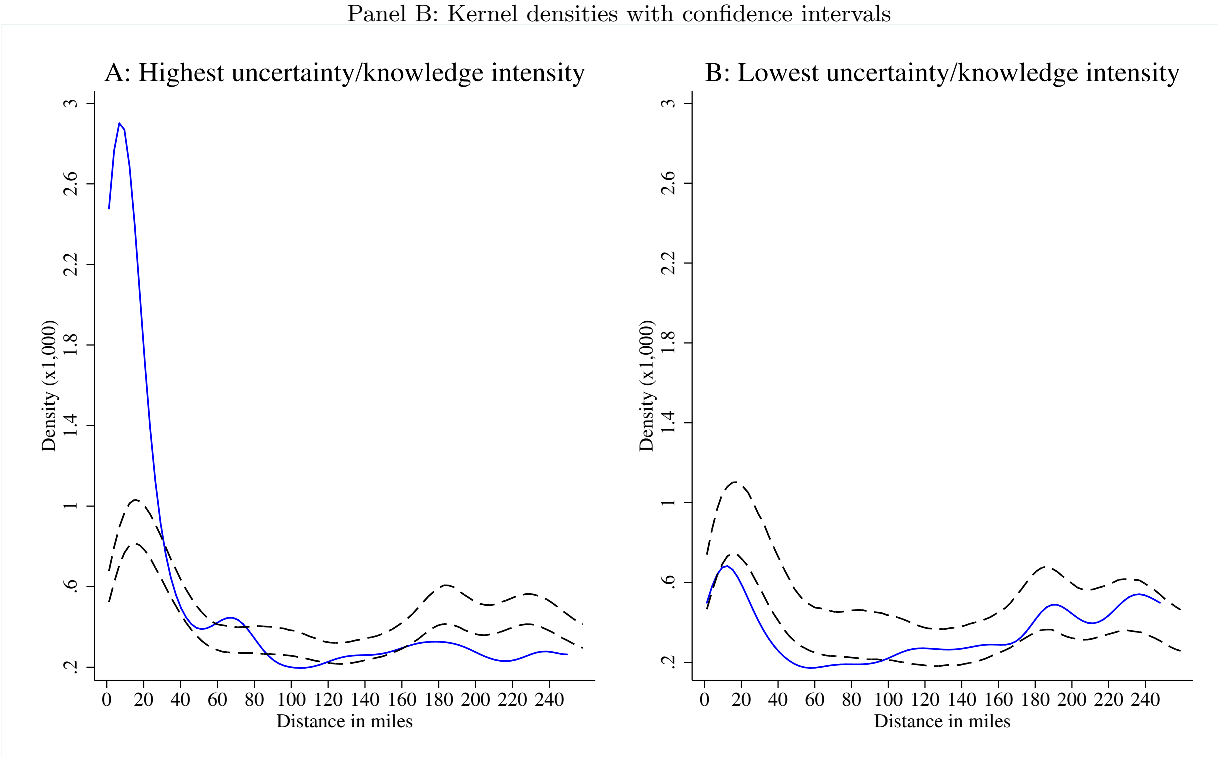
Storper, Michael, and Anthony J. Venables, 2004, Buzz: Face-to-face contact and the urban economy, *Journal of Economic Geography* 4, 351–370.

Figure 1. Kernel densities

Panel A: Kernel densities of highest vs. lowest uncertainty/knowledge intensity industry

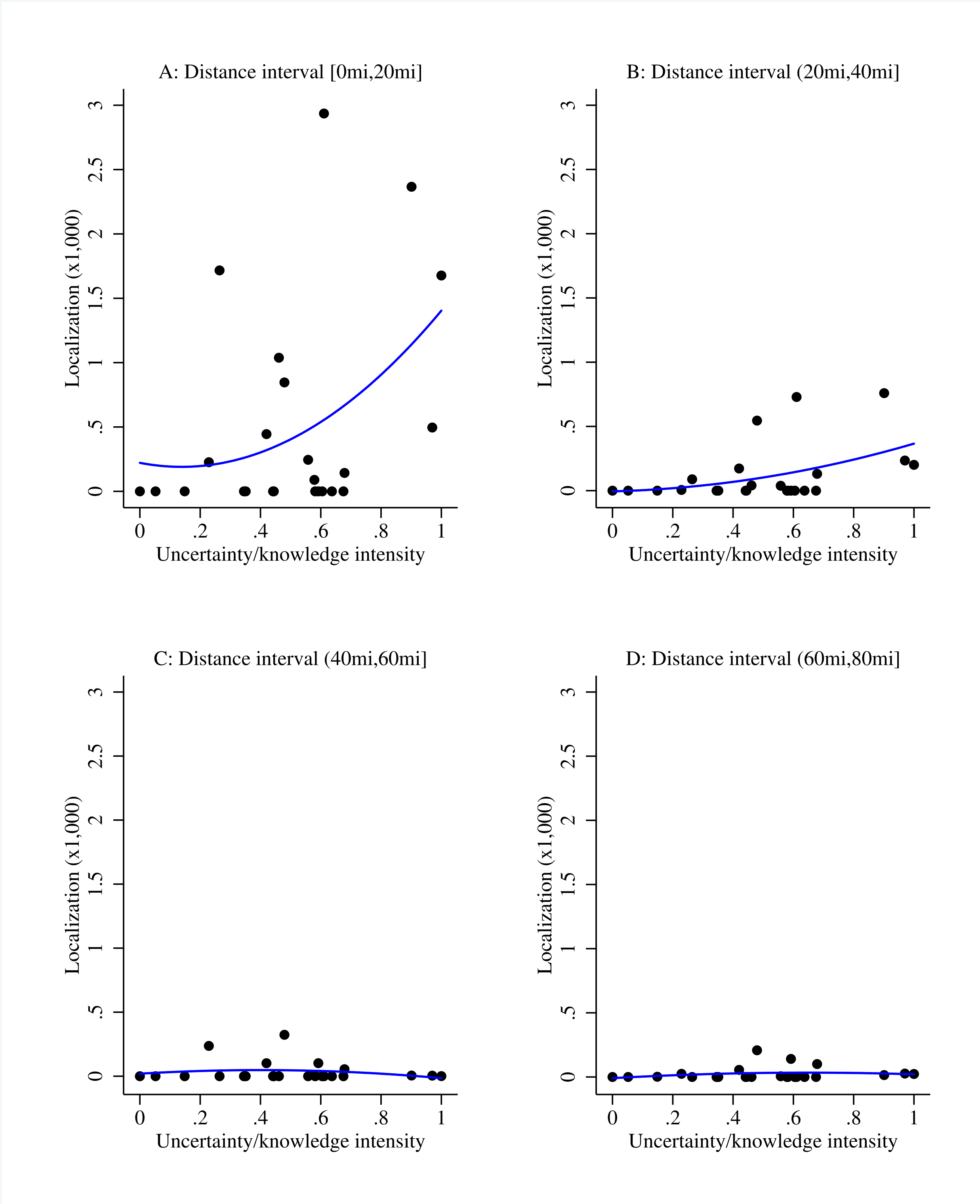


Panel B: Kernel densities with confidence intervals



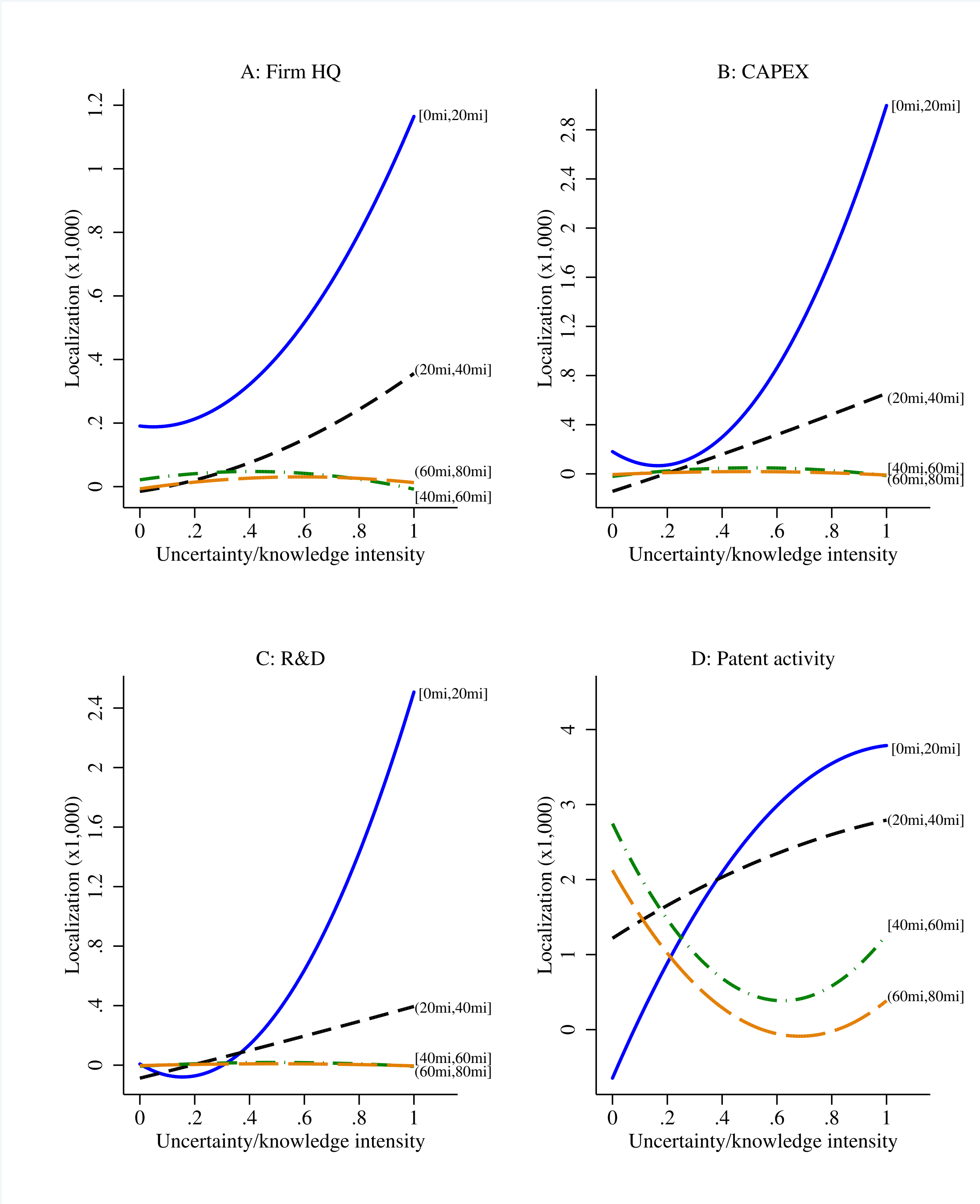
This figure plots the kernel densities of the highest and lowest uncertainty/knowledge intensity industries (×1,000 for scale). Industries are based on the Fama and French 48 industry classification (Finance and Utilities industries excluded). Industries are ranked by uncertainty/knowledge intensity based on stock index volatility and skill requirements. In Panel B, we plot the kernel densities of the highest and lowest uncertainty/knowledge intensity industries separately with their respective 95% confidence intervals.

Figure 2. Industry localization index and uncertainty/knowledge intensity by distance intervals



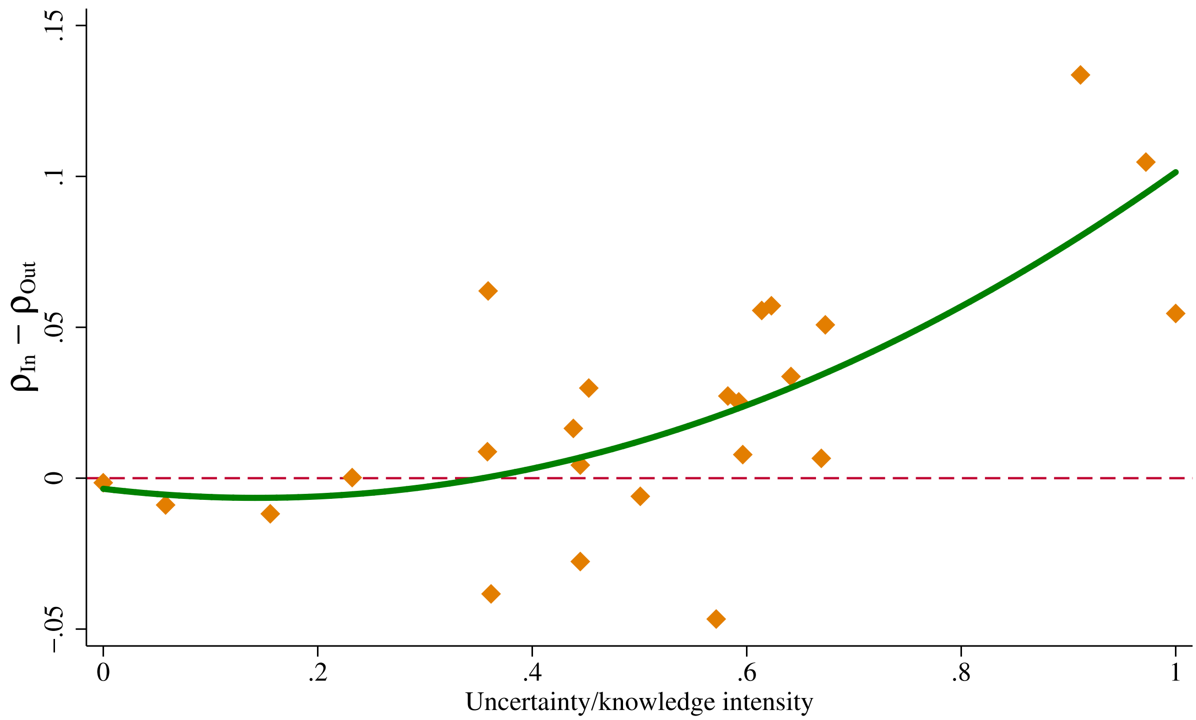
This figure plots the industry localization index (defined in Equation (9)) against the UKI index for various distance intervals. The solid line represents a quadratic interpolation.

Figure 3. Industry localization index and uncertainty/knowledge intensity by distance



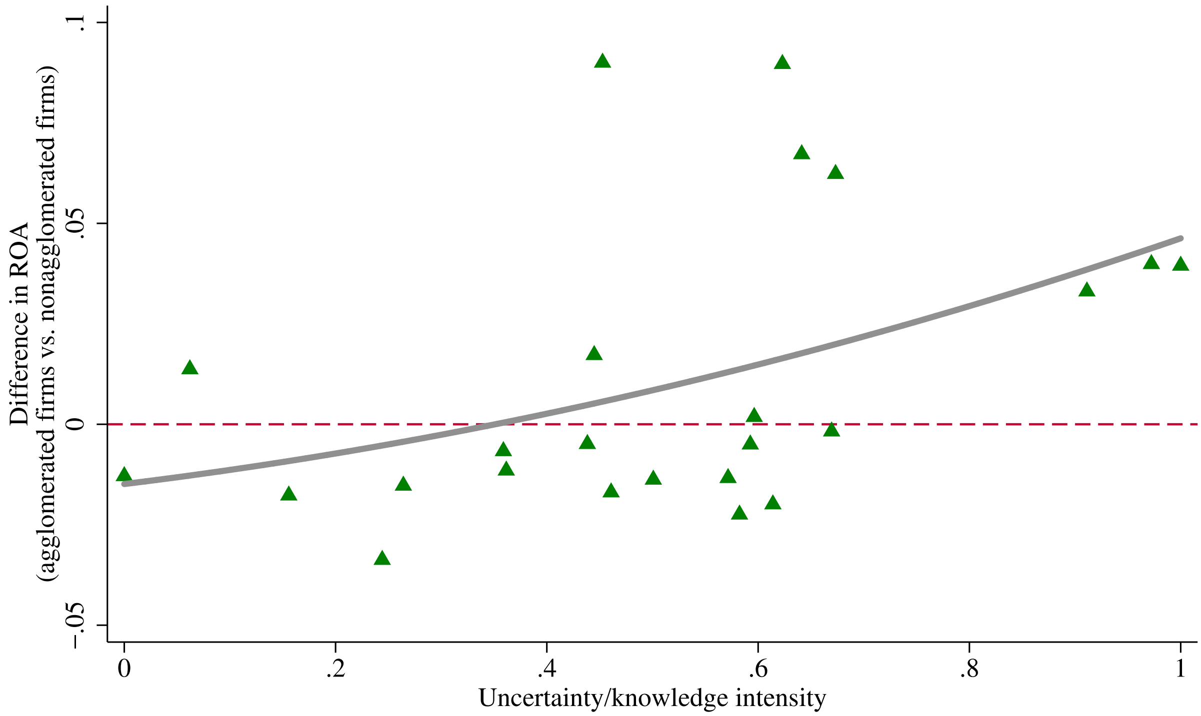
This figure plots the quadratic interpolation of industry localization index (defined in Equation (9)) against the UKI index for various distance intervals. Panel A considers firm headquarters. Panels B and C consider CapEx and R&D weighted headquarters locations, respectively. Panel D considers patent investor locations.

Figure 4. Strength of interaction among agglomerated firms



This figure plots differences in the strength of interaction among agglomerated firms (i.e., firms located within 20 miles of each other) and firms located at longer distances. Industries are sorted according to the UKI index (x-axis). The strength of firm interaction is measured by the parameter ρ in Equation (11).

Figure 5. Firm agglomeration and performance



This figure plots differences in the return on assets (ROA) between agglomerated firms (i.e., firms located within 20 miles of each other) and firms located at longer distances. Industries are sorted according to the UKI index (x-axis).

Table 1. Summary

Industry	Annualized volatility	Skill	Rank (volatility)	Rank (skill)	UKI index
Electronic equipment	0.339	3.508	1	4	1
Measuring and control equipment	0.322	3.556	2	2	0.970
Computers	0.271	3.763	6	1	0.901
Automobiles	0.320	2.734	4	17	0.679
Steel	0.320	2.718	3	18	0.675
Machinery	0.262	3.071	7	10	0.637
Oil	0.215	3.368	13	6	0.611
Personal Services	0.230	3.229	11	9	0.605
Electrical equipment	0.255	2.997	9	11	0.592
Construction	0.271	2.836	5	16	0.581
Healthcare	0.202	3.372	15	5	0.579
Telecommunications	0.200	3.328	16	7	0.558
Pharmaceuticals	0.147	3.524	22	3	0.479
Entertainment	0.260	2.580	8	20	0.461
Chemicals	0.211	2.920	14	13	0.444
Construction materials	0.239	2.688	10	19	0.442
Medical equipment	0.154	3.299	21	8	0.420
Transportation	0.185	2.853	18	14	0.351
Wholesale	0.168	2.969	20	12	0.345
Clothing	0.229	2.255	12	23	0.264
Household consumer goods	0.141	2.844	24	15	0.229
Retail	0.173	2.367	19	21	0.149
Food	0.142	2.329	23	22	0.052
Meals, restaurants, and hotels	0.188	1.812	17	24	0

This table ranks the various industries based on stock index volatility and skill requirements. To construct the uncertainty/knowledge-intensity index (UKI index), the volatility and skill metrics are standardized and averaged. Then, the resulting values are normalized so that the UKI index ranges from 0 to 1.

Table 2. Agglomeration, UKI, and Investment Expenditures

Panel A: 20-mile Concentric Geographic Areas

	Capex			R&D		
	(1)	(2)	(3)	(4)	(5)	(6)
Industry Inv. within 20mi	0.0052***	0.0077***	0.0006	0.0021***	0.0018***	-0.0237***
	(0.0012)	(0.0016)	(0.0038)	(0.0004)	(0.0004)	(0.0033)
Industry Inv. Outside 20mi	0.0034***	0.0054***	0.0054***	-0.0006***	-0.0007***	-0.0006***
	(0.0002)	(0.0008)	(0.0008)	(0.0001)	(0.0001)	(0.0001)
Industry Inv. within 20mi			0.0112**			0.0404***
×UKI index			(0.0046)			(0.0054)
Year FE	yes	yes	yes	yes	yes	yes
Firm FE	no	yes	yes	no	yes	yes
R^2	0.0282	0.6657	0.6657	0.0059	0.6732	0.6739
N	96,002	96,002	96,002	96,002	96,002	96,002

Panel B: 40-mile Concentric Geographic Areas

	Capex			R&D		
	(1)	(2)	(3)	(4)	(5)	(6)
Industry Inv. within 40mi	0.0062***	0.0084***	0.0032	0.0018***	0.0016***	-0.0166***
	(0.0011)	(0.0013)	(0.0033)	(0.0003)	(0.0003)	(0.0024)
Industry Inv. Outside 40mi	0.0031***	0.0052***	0.0052***	-0.0006***	-0.0007***	-0.0006***
	(0.0002)	(0.0008)	(0.0008)	(0.0001)	(0.0001)	(0.0001)
Industry Inv. within 40mi			0.0083*			0.0290***
×UKI index			(0.0043)			(0.0038)
Year FE	yes	yes	yes	yes	yes	yes
Firm FE	no	yes	yes	no	yes	yes
R^2	0.0279	0.6625	0.6625	0.0058	0.6741	0.6745
N	96,002	96,002	96,002	96,002	96,002	96,002

This table reports estimates for the relation between investment comovement, regional proximity, and uncertainty/knowledge intensity. The dependent variable in Columns (1)–(3) ((4)–(6)) is the natural log of capital expenditures (R&D). The independent variable *Industry Inv within 20mi (40mi)* is an equally weighted portfolio (p) of firms within firm $j's$ industry (i) and its 20-mile (40-mile) area (a). Similarly, *Industry Inv Outside 20mi (40mi)* is the equally weighted portfolio of firms within firm $j's$ industry (i) but located outside its area (a). Firm j is excluded from the calculation of each portfolio. Columns 3 and 6 include the interaction term ($Investment^{i,a}_{p,t} \times UKI_{index}$). All columns include year fixed effects, and Columns 2-3 and 5-6 include firm fixed effects. Standard errors are clustered at the Fama–French 48 industry level and are reported in parentheses, below the coefficient estimates. ***p<0.01, **p<0.05, *p<0.1.

Table 3. Excess Clustering of CapEx and R&D

Panel A: Residual Distance

	CapEx			R&D		
	(1)	(2)	(3)	(4)	(5)	(6)
1(d≤20mi)×UKI index	-0.0339***	-0.0354***	-0.0299***	-0.0491***	-0.0512***	-0.0355***
	(0.0121)	(0.0121)	(0.0081)	(0.0143)	(0.0154)	(0.0098)
1(d≤20mi)	0.0340***	0.0384***	0.0307***	0.0243**	0.0261**	0.0232***
	(0.0118)	(0.0117)	(0.0080)	(0.0098)	(0.0105)	(0.0068)
UKI Index	-0.0116**	-0.0100*	-0.0039	0.0311**	0.0394**	0.0199*
	(0.0051)	(0.0051)	(0.0035)	(0.0156)	(0.0172)	(0.0107)
First stage specification:						
Firm controls	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
CBSA FE	yes	yes	no	yes	yes	no
Industry FE	no	yes	no	no	yes	no
Firm FE	no	no	yes	no	no	yes

Panel B: Residual Changes

	CapEx			R&D		
	(1)	(2)	(3)	(4)	(5)	(6)
1(d≤20mi)×UKI index	-0.0385***	-0.0382***	-0.0378***	-0.0403***	-0.0380***	-0.0376***
	(0.0105)	(0.0105)	(0.0101)	(0.0127)	(0.0129)	(0.0121)
1(d≤20mi)	0.0369***	0.0367***	0.0362***	0.0275***	0.0254***	0.0251***
	(0.0103)	(0.0103)	(0.0099)	(0.0089)	(0.0092)	(0.0086)
UKI Index	-0.0019	-0.0019	-0.0021	0.0230*	0.0232*	0.0177
	(0.0048)	(0.0048)	(0.0048)	(0.0139)	(0.0135)	(0.0126)
First stage specification:						
Firm controls	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes
CBSA FE	yes	yes	no	yes	yes	no
Industry FE	no	yes	no	no	yes	no
Firm FE	no	no	yes	no	no	yes

This table reports the results of the second stage of the 2-step test of residual distances. The results of the first step, in which CapEx and R&D are regressed on lagged firm controls to obtain the residuals, are reported in Table IA.2. Panel A reports the results of the estimation of Equation (14), in which the dependent variable is the absolute value of the residual distances for all possible pairs of firms in an industry. The explanatory variable of interest is $1(d \leq 20mi) \times UKI\ index$, which is the interaction between the UKI index and a dummy variable that takes the value of 1 if the pair of firms are headquartered within 20 miles of each other and 0 otherwise. Panel B reports the results of the estimation of Equation (15), where the dependent variable is the absolute pair difference in changes in the first-stage residuals. Standard errors are clustered at the Fama–French 48 industry–year level and are reported in parentheses, below the coefficient estimates. ***p<0.01, **p<0.05, *p<0.1.

Table 4. Location and Uncertainty

Panel A: Industry

	log(distance)		Within 20mi		20-40mi	40-60mi	60-80mi
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
UKI index$_{customer}$	-0.0503***	-0.1363***	0.0434***	0.0541***	0.0047***	0.0007	-0.0023***
	(0.0143)	(0.0228)	(0.0031)	(0.0041)	(0.0017)	(0.0010)	(0.0008)
log(Sales)$_{Firm1}$		-0.0171***		0.0014***	0.0004*	-0.0003***	-0.0004***
		(0.0025)		(0.0003)	(0.0002)	(0.0001)	(0.0001)
log(Sales)$_{Firm2}$		-0.0172***		0.0014***	0.0004***	-0.0003***	-0.0004***
		(0.0008)		(0.0001)	(0.0001)	(0.0001)	(0.0001)
Number of firms$_{FF48}$		0.0120***		-0.0013***	-0.0005***	-0.0001***	0.0000
		(0.0008)		(0.0001)	(0.0001)	(0.0000)	(0.0000)
R^2	0.0029	0.0038	0.0016	0.0035	0.0008	0.0006	0.0001
N	3,730,680	3,730,680	3,730,680	3,730,680	3,567,790	3,491,016	3,457,266

Panel B: Supplier Locations

	log(distance)		Within 20mi		20-40mi	40-60mi	60-80mi
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
UKI index$_{customer}$	-0.4288***	-0.3721***	0.0942***	0.0785**	0.0132	0.0141	-0.0148*
	(0.1274)	(0.1361)	(0.0321)	(0.0323)	(0.0085)	(0.0119)	(0.0084)
log(Sales)$_{customer}$		0.0767		-0.0121*	0.0004	-0.0050	0.0015
		(0.0469)		(0.0063)	(0.0021)	(0.0036)	(0.0018)
log(Sales)$_{supplier}$		-0.0465**		0.0037*	-0.0029*	0.0008	-0.0026
		(0.0215)		(0.0019)	(0.0015)	(0.0017)	(0.0030)
Number of firms$_{customer}$		0.0288		-0.0060	0.0044	-0.0024	0.0000
		(0.0349)		(0.0063)	(0.0029)	(0.0040)	(0.0014)
Number of firms$_{supplier}$		0.0202		0.0005	0.0004	0.0016	-0.0023
		(0.0269)		(0.0051)	(0.0025)	(0.0025)	(0.0029)
R^2	0.0445	0.0514	0.0694	0.0747	0.0241	0.0260	0.0167
N	2,323	2,323	2,323	2,323	2,148	2,096	2,068

This table reports estimates for the relation between customer and supplier locations and customer uncertainty/knowledge intensity. Customers are identified from the Compustat segment files. The dependent variable in Columns (1)–(2) is the natural log of the distance (in miles) between a customer and a supplier HQ location.The dependent variables in Columns (3)–(7) are indicators of whether a customer and supplier are located within 20 miles, between 20 miles and 40 miles, between 40 miles and 60 miles, or between 60 miles and 80 miles, respectively. Distances are calculated from geographic coordinates for corporate HQ ZIP codes. The customer–supplier pair is the unit of observation in this panel. Each customer–supplier pair is represented by a single observation (i.e., this is a cross-sectional analysis). Standard errors are clustered at the Fama–French 48 industry level and are reported in parentheses, below the coefficient estimates. ***p<0.01, **p<0.05, *p<0.1.

Appendix

A. Proofs

Proof of Proposition 1

We must prove that the gains from clustering are higher for riskier industries and more complex industries. To do so, we show that the gains from clustering are higher for firms in industries with higher risk.

Let us denote the gains from clustering as the difference between the expected value of being clustered versus the expected value of not being clustered. Thus,

$$f(\tau,\sigma) = E(V_C) - E(V_{NC}) = E(\kappa = 1/2) - \int [\bar{\theta} - \Phi(\frac{\kappa + \frac{\sigma^2}{\tau^2}(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}})] f(\bar{\theta}) d\bar{\theta}. \tag{A.1}$$

We need to show that for any $\tau_H > \tau_L \Rightarrow f(\tau_H, \sigma) > f(\tau_L, \sigma)$. We do this as follows. First, we show this is the case in the limit, then we show that gains are a continuous function, thus monotonicity entails

$$f(\tau_H \to \infty, \sigma) > f(\tau_L \to 0, \sigma) = 0. \tag{A.2}$$

This is equivalent to showing that

$$E(\kappa = 1/2) - lim_{\tau\to\infty} \int [\bar{\theta} - \Phi(\frac{\kappa + \frac{\sigma^2}{\tau^2}(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}})] f(\bar{\theta}) d\bar{\theta} > 0, \tag{A.3}$$

and

$$E(\kappa = 1/2) - \int \bar{\theta} f(\bar{\theta}) d\bar{\theta} + \int [\Phi(lim_{\tau\to\infty} \frac{\kappa + \frac{\sigma^2}{\tau^2}(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}})] f(\bar{\theta}) d\bar{\theta} > 0. \tag{A.4}$$

Using L'Hospital's rule to calculate the limit, we obtain

$$E(\kappa = 1/2) + [\Phi(\frac{\kappa - \bar{\theta}}{\sqrt{2\sigma^2}})] > 0, \tag{A.5}$$

which is positive. Notice that Φ, which is the CDF of a normal random variable, is a continuous function, thus the increase in gains from agglomeration is also a continuous function. This implies that $f(\tau_H, \sigma) > f(\tau_L, \sigma)$.

Similarly, we must show that the gains from clustering are higher for firms in industries that are more complex. We define a *complex* industry as one in which the cost of privately learning the state of the world is very high, thus private assessments are very noisy. Specifically, we must show that for any $\sigma_H > \sigma_L \Rightarrow f(\tau, \sigma_H) > f(\tau, \sigma_L)$. We first show that this is the case in the limit. Then, we show that the increase in gains is a continuous function and thus entails monotonicity. We want to show $f(\tau, \sigma_H \to \infty) > f(\tau, \sigma_L \to 0) = 0$. This is equivalent t to showing that

$$E(\kappa = 1/2) - lim_{\sigma\to\infty} \int [\bar{\theta} - \Phi(\frac{\kappa_H + \frac{\sigma^2}{\tau^2}(\kappa_H - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}})] f(\bar{\theta}) d\bar{\theta} > 0, \tag{A.6}$$

and

$$E(\kappa = 1/2) - \int \bar{\theta} f(\bar{\theta}) d\bar{\theta} + \int [\Phi(lim_{\sigma\to\infty} \frac{\kappa + \frac{\sigma^2}{\tau^2}(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}})] f(\bar{\theta}) d\bar{\theta} > 0. \tag{A.7}$$

Using L'Hospital's rule to calculate the limit, we obtain

$$E(\kappa = 1/2) + [\Phi(\frac{\kappa - y}{\tau^2})] > 0, \tag{A.8}$$

which is positive. Notice that Φ, which is the CDF of a normal random variable, is a continuous function, thus gains from agglomeration is also a continuous function. This implies that $f(\tau, \sigma_H) > f(\tau, \sigma_L)$.

Proof of Proposition 2

Let us define the expected gains from co-location in industry i as the difference between the expected value of being co-located versus not being co-located, which is

$$E(V_C^i) - E(V_{NC}^i).$$

We assume that by being co-located, firms can share knowledge, thus they operate on the limit equilibrium where there is no strategic miscoordination. On the other hand, firms that do not co-locate cannot share their private signals with nearby firms. Thus,

$$E(V_C^i) = E(\kappa = 1/2)$$

and

$$E(V_{NC}^i) = \int [\bar{\theta} - \Phi \left\{ \frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] f(\bar{\theta}) d\bar{\theta},$$

where $\bar{\theta} \sim N(0, \hat{\sigma})$.

Note that since equilibrium firms follow a cutoff strategy in which investment occurs if and only if $\bar{\theta} > \kappa$. Given that the value of not investing is equal to zero, then the expected value of not co-locating is given by

$$E(V_{NC}^i) = \int_{\kappa}^{\infty} [\bar{\theta} - \Phi \left\{ \frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] f(\bar{\theta}) d\bar{\theta}.$$

We evaluate how the marginal value of co-locating varies for industries with different uncertainty levels. For this, we take the derivative of the gains from co-locating with respect to κ:

$$\frac{d}{d\kappa}[E(V_C^i) - E(V_{NC}^i)] = \frac{d}{d\kappa}[E(\kappa = 1/2) - E(V_{NC}^i)] = -\frac{d}{d\kappa}\{ \int_{\kappa}^{\infty} [\bar{\theta} - \Phi \left\{ \frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] f(\bar{\theta}) d\bar{\theta}\},$$

since at the limit, the expected value of co-locating is not a function of the cutoff.

Using the Leibniz formula, we can represent the above expression as

$$= -[\kappa - \Phi \left\{ \frac{(\sigma^2/\tau^2)(\kappa - y)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] + \int_{\kappa}^{\infty} \frac{d}{d\kappa}[\bar{\theta} - \Phi \left\{ \frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] f(\bar{\theta}) d\bar{\theta}\}.$$

Notice that the equilibrium condition states that $[\kappa - \Phi \left\{ \frac{(\sigma^2/\tau^2)(\kappa - y)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] = 0$, thus

$$= -[\frac{1-(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}] \int_{\kappa}^{\infty} \phi \left\{ \frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] f(\bar{\theta}) d\bar{\theta}.$$

Changing variables results in

$$\frac{d}{d\kappa}\{\int_{\kappa}^{\infty} [\bar{\theta} - \Phi \left\{ \frac{\kappa + (\sigma^2/\tau^2)(\kappa - y) - \bar{\theta}}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}} \right\}] f(\bar{\theta}) d\bar{\theta}\} = -[\frac{1-(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}] \int_{\kappa}^{\infty} \phi[\zeta - \sqrt{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}}\bar{\theta}]\phi[\bar{\theta}]d\bar{\theta}.$$

Since $\bar{\theta} \sim N(0, \hat{\sigma})$ and $\zeta - \sqrt{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}}\bar{\theta} \sim N(\zeta, \frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}\hat{\sigma})$, the product of two Gaussian distributions is a scaled Gaussian distribution where

$$\mu_S = (\frac{\zeta}{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}\hat{\sigma}})(\frac{1}{\hat{\sigma}} + \frac{1}{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}\hat{\sigma}})$$

and

$$\sigma_S = \sqrt{1/(\frac{1}{\hat{\sigma}} + \frac{1}{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}\hat{\sigma}})},$$

with a scale factor of

$$S = \frac{1}{\sqrt{2\pi\frac{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}\hat{\sigma}^2}{\sigma}}} exp\{-\frac{1}{2}(\frac{\zeta^2}{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}\hat{\sigma}^2})\sigma\}.$$

Therefore,

$$-[\frac{1-(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}] \int_{\kappa}^{\infty} \phi[\zeta - \sqrt{\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2+\sigma^4}}\bar{\theta}]\phi[\bar{\theta}]d\bar{\theta} = [\frac{1-(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}][1 - \Phi_S(\kappa)],$$

which means that the marginal gain of co-locating is given by

$$\frac{d}{d\kappa}[E(V_C^i) - E(V_{NC}^i)] = [\frac{1-(\sigma^2/\tau^2)}{\sqrt{\frac{2\sigma^2\tau^2+\sigma^4}{\sigma^2+\tau^2}}}][1 - \Phi_S(\kappa)] > 0.$$

The above expression means that the value of co-locating is higher for more uncertain industries. Moreover, the gains from co-locating are monotonic in τ. This is the case, because $\frac{\partial\kappa}{\partial\tau} > 0$.

The above expression also means that the value of co-locating is higher for more complex industries. Furthermore, from the above expression, we see that the gains from co-locating are monotonic in σ. This is the case, because $\frac{\partial \kappa}{\partial \sigma} < 0$

Let us assume a fixed cost of co-locating. Then, for any fixed cost of co-location C, there exists $\bar{\tau}$ such that $\forall \tau^i > \bar{\tau}$, and there exists $\bar{\sigma}$ such that $\forall \sigma^i > \bar{\sigma}$. This means

$$[E(V_C^i) - E(V_{NC}^i)] > C.$$

Then, only the industries with an uncertainty and complexity level above the cutoff should co-locate. All other industries would decide not to co-locate.

Proof of Proposition 3

We have two industries, in which one industry features higher uncertainty than the other, i.e., $\overline{\tau} > \underline{\tau}$.

The investment intensity of any given a firm (given that all firms use a switching strategy around κ) is determined by the proportion of firms that receive a private signal higher than κ.

This is given by

$$1 - \Phi(\frac{\kappa - \theta}{\sigma}),$$

where θ is the true state of the world and σ captures the noisiness of a private signal. We must show that

$$lim_{\sigma \to 0}[1 - \Phi(\frac{\overline{\kappa} - \theta}{\overline{\sigma}})] - [1 - \Phi(\frac{\overline{\kappa} - \theta}{\overline{\sigma}})] > lim_{\sigma \to 0}[1 - \Phi(\frac{\underline{\kappa} - \theta}{\underline{\sigma}})] - [1 - \Phi(\frac{\underline{\kappa} - \theta}{\underline{\sigma}})].$$

Notice that

$$lim_{\sigma \to 0}[1 - \Phi(\frac{\overline{\kappa} - \theta}{\overline{\sigma}})] = lim_{\sigma \to 0}[1 - \Phi(\frac{\underline{\kappa} - \theta}{\underline{\sigma}})]$$

since, in the limit, all firms follow the same strategy. That is, they observe a cutoff equal to 1/2. Thus, we must show that

$$[1-\Phi(\frac{\underline{\kappa}-\theta}{\underline{\sigma}})] > [1-\Phi(\frac{\overline{\kappa}-\theta}{\overline{\sigma}})] \Rightarrow \underline{\kappa}-\theta < \frac{\underline{\sigma}}{\overline{\sigma}}(\overline{\kappa}-\theta)$$

and

$$\Leftrightarrow \underline{\kappa}-\theta < \sqrt{\frac{\sigma^2\bar{\tau}^2+\underline{\tau}^2\overline{\tau}^2}{\sigma^2\underline{\tau}^2+\underline{\tau}^2\overline{\tau}^2}}(\overline{\kappa}-\theta).$$

We must ensure this condition always holds. First, notice that when $\sigma \to 0$, the above expression converges to $\underline{\kappa} < \overline{\kappa}$. And, when $\sigma \to \infty$, the above expression converges to $\underline{\kappa} < \frac{\overline{\tau}}{\underline{\tau}}\overline{\kappa}$.

Morris and Shin (2002) showed that investment intensity is lower for higher uncertainty. Thus, in equilibrium, there must be a higher cutoff. This means that the above conditions hold.

Agglomeration, Coordination, and Corporate Investment

Internet Appendix

This appendix is divided into two sections. The first section provides supplementary tables and figures. The second section provides an alternative model of information sharing.

A. Supplementary Tables and Figures

Table IA.1. Firm summary statistics

	Mean	SD	p25	p50	p75
Log(Assets)	3.23	2.76	1.60	3.44	5.09
ROA	-0.13	0.30	-0.20	0.01	0.07
Log(Size)	3.28	2.57	1.50	3.38	5.08
Market leverage	0.16	0.19	0.00	0.09	0.25
Investment	0.34	22.70	0.01	0.03	0.07
R&D	7.4	13.1	0.0	0.0	7.4

This table describes the 9,167 firms in the main sample. The variables *Assets*, *Size* (Market Cap), and *R&D* are denominated in millions of dollars.

Table IA.2. First step of residual regressions

	CapEx			R&D		
	(1)	(2)	(3)	(4)	(5)	(6)
Leverage	0.0073***	-0.0031**	-0.0326***	0.0029	0.0261***	0.0186***
	(1.6036e-03)	(1.5245e-03)	(2.1593e-03)	(2.5763e-03)	(2.4257e-03)	(2.6030e-03)
log(Sales)	-0.0031***	-0.0005	-0.0004	-0.0302***	-0.0144***	0.0013*
	(3.5809e-04)	(3.7145e-04)	(5.5686e-04)	(5.7601e-04)	(5.9133e-04)	(6.7027e-04)
log(Assets)	0.0033***	-0.0010**	-0.0217***	0.0099***	-0.0013**	-0.0392***
	(3.9957e-04)	(4.0178e-04)	(6.8945e-04)	(6.4191e-04)	(6.3896e-04)	(8.2467e-04)
Market-tobook	0.0000***	0.0000***	0.0000***	0.0000**	0.0000	0.0000**
	(3.7120e-06)	(3.4330e-06)	(3.2339e-06)	(5.7933e-06)	(5.3043e-06)	(3.7767e-06)
Z-score	-0.0000**	-0.0000*	0.0000**	0.0000	0.0000	0.0000***
	(3.5313e-07)	(3.2620e-07)	(3.1788e-07)	(5.6250e-07)	(5.1462e-07)	(3.7659e-07)
ROA	-0.0000***	-0.0000***	-0.0001***	0.0000**	0.0000	-0.0000***
	(4.9285e-06)	(4.5499e-06)	(6.7358e-06)	(7.8685e-06)	(7.1939e-06)	(7.8780e-06)
Year FE	yes	yes	yes	yes	yes	yes
CBSA FE	yes	yes	no	yes	yes	no
Industry FE	no	yes	no	no	yes	no
Firm FE	no	no	yes	no	no	yes
N	45308	45308	45308	45640	45640	45640
R^2	0.127	0.257	0.593	0.251	0.376	0.803

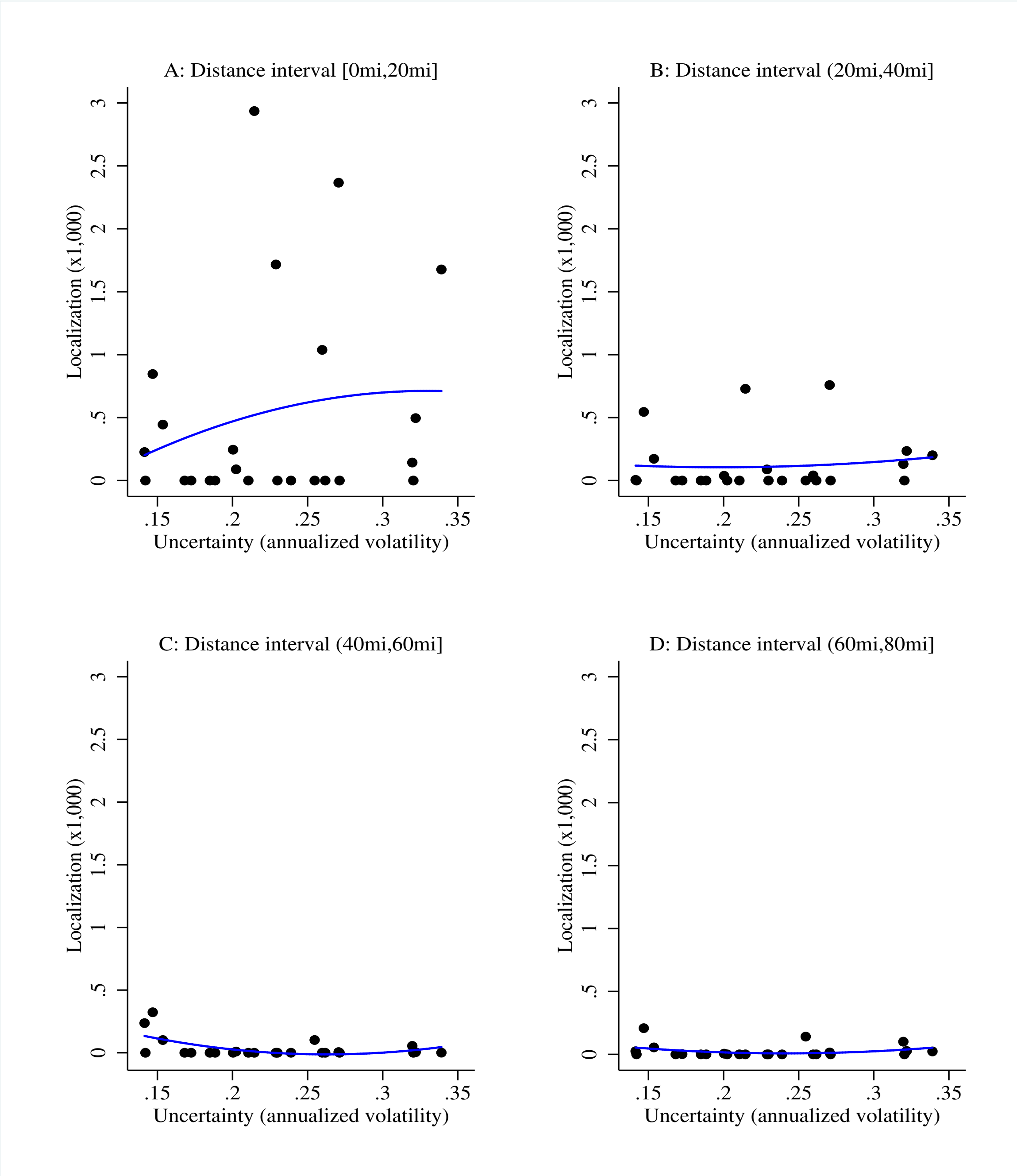
This table reports the results of the first step of the 2-stage procedure presented in Table 3. Specifically, CapEx and R&D are regressed on lagged firm controls to obtain the residuals for Step 2.

Table IA.3. Robustness for Table 4

	log(distance)		Within 20mi	
UKI index$_{combined}$	-0.2921*	-0.2760	0.0753***	0.0642**
	(0.1691)	(0.1885)	(0.0275)	(0.0311)
log(Sales)$_{customer}$		0.0785***		-0.0125**
		(0.0287)		(0.0051)
log(Sales)$_{supplier}$		-0.0458**		0.0036
		(0.0194)		(0.0029)
Number of firms$_{customer}$		0.0375		-0.0075
		(0.0344)		(0.0059)
Number of firms$_{supplier}$		0.0251		-0.0009
		(0.0343)		(0.0054)
R^2	0.0426	0.0503	0.0671	0.0734
N	2,323	2,323	2,323	2,323

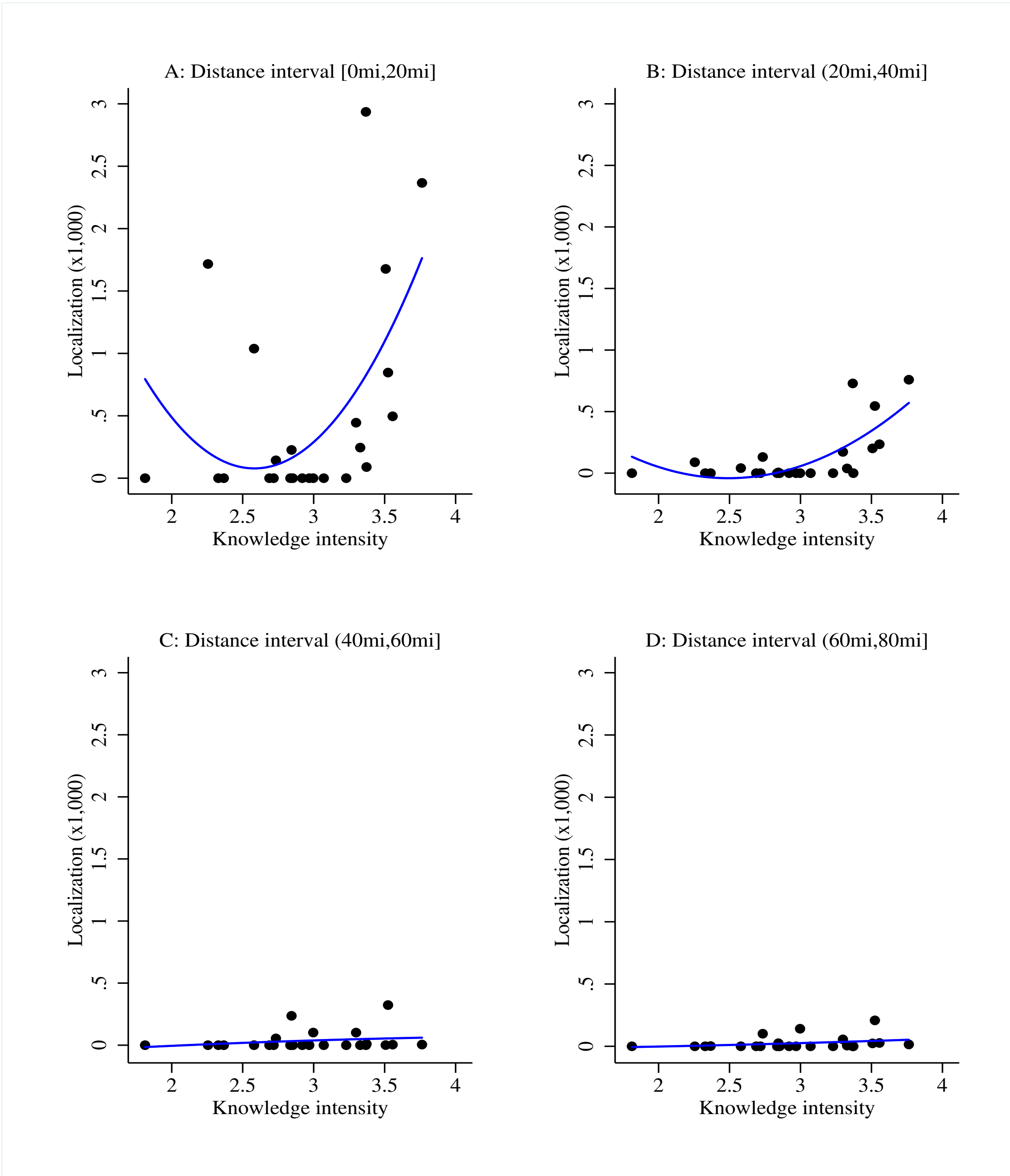
This table repeats the estimations in Table 4, with the only difference being that the UKI index of the customer and the UKI index of the supplier are averaged. Standard errors clustered at the Fama–French 48 industry level and are reported in parentheses, below the coefficient estimates. ***p<0.01, **p<0.05, *p<0.1.

Figure IA.1. Industry localization index and uncertainty



This figure plots the industry localization index (defined in Equation (9)) against industry uncertainty for different distance intervals. The solid line represents a quadratic interpolation.

Figure IA.2. Industry localization index and knowledge intensity



This figure plots the industry localization index (defined in Equation (9)) against industry knowledge intensity for different distance intervals. The solid line represents a quadratic interpolation.

B. Alternative Model of Information Sharing: The Pure Information Channel

Let us assume that a principal must choose whether to invest in an uncertain project in the presence of asymmetric information regarding the quality of the project. In this setting, we do not assume any externalities from investment; therefore, we denominate the model as a *purely informational narrative*. The timing of this game is as follows:

1. The principal decides to evaluate an investment opportunity.
2. In the research process, the principal observes an imperfect signal of the project's quality.
3. Conditional on the signal, the principal decides whether she wants to invest in the project or in a safe asset instead (i.e., replacement).
4. Production takes place.

There is a common prior:

$$\beta \sim N(B, 1/\tau),$$

where β is the actual quality of a project, and B is the expected quality of projects in a given industry. $1/\tau$ is the precision of the distribution (or the inverse of the second moment-variance), and τ captures the uncertainty of the industry. We assume this is all common knowledge.

After the principal evaluates the project, she observes a private and imperfect signal y about the project's quality, where

$$y \sim N(\beta, 1/\rho).$$

That is, on average, the principal is expected to make the right inference about the project's quality, but with noise. The precision of the principal's assessment is given by $1/\rho$.

The principal can always choose not to invest, and thus receive a normalized value equal to zero. Each financed project yields a return equal to x. In particular, $E(x) = \beta$. Markets are competitive, so they clear at the expected value of the project's quality. The principal's payoff is given by $x - Investment = x - E(\beta)$.

Let us define $\hat{\beta}$ as the conditional posterior of the project's quality that the principal acquires after researching. This means that:

$$\hat{\beta} = E(x|y).$$

Using the properties of the normal distribution,

$$\hat{\beta} = \frac{\tau B + \rho y}{\tau + \rho}.$$

Funding rule

A project is financed as long as $\hat{\beta} > 0$, which necessarily implies the following funding rule:

$$y \geq -\frac{\tau B}{\rho} = C.$$

Therefore, whenever the principal observes a signal y higher than C, then the principal funds the project. Otherwise, the principal invests in a safe asset instead, with a normalized expected return of zero.

The expected value of investment

Let y be the signal. We must characterize the distribution of y before observing the signal.

$E(y) = B$ and $Var(y - B) = Var(y - \beta) + Var(\beta - B) = 1/\rho + 1/\tau = \frac{\rho+\tau}{\rho\tau} \equiv H$, thus $y \sim N(B, 1/H)$. The value of investment before observing y is given by

$$V = \int_{-\infty}^{\infty} max\{0, \hat{\beta}\} \sqrt{\frac{H}{2\pi}} exp(-\frac{H}{2}[y - B]^2) dy > B.$$

Thus, there is a value option of investing ex ante. Let $Z \equiv \sqrt{H}(y - B) \Rightarrow dy = dZ/\sqrt{H}$. Then

$$V = \int_{-\infty}^{\infty} max\{0, B + \frac{Z\sqrt{H}}{\tau}\} \sqrt{\frac{H}{2\pi}} exp(-\frac{1}{2}[Z]^2) dZ/\sqrt{H} = \int_{-\infty}^{\infty} max\{0, B + \frac{Z\sqrt{H}}{\tau}\} \sqrt{\frac{1}{2\pi}} exp(-\frac{1}{2}[Z]^2) dZ,$$

where $\phi(Z) = \sqrt{\frac{1}{2\pi}} exp(-\frac{1}{2}[Z]^2)$ is the probability density function of a standard normal distribution. Note that $B + \frac{Z\sqrt{H}}{\tau} > 0$ when $Z \geq \frac{-B\tau}{\sqrt{H}}$. Thus, the integral is given by:

$$V = \int_{\frac{-B\tau}{\sqrt{H}}}^{\infty} (B + \frac{Z\sqrt{H}}{\tau}) \phi(Z) dZ.$$

Furthermore,

$$\frac{-B\tau}{\sqrt{H}} = -B\frac{\tau+\rho}{\tau+\rho}\frac{\tau}{\sqrt{H}} = -B\frac{\tau+\rho}{\rho}\sqrt{H} = (C-B)\sqrt{H}$$

and

$$V = \int_{\sqrt{H}(C-B)}^{\infty}(B+\frac{Z\sqrt{H}}{\tau})\phi(Z)dZ = B[1-\Phi(\sqrt{H}(C-B)] + \frac{\sqrt{H}}{\tau}\phi(\sqrt{H}(C-B).$$

Let us define $\theta = V - B$ as the option value of investment before observing the realization of a signal.

Proposition B.1

The value V and the option value θ of investing are increasing in the precision of the signal ρ.

Proof:

$$\frac{\partial V}{\partial \rho} = \frac{\partial \theta}{\partial \rho} = \frac{\tau}{2(\rho+\tau)}\frac{1}{\sqrt{H}}\phi(\sqrt{H}(C-B)) > 0.$$

Therefore, information sharing, which can increase the precision of private signals, translates into ex-ante value for firms.

Proposition B.2

The value V and option value θ of investing are decreasing in the precision of the prior τ.

Proof:

$$\frac{\partial V}{\partial \tau} = \frac{\partial \theta}{\partial \tau} = (\frac{\rho^2}{2\tau(\rho+\tau)^2\sqrt{H}} - \frac{\sqrt{H}}{\tau^2})\phi(\sqrt{H}(C-B)) < 0$$

This means that option value of investing is higher for more uncertain industries.

Proposition B.3

The gains from increasing the signal ρ (agglomeration) precision in V and θ can be larger or smaller for more uncertain industries.

Proof:

$$\frac{\partial^2 V}{\partial \tau \partial \tau} = \frac{1}{2\tau(\rho+\tau)^2\sqrt{H}}[\frac{2\rho+\frac{1}{\tau H}}{2(\rho+\tau)}\phi(\sqrt{H}(C-B)) - \tau\frac{\rho+2\tau^2 BH}{2\tau^2\rho\sqrt{H}}\phi^{'}(\sqrt{H}(C-B))],$$

which can be positive or negative depending on parameter values. The crucial implication of these results is that firms in more uncertain industries do not always obtain larger gains from agglomeration, because in a pure information setting, the option value of investing is already higher for more uncertain industries. Therefore, it is not clear that gains from informational sharing are more important for firms in those particular industries.