NBER WORKING PAPER SERIES

PROXIMITY AND ECONOMIC ACTIVITY: AN ANALYSIS OF VENDOR UNIVERSITY TRANSACTIONS

Nathan Goldschlag
Julia Lane
Bruce Weinberg
Nikolas Zolas

Working Paper 23018
http://www.nber.org/papers/w23018

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2017, Revised April 2018

Previously circulated as "Research Funding and Regional Economies." Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau or the National Bureau of Economic Research. All results have been reviewed to ensure that no confidential information is disclosed. This research was supported by the National Center for Science and Engineering Statistics. NSF SciSIP Awards 1064220 and 1262447; NSF Education and Human Resources DGE Awards 1348691, 1547507, 1348701, 1535399, 1535370; NSF NCSES award 1423706; The National Institute on Aging and the Office of Behavioral and Social Science Research via P01AG039347 (on which Weinberg and his work were supported directly by the National Bureau of Economic Research and indirectly by Ohio State); and the Ewing Marion Kaufman and Alfred P. Sloan Foundations. Lane was supported through an Intergovernment Personnel Act assignment to the US Census Bureau. We thank Cameron Conrad, Ahmad Emad, Evgeny Klochikhin, for research support, Greg Carr, Marietta Harrison, David Mayo, Mark Sweet, Jeff Van Horn and Stephanie Willis for help with data issues, and, Barbara McFadden Allen, Jay Walsh, Roy Weiss, and Carol Whitacre for their continuing support. Chris Goetz provided extremely helpful comments and suggestions. The research agenda draws on work with many coauthors, but particularly Patrick Llerena, Ron Jarmin, and Jason Owen Smith.

Proximity and Economic Activity: An Analysis of Vendor University Transactions
Nathan Goldschlag, Julia Lane, Bruce Weinberg, and Nikolas Zolas
NBER Working Paper No. 23018
January 2017, Revised April 2018
JEL No. O31,R1

ABSTRACT

This paper using transaction based data to provide new insights into the link between the geographic proximity of businesses and associated economic activity. It contributes to the literature by developing both two new measures of distance and a set of stylized facts on the distances between observed transactions between vendors and customers for a research intensive sector – universities. We show that spending on research inputs is more likely to be expended at businesses physically closer to universities than those farther away. That relationship is stronger for High Tech and R&D performing businesses than businesses in general, which is consistent with theories emphasizing the role of tacit knowledge. We find that firms behave in a way that is consistent with the notion that propinquity is good for business: if a firm supplies a project at a university in a given year, it is more likely to open an establishment near that university in subsequent years than other firms. We also investigate the link between transactional distance and economic activity and show that when a vendor has been a supplier to a project at least one time, that vendor is subsequently more likely to be a vendor on the same or related project.

Nathan Goldschlag
Center for Economic Studies
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
Nathan.Goldschlag@census.gov

Julia Lane
Wagner School of Public Service and CUSP
New York University
295 Lafayette Street
New York, NY 10012-9604
julia.lane@nyu.edu

Bruce Weinberg
Department of Economics
Ohio State University
410 Arps Hall
1945 North High Street
Columbus, OH 43210
and NBER
weinberg.27@osu.edu

Nikolas Zolas
Center for Economic Studies
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233
nikolas.j.zolas@census.gov

1. Introduction

While there is now a large literature studying the role of geographic proximity in economic activity, a paucity of data means that links between R&D activity and economic activity are generally inferred rather than measured formally (Doloreux and Shearmur 2011). This article uses a unique new dataset matched to rich Census data on establishments to directly connect economic activity to research. Specifically, instead of starting from observed distance and then seeking to infer an effect of geography on economic activity (e.g. based on the joint distribution of firm locations), we observe actual economic transactions between research-intensive projects within universities and businesses supplying inputs to those projects. We then use these observed transactions to examine the role of distance in the probability of a transaction and the type of transaction. We also relate firm behavior, including the opening of new establishments and closing of existing establishments to transactions. Our contribution is that we provide a new, albeit noncausal, approach to connecting the dots between R&D activity at universities and economic activity at businesses. In addition, we shed some, but far from complete, light on mechanisms.

The economic geography literature has identified several possible pathways whereby geographical proximity can affect economic activity. One is the importance of physical closeness for individuals to meet and transmit tacit knowledge about complex processes (Boschma 2005; Gertler 2003). Another is that the networks important for business transactions are more readily formed through face-to-face interactions (Boschma 2005; Owen-Smith and Powell 2004). Informal knowledge spillovers, transportation costs, and land costs are also often cited as important factors (McCann 2007). There is also evidence that geographic closeness matters more for firms with short-term product life cycles, in which customer contact is important, and those with short-term contracts are more likely to benefit from face-to-face interactions than others (McCann 2007).

A more narrowly focused, parallel literature has attempted to describe the link between the R&D-intensive research and education industries and economic growth. The public support of research typically relies on the notion that universities are engines of economic development, and that university research is a primary driver of R&D-intensive, high-wage, localized, economic activity. In that literature, the link between university research and economic activity has been identified as a key driver of regional economic growth (Nelson 2005; Saxenian 1996) for many of the same reasons identified in the economic geography literature.

As indicated, we contribute to the literature by using data on actual transactions between universities and businesses, which is possible because of access to a unique dataset. It combines longitudinal transactional data from 13 research-intensive universities on all research related purchases from suppliers with longitudinal data on both those suppliers and all other US businesses from the US Census Bureau. The resulting transactional data set includes information about all research purchases from all vendors so that direct economic links can be measured in terms of volume, value, and persistence. Project-level transactional information about the people on each project can be used to study the potential role of networks in the purchasing patterns of research inputs. The links to US Census data enable us to see how distance is related to purchasing decisions controlling for a range of potential confounding factors, including industry, R&D activity and firm characteristics. The Census links also permit an examination of the role of distance in firm behavior – such as placing new establishments close to the universities they are supplying, as well as the link between geographic proximity and survival.

The paper begins by establishing some basic facts. We compare the characteristics of businesses that supply research at universities to the characteristics of all US firms and establishments. Establishments supplying research universities with inputs are systematically different from other establishments. They tend to be older, larger, pay higher wages, and are much more likely to be operated by an R&D performing firm than non-vendor establishments.

We then examine the role of geographical propinquity. Here we use our rich data to make four interrelated points. First, we show that vendors supplying research inputs are significantly more likely to be located near the university with which they do business than non-vendors. That relationship becomes stronger when the business relationship, as measured by the value of the transactions, is stronger.

Second, we show transactions with universities are related to economic activity, Specifically, vendors are more likely to locate new establishments near the universities with which they do business, and the strength of the relationship increases in accordance with the value of the transaction. Similar results hold among the subset of firms that open new establishments in a given year—those newly opened establishments are more likely to be near the universities with which they do business, especially if the transactions are large. Being a vendor to a university is also associated with a lower exit rate for establishments. While our estimates are not causal, these findings are consistent with universities stimulating local economic activity through their vendor purchases.

Third, we seek to shed some light on the mechanisms that result in the location of businesses to the extent that our data permits. Consistent with the notion that geographic closeness is important for the transmission of complex ideas, we show the relationship is also stronger for vendors that are High Tech or in establishments that are part of R&D performing firms. Our finding of greater localization for High Tech and at R&D performing firms is significant because it may indicate that they are able to leverage their work for universities to be more competitive. [2]

Lastly, we provide evidence consistent with the importance of networks in generating persistent business relationships. In doing so, we posit the importance of prior business (as opposed to social) relationships as a determinant related but distinct activity. Specifically, because we have longitudinal data on all transactions and all individuals on all funded projects, we are able to show substantial persistence in purchases from a given vendor. If researchers on a funded project purchase from a given vendor, they are 35-42% more likely to purchase from that same vendor in the following year. Additionally, if one project purchases from a given vendor other projects with overlapping faculty are also more likely to purchase from that vendor. Such a high degree of persistence can help explain why firms would locate their establishments based on their existing vendor relationships.

2. Background

There is an extensive literature on the links between geographical closeness and economic growth. The increasing importance of cities as loci of innovation is often cited as evidence that physical propinquity leads to greater economic growth (E. Glaeser 2011; Moretti 2012). The reasons range from the importance of *tacit knowledge* to the role of *networks* to the importance of informal *spillovers*.

[2] Unfortunately, we are not able to test this hypothesis further, nor are we able to show whether or how activity at vendor firms generates a multiplier by stimulating business at their suppliers because we do not have links between vendors and their suppliers. That said, making a direct link between customers and vendors is, in itself, an advance.

The transmission of *tacit knowledge* is thought to be a major reason for the importance of human contact, particularly when it is important to transmit complex ideas. In other words, when people are physically close to each other, skills are easier to acquire and knowledge is easier to exchange (Duranton and Puga 2004; Gertler 2003). In a wide ranging overview, Morgan (2004) points out that the notion that innovations in information and communication technology have resulted in the "death of geography" is simply incorrect (Morgan 2004). He argues that while codified information can readily be spread across distance, understanding can not. This view has important implications for economic activity since firms that are able to set up mechanisms whereby tacit knowledge can readily be transferred are more likely to succeed. Indeed, this view is consistent with the parallel literature on the importance of intangible assets as part of firm capital measurement (Van Ark et al. 2009; Corrado, Haskel, and Jona-Lasinio 2017; Corrado, Hulten, and Sichel 2009).

A closely related notion is that the *networks* important for business transactions are more readily formed through face-to-face interactions (Boschma 2005; Owen-Smith and Powell 2004). There is empirical support for the idea that geographic closeness matters more for firms with short-term product life cycles, in which customer contact is important, and for those firms with short-term contracts which are more likely to benefit from face-to-face interactions than others (McCann 2007). Informal knowledge *spillovers* are also often cited as important factors that are closely related to geography (McCann 2007).

Another strand of the literature discusses the complex measurement issues. Boschma (2005) points out that there may be alternative transmission mechanisms other than close contact, and describes four other related but distinct measures of closeness, such as cognitive, organizational, social and institutional propinquity(Boschma 2005).

The parallel literature on the regional economic impact of universities mentioned in the introduction is of particular interest in this context, because universities perform both basic and applied research. The existence of Silicon Valley has been traced to its propinquity to Stanford, Boston's growth has been attributed to its great research universities, and the Research Triangle to the research performed at Duke, University of North Carolina and North Carolina State University (Liu 2015). Indeed, Glaeser (2010) points out the importance of face-to-face interactions of researchers (E. L. Glaeser and Ponzetto 2010). Furthermore, university researchers are important customers of R&D performing firms; in life sciences, for example, about 8% of R&D funds are directly spent on equipment, while core research expenditures are likely to be much more (Stephan 2014).

These findings are consistent with work by Doloreux and Shearmur (2011), who use rich survey data to argue that innovation occurs differently depending on the distance from urban centers. They find that collaboration/competition are particularly important where there are specialized inputs, services and resources and knowledge, as is likely the case with high tech and R&D performing firms, but do not find collaboration and competition to vary continuously across space (Doloreux and Shearmur 2011). They posit two ideas. One is that geography acts to change the ability of firms to specialize in niche areas. The second is that geography affects the quality and frequency of collaboration. They draw heavily on McCann's view that "the frequency of face-to-face interaction between agents is itself a decision-variable in the innovation objective function of the firms, and the response to this variable itself depends on the geographical distance between the location of the firm and the location of face-to-face interaction " – and this is particularly true with "knowledge-exchange activities associated with core knowledge centres"(McCann 2007)

Relatively little is known about the role of geography in the purchasing patterns and economic impacts of university-based research. Much of the literature that has addressed the role of geography has relied on aggregate data with links inferred rather than directly measured(Hausman 2012; Kantor and Whalley 2014; Saha, Staudt, and Weinberg 2015). In the analyses that follow, we will leverage data that directly captures the transactions made by university researchers to investigate the role of geography.

3. The Conceptual and Empirical Framework

Our conceptual framework builds on the literature described above in two ways. The first is by extending the literature to incorporate a different unit of analysis. While much of the literature has focused on the role of cities as knowledge centers, our focus is on a different type of knowledge center - research universities. The second is contributing to the discussion of the mechanisms – the role of tacit knowledge, networks and informal spillovers. In what follows below, we will compare the establishments of high tech and R&D performing firms, for which tacit knowledge is likely to be most important, to other types of businesses. We connect to the spillover and network literature through an analysis of the propagation of vendor relationships between networks of researchers at universities.

The empirical framework adds an additional dimension of distance to those identified by Boschma (2005). Each of his dimensions of cognitive, organizational, social and institutional propinquity should exogenously determine economic activity[3]. However, because we have repeated data on transactions across economic units, we are able to extend the measurement approach to examine whether economic activity induces its own form of business relationship

[3] We are grateful to an anonymous referee for pointing this distinction out.

distance (which is a function of past interactions)[4]. As such, in our empirical framework we introduce two new measures of business relationship distance. The first is direct, and captures the persistence of transactions between projects and vendors. The second is indirect, and captures the initiation of transactions between a project and a vendor which had no previous direct ties but where related projects had transactions with the vendor. Whereas social distance has been seen to be exogenous to economic activity, our concept of business relationship distance is, by its very nature, endogenous to economic activity. Although one might seek to identify quasi-random fluctuations in business relationship distance, our goal here is to provide an initial descriptive characterization. The second part of this section describes the structure of such models.

3.1 Physical Distance and Economic Activity

We start by investigating the role of physical distance of the establishment on the propensity for a random establishment to have a transaction with a given university. One prominent approach in the literature is to estimate the distribution of distances between establishments in the same industry (e.g. (Duranton and Puga 2004)). As indicated and given our transaction data, we turn this structure around and ask the question – how is distance related to the probability that a transaction occurred between a research project at a given university and any one of the roughly 6.7 million establishments in the United States?

This formulation leads to a model, which in a fairly general form, can be specified as

$$T_{jut} = \beta_0 + f\left(D_{ju}, \beta_1\right) + \beta_2 V_{jt} + \beta_3 G_{jt} + \gamma_u + \delta_t + \lambda_x + \varepsilon_{jut}. \qquad (1)$$

---

[4] Again, we are grateful to an anonymous referee for this insight.

Here, T_{jut} is a dummy variable for whether a transaction occurred between vendor j and university u, at time t; D_{ju} is the distance of vendor j to university u (note that by construction, establishment locations are fixed, so that D_{ju} does not vary with t); V_{jt} are the characteristics of vendor j at time t (some of which are time-varying such as firm age and size); G_{jt} are local economic characteristics pertaining to vendor j at time t (e.g. population in the 5-digit zip code in which j is located); γ_u denotes time-invariant university characteristics for university u; δ_t captures year-specific variation for year t; and λ_x captures time-invariant characteristics of 3-digit NAICS industries for industry x. For the distance function, f(.), we employ a flexible polynomial function in distance. We also provide tabulations by distance bins. The reason for doing so is to identify potential nonlinearities on the impact of distance.[5]

The model can be modified to address the possibility that establishments doing business with universities differ in ways that are unobservable from those that do not. To address this concern, we can take all the vendors in our data and estimate the probability that an establishment that sells to at least one university sells to each of the other universities. This approach ensures that (aside from proximity), all the establishments are plausible university vendors.[6]

As indicated, a frequent theme in the existing literature is the importance of distance where complex, tacit knowledge (Morgan 2004; Whittington, Owen-Smith, and Powell 2009) or

---

[5] For our discrete binary measure, we use a 100 mile radius. It should be noted however that any radius is arbitrary; while we report the 100 mile radius below, we experimented with both shorter and larger radii and found qualitatively similar results.

[6] Of course, we only observe transactions for the universities in our sample, which means that with data that are more comprehensive we could include vendors to universities that are not in our sample. This does not, however, invalidate our approach. Fortunately, our universities are geographically clustered and similar in many dimensions, which means that, aside from their location, the vendors to one university are plausible vendors for other universities.

short product cycles (McCann 2007) are important. While we do not have direct measures of tacit knowledge or product cycle length, it is plausible that establishments that are part of R&D performing firms or in High Tech industries are associated with the production of complex knowledge and have shorter product cycles. To get at this mechanism, we can both include interactions with industry and subset the sample to only include certain types of firms and establishments.

The second possible observable link between physical distance and economic activity is the opening or closure of establishments close to major customers. We look at whether and where multi-establishment firms that are vendors to a university open new establishments and how being a vendor to a university affects survival of establishments. While these results do not address causality they can be used to determine whether businesses behave in a way consistent with university purchases stimulating economic activity.[7]

The baseline model for establishment openings involves regressing whether a multi-establishment firm opens a new establishment near each university in 2014 as a function of whether that firm has been a vendor to the university in the prior year. Formally, our model is

$$E_{fu2014} = \beta_0 + \beta_1 T_{fu2013} + \beta_2 P_{fu2013} + \beta_3 \overline{D_{fu2013}} + \gamma_u + \theta_f + \varepsilon_{fut}. \qquad (2)$$

Here E_{fu2014} is a dummy variable equal to 1 if multi-establishment firm *f* opens a new establishment within 50 miles of university *u* in 2014.[8] P_{fu2013} captures the value of vendor payments from university *u* to firm *f* in 2013. T_{fu2013} indicates whether firm *f* was a vendor to

[7] Unfortunately, our research design, which uses a pre-existing relationship between a university and a vendor as a primitive to estimate whether that vendor open an establishment close to a university precludes us from looking at spinoffs directly.

[8] Notice that the subscript for vendors has changed to *f* to designate that we are focusing on the firm and no longer the establishment. Also note that in the analyses that follow, we focus exclusively on establishment openings in 2014, which is why we drop the time dummy.

university u in 2013. We control for time-invariant firm specific characteristics with θ_f. We measure transactions, distance (here $\overline{D_{fu2013}}$, the mean distance from each of the firm f's establishments to university u), and vendor characteristics with a one-year lag, so that they are not directly affected by the outcome variable. These estimates give a composite effect – the effect of doing business with a university on opening an establishment and on the location of that establishment. Because opening an establishment is a rare event, we also provide a separate set of estimates for the sample of multi-establishment firms that opens an establishment in year 2014. While this sample is selected, these estimates isolate location from the opening decision.

Our estimates for establishment exit are in the same spirit. Formally, our model is

$$X_{jut} = \beta_0 + \beta_1 T_{jut-1} + \beta_2 D_{ju} + \beta_3 V_{jt-1} + \gamma_u + \delta_t + \lambda_x + \varepsilon_{jut}. \quad (3)$$

Here X_{jut} is an indicator variable for whether establishment j closes at time t (note that a separate observation is included for each establishment-university pair). As with establishment entry, we lag the transaction variable and vendor characteristics.

As with equation (1), we can get at mechanisms by estimating a range of specifications of equations (2) and (3). The sample can be subset to only High Tech and R&D performing firms to assess whether the relationships vary for firms that are more likely to benefit from knowledge transmission. Additionally, the level of economic interaction can be quantified using both the frequency and the dollar value of transactions. In sum, these different specifications will provide direct evidence of the links between physical distance and economic transactions. The next section focuses on the role of networks on these same transactions.

3.2 Networks and Economic Activity

The literature suggests that ongoing customer relationships, often developed through face-to-face meetings, generate persistent, repeated transactions. Relationships can also propagate within a firm through networks (Gordon and McCann 2000). The models developed above, in which transactions between a vendor and a university are an observed outcome, can be unpacked in more detail. In particular, since our data (described below) include both information about individual projects in each university and the people on each project, it is possible to examine both direct and indirect links between people and purchases.

We develop our models here into two steps. The first step, which we lay out for expositional reasons but estimate as part of a larger model, is to examine the persistence of transactional activity between vendors and individual projects at universities – a direct link. An example might help fix ideas. A project might use a particular vendor for designer mice and be pleased with the experience. That would lead to repeat purchasing – and therefore we would observe persistence in transactions between the project and the vendor. This unpacking could be written in equation (4), as a transaction occurring between vendor *j* and project *i* at university *u* at time *t* as a function of a previous transaction that occurred between the same project and vendor. (Note that in this model, we include *u* to index universities, but each project *i* is university specific.) Formally,

$$T_{jui\,t} = \beta_0 + \beta_1 T_{jui\,t-1} + \epsilon_{juit}. \quad (4)$$

This temporal transactional information network can be further extended to include an indirect link, where people on different projects are connected in some fashion through a network. Because our data include (deidentified) information on each individual researcher and each funded project with which they are involved in each time period, we are able to trace the human links between purchases with a given vendor on one project and purchases with that same vendor on another project. Another example might help fix ideas here. Suppose project *i* needs a particular type of microscope. Further suppose some other project, *k*, at the same university has purchased similar microscopes in the past and the associated researchers were pleased with the experience. If a faculty member is on both projects *i* and *k*, there might well be information transmitted from project *k* to project *i* about the quality of that vendor. More formally, we can extend equation (4) to include $T_{jit-1}^{SD=1}$, a measure that is one for all projects, *i*, that have a network distance of 1 (i.e. 1 degree of separation via the network created by collaborative research projects) from another project that transacted with vendor *j* at time *t-1*. Formally

$$T_{juit} = \beta_0 + \beta_1 T_{juit-1} + \beta_2 T_{juit-1}^{SD=1} + \beta_3 D_{ju} + \beta_4 T_{juit-1}^{oSD=1} \times D_{ju} + \alpha_{ju} + \tau_i + \varepsilon_{jit}. \quad (5)$$

Here we include fixed effects for university-vendor pairs (α_{ju}) and project specific characteristics (τ_i). (Note that the university-vendor pair effects subsume the direct effect of distance.) We also include an interactive term between physical distance and network distance to control for the negative effect of distance on transactions and how it may negate the distance impact for vendors located further away from the university. As indicated earlier, our novel measure of network distance is based on prior business relationships (as opposed to social relationships), which makes it a powerful predictor, although we do not make a claim that it is exogenous.

Measuring such connections requires constructing network units of observation based on both project-vendor relationships and project-researcher relationships. We provide more detail about the structure of the data we have available to us in the next section, but briefly, the project level data are available at the level of funded research projects. The data consist of all purchases by all research projects from all vendors, as well as information about which faculty are involved in each project. As such, it is possible to determine from the data that a purchase from a vendor was made by a funded project, but, since there can be several faculty members on a funded project, any or all of the faculty could have made the purchase decision. Transaction data permit the construction of network measures where the nodes are projects and the edges are defined where the projects share any faculty in common (see Appendix 1 for more details on the construction of these measures).

4. Data

The source of university transactions data is the enhanced STAR METRICS data, or UMETRICS data (Lane et al. 2015). We use UMETRICS data from thirteen major research universities (Michigan State University, Northwestern University, Ohio State University, Pennsylvania State University, Purdue University, and the Universities of Arizona, Illinois, Indiana, Iowa, Kansas, Michigan, Missouri and Wisconsin), provided as a result of a collaboration with the Institute for Research on Innovation and Science. They are not representative of the universe of all research universities, although they do account for more than 20% of federal academic R&D expenditures.

The vendor data are unusually rich. For every purchase on every funded project, the data give the date of each transaction, the amount and identifying information about the vendor (a unique vendor ID, the name, address and, when available, the DUNS Number). Although we do not use these data here, there is also information on the type of good or service purchased. For cross-university consistency, we use data from 2012 to 2014, although the results are consistent in previous years. Over that period, the data cover almost 1.2 million transactions totaling over $1.67 billion on 22,800 funded projects from over 50,500 vendor establishments (an establishment is the physical place where business is conducted, and the unit of observation at which industry and geographic location are defined; firms can own one or more establishments).

The transaction level data are extremely granular: every purchase made on every funded project is included in the data, including very small transactions of just a few dollars. Most transactions are small, and reflect small-scale purchases such as office supplies or trips of short duration. There are also a number of very large transactions from important scientific vendors. Some vendors are responsible for tens of thousands of transactions, but the simple mean of transactions per vendor ranges from 5 to 68, depending on the university. Thus, the distribution of the number of transactions and of transactions by value is right skewed.

We take these data on purchases and match them to data on all US non-agricultural employer establishments with their geographic location (longitude and latitude) from the US Census Bureau's Business Register and Longitudinal Business Database (LBD) (Jarmin and Miranda 2002). This permits us to construct comparison groups of businesses and also directly estimate the likelihood that a given establishment is selected to be a vendor as a function of distance. It is worth noting that distance effects are likely to be nonlinear for two reasons. First, most of the universities in our sample are in the Midwest, so a natural boundary is the two coasts. Second, California, a particularly R&D-intensive state, is between 1,500 and 2,000 miles away from the sample universities.

To construct the US comparison groups, we use two subsamples of the LBD: a subset of firms performing R&D and a subset of High Tech firms.

The R&D sample includes all establishments associated with R&D performing firms. These are identified from surveys of R&D performing firms - which include all firms that report non-zero expenditures in R&D in any given year between 1976 and 2012. The firm identifiers and R&D expenditures are collected from two separate surveys collected over two separate time periods. The R&D data from 1976 to 2007 are collected from the Standard Industrial Research and Development Survey (SIRD) and the R&D data from 2008 to 2012 are collected from the updated version of this survey called the Business Research and Development and Innovation Survey (BRDIS)[9]. Both surveys are jointly administered by the US Census Bureau and the National Science Foundation and represent a national sample of firms beginning in 1992. All firms in the surveys that report conducting R&D in one year are retained to the next year, with

[9] https://bhs.econ.census.gov/bhs/brdis/about.html

additional firms sampled (based on survey weights). There are 12,500 such firms.[10] These firms are known to significantly differ from the typical U.S. firms in a number of dimensions, including being significantly larger and much more likely to engage in international trade and multinational activity (Davis et al. 2007).

Second, we identify establishments in High Tech industries, relying on the classification developed in Goldschlag and Miranda(Goldschlag and Miranda 2016) which is based on the union of industries with the highest proportion of STEM employment in 2005, 2012, and 2014.

5. Basic Facts

Before presenting our results, we begin by showing that vendors have characteristics that make them particularly likely to be engines of growth for the new economy. Specifically, we show that university vendors have more characteristics associated with higher levels of productivity than do non-vendors. Next we show that vendors are more likely to be geographically closer to the university than non-vendors.

Vendors are disproportionately in High Tech industries. Almost 15% of vendor establishments are in Professional and Commercial Equipment and Supplies and Scientific Research and Development Services (NAICS code 4234 and 5417 respectively), while only 0.7% of US establishments are from those industries. Other major industries include Navigational, Measuring, Electrometrical, and Control Instruments Manufacturing (4.8% vs. 0.07% nationally), Pharmaceutical and Medicine Manufacturing (3.4% vs. 0.03% nationally), and Electrical and Electronic Goods Merchant Wholesalers (2.4% vs. 0.38% nationally).

---

[10] Note that the R&D data come from a survey of 45,000 firms that are in industries likely to be doing R&D: there is no census of R&D activity.

Table 1: Industry Concentration of Vendor and All US Establishments

Detailed Industry	LBD Estab. Share	LBD Emp Share	Vendors Estab. Share	Vendors Vendor Dollar Share	Difference in Estab. Share (Vendors-LBD)
Professional and Commercial Equipment and Supplies	0.48	0.54	8.28	11.45	7.80
Scientific research and development services	0.23	0.58	6.54	5.73	6.30
Navigational, measuring, electromedical, and control instruments manufacturing	0.07	0.33	4.81	4.33	4.74
Pharmaceutical and medicine manufacturing	0.03	0.20	3.44	2.49	3.41
Electrical and Electronic Goods Merchant Wholesalers	0.38	0.43	2.37	1.47	1.99

Source: UMETRICS and LBD, author's calculations.
Note: Statistics calculated pooling 2012-2014 for all universities in the sample. Establishment, employment, and vendor dollar shares are calculated by 4-digit 2007 NAICS industries.

Interestingly, the vendors also have characteristics that are associated with greater firm level productivity and with the importance of tacit knowledge. Shown in Table 2, the vendor establishments are larger (>14 times larger) than the typical U.S. establishment, have higher wages, are older, are more likely to be owned by an R&D performing firm, and are more likely to be owned by a firm that patents. One question that arises is whether the vendors maybe spinoffs from the universities. While some are likely to be, the mean age of the vendor establishment is 23.1 years (median 25.5), both of which are considerably older than all U.S. establishments, making it unlikely that many are recent spinoffs.(Klepper and Sleeper 2005)

Table 2: Comparison of Vendor Characteristics with All US Establishments

		Vendors	US	Ratio
Employment	Mean	228	16	14.2
	Median	21.0	4.0	5.3
Payroll per worker (000s)	Mean	69.9	40.3	1.7
	Median	56.1	27.1	2.1
Firm Age	Mean	23.1	17.5	1.3
	Median	25.5	14.5	1.8
Percent within R&D Performing Firm		13.0	5.0	2.60
Percent within Patenting Firm		35.0	17.0	2.06
Percent within 50 Miles of University		12.7	0.7	19.3
Percent within 50-100 Miles of University		4.9	1.0	5.1

Percent within 100-250 Miles of University	13.2	7.7	1.7
Percent within State of University	23.3	2.5	9.4

Source: UMETRICS and LBD, author's calculations.
Note: Statistics calculated pooling 2012-2014 for all vendors in the sample and all establishments in the LBD. Percent within distance bins and within state give the average across all vendors - whether the vendor is within the distance bin or state of the university. In the case of establishments, we calculate whether each establishment is within the distance bin or state of each university and then average across universities. Employment growth is calculated using the standard Davis, Haltiwanger and Schuh (1996) method weighting by the average of employment in t and t-1. Medians calculated as the mean of the 45th and 55th percentiles(Davis, Haltiwanger, and Schuh 1998).

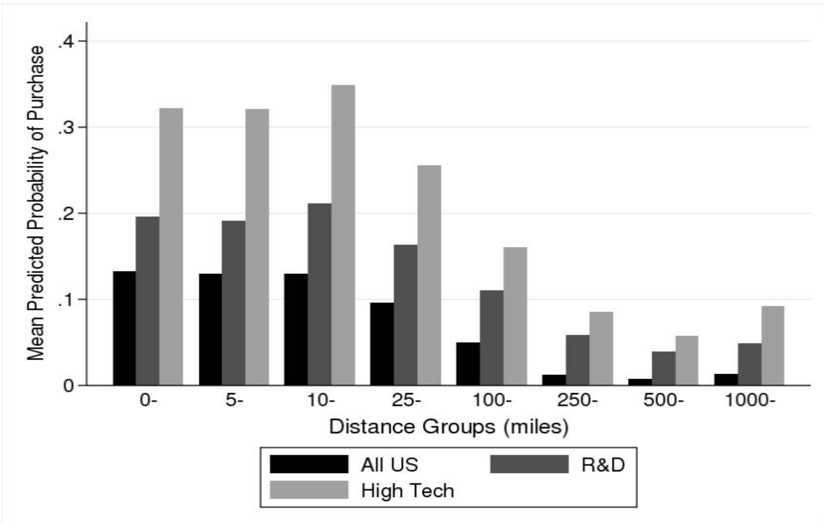
6. Link between Research Expenditures and Regional Activity

Having shown that university vendors differ from the typical US establishment, we now link to Census data and begin characterizing the relationship of physical distance and the probability an establishment supplies inputs for university-based research. Establishments physically closer to each university are disproportionately more likely to be vendors to the university's researchers than are other US establishments. Table 2 shows that vendor establishments are more than 19 times more likely to be within 50 miles of the university than are all US establishments. Vendors are 5 times more likely to be between 50 and 100 miles and 1.7 times more likely to be between 100 and 250 miles than are all US establishments.

Figure 1 shows results from a regression of the probability of a given US establishment being a vendor to a given university as a function of distance, controlling for potential mediating factors (equation (1)). It shows that an establishment within 5 miles of each university has about a .15% chance of being a vendor; this figure drops to about .10% if the establishment is between 25 and 100 miles away. The probabilities are much higher for establishments associated with R&D performing firms and higher still for establishments in High Tech industries. High Tech establishments that are closest have a .32% chance of doing business with the university, and the probabilities, while remaining higher than those for all establishments, still decline with distance.

However, in all cases there is an uptick in probability for vendors 1000-2500 miles away, which is consistent with the distance from our Midwestern universities and the East and West coasts.

Figure 1: Vendor purchase probability and distance from the university.



Source: UMETRICS and LBD, author's calculations.
Note: Mean predicted probability calculated as the estimated probability from a regression of whether an establishment is a vendor in for a given university as a function of distance, distance to the 2nd, 3rd, 4th, 5th, and 6th power[11], 2010 population within the establishment's zip code and population squared, year, and university fixed effects with robust standard errors.

The data are also rich enough to allow us to control for multiple confounding factors—and the results continue to be robust. Because universities vary greatly in size, we include university fixed effects and include fixed effects to control for year-to-year changes in sponsored research funding. We also control for density of people per zip code to capture city agglomeration effects (Chatterji, Glaeser, and Kerr 2013).

Table 3 reports the results of a linear probability model examining the probability of being a vendor based on the distance of establishments from of the university in 100-mile increments, including all controls and with four different samples of establishments. First, we

[11] We experimented with different polynomial functions and settled on using a 6th degree polynomial since the coefficients were still significant and with a nontrivial value (i.e. greater than 1E-10).

regress distance on whether or not a purchase was made for all US establishments. Second, we use the binary outcome measure for the subset of establishments within R&D performing firms. The third sample includes establishments in High Tech industries. Finally, the fourth sample includes only establishments that are vendors to at least one university.

The estimated effects of distance are strongly statistically significant and are robust to the choice of polynomial. The effects of distance are even more pronounced with establishments in both High Tech industries and that are part of R&D performing firms. These results suggest the importance of tacit knowledge for businesses close to the knowledge frontier. Transactions are also more likely to be with establishments that are older and larger, with the exception that, among establishments in R&D performing firms, transactions tend to be with establishments in younger firms.

Among known vendors, for establishments that sell to at least one university in our sample, the estimates (in the last panel) show a much stronger negative relationship between distance and the probability of being a vendor (note that the implied minimum point is quite similar to that for the other models). This is expected given that one reason that the previous coefficients are so small is because the probability that any given establishment is a vendor to a university is very low and all establishments in this sample are vendors to at least one university.

Table 3: Physical Distance Regressions (Linear Probability Model of whether establishment is a vendor)

	All Establishments		Establishment in R&D Performing Firms		Establishments in High Technology Industries		All Establishments that are Vendors to at least 1 University	
Distance (00's miles)	-0.000893***	-0.00302***	-0.00122***	-0.00449***	-0.000104	-0.00378***	-0.150***	-0.763***
	(1.61e-05)	(4.20e-05)	(0.000131)	(0.000298)	(0.000171)	(0.000394)	(0.00677)	(0.0201)
Distance squared (00's miles)		8.21e-05***		0.000128***		0.000143***		0.0277***
		(1.13e-06)		(7.80e-06)		(1.06e-05)		(0.000834)
Log of population in zip code	0.000992***	0.00114***	0.00139	0.00164*	0.00263***	0.00284***	0.0999***	0.0635*
	(8.40e-05)	(8.43e-05)	(0.000854)	(0.000855)	(0.000993)	(0.000994)	(0.0346)	(0.0346)
Firm Size	0.000229***	0.000229***	0.000134***	0.000134***	0.000484***	0.000484***	0.000524***	0.000539***
	(7.92e-06)	(7.92e-06)	(7.76e-06)	(7.76e-06)	(2.41e-05)	(2.41e-05)	(4.92e-05)	(4.91e-05)
Firm Age	0.000193***	0.000176***	-0.00424***	-0.00425***	0.00175***	0.00172***	0.00109	-0.00297
	(7.68e-06)	(7.67e-06)	(0.000210)	(0.000210)	(6.70e-05)	(6.70e-05)	(0.00272)	(0.00272)
Observations	237,585,000	237,585,000	11,104,000	11,104,000	10,850,000	10,850,000	553,500	553,500
R-squared	0.001	0.001	0.004	0.004	0.001	0.001	0.022	0.025

Source: UMETRICS and LBD, author's calculations.

Note: Dependent variable is whether or not a given establishment purchased from a university in a given year (2012-2014). Robust standard errors in parentheses. All regressions include year and university fixed effects and 3-digit NAICS fixed effects. Observations are university-establishment pairs. Observation counts rounded. * denotes significance at the 10% level, ** significance at the 5% level, *** significance at the 1% level.

7. Is Business with a University Related to Local Economic Activity?

If propinquity is indeed important in creating and sustaining successful business relationships, firms should be aware of it and should be more likely to locate near universities. Moreover vendors that do business with and are close to universities should be less likely to close. The data permit an examination of these outcomes because the LBD allows us to track entry and exit among all establishments in the United States. It is thus possible to examine the opening of all new establishments and closing of all existing establishments by all firms, including those that have business relationships with universities. As above we gain some leverage on mechanisms, but looking separately at High Tech establishments and R&D performing firms.

An analysis of the 2014 LBD data supports the notion that firms pay attention to previous business relationships in making location decisions. Opening a new establishment within 50 miles of the universities in the sample is a rare event: only 0.2% of all multi-unit firms in the US did so. However, it is less rare for firms that did business with one of the universities in 2013 – of those, 5.9% opened an establishment within 50 miles of that university in 2014. The rate is even higher among the R&D performing firms that did business with the universities in 2013 - 130 firms of the 1770 such firms (7.2%) opened one or more establishments within 50 miles of the universities. In addition, the greater the value of the business transactions, the higher the probability of opening an establishment. The 130 R&D performing firms who opened an establishment near the university averaged transaction revenues of over $396,780 in 2012 and $488,320 in 2013 with the university near their new establishment. The 1,640 R&D vendor firms that did not open an additional establishment had much lower valued business transactions: the mean value was $51,550 in 2012 and $53,540 in 2013.

Table 4 summaries the order of magnitude of this relationship. It presents a regression analysis of the probability that a firm with multiple establishments in 2013 opens a new establishment within 50 miles of a university in 2014, based both on whether the firm was a vendor in 2013 as well as the amount of the 2013 transaction. Being a vendor in the prior year increases the probability that a firm colocates its new establishment near the university by about 2.3% for the full sample, 1.5% among R&D performing firms, and 1.1% for High Tech firms. A higher value of transactions is associated with a higher probability of collocating with the university, while distance (measured using the average distance of the firm's establishments) to the university is negatively related with colocation.

Table 4. Colocation of New Establishments with University

VARIABLES	All	R&D performing	High Tech
Vendor in 2013	2.280***	1.453**	1.139**
	(0.309)	(0.587)	(0.512)
Payments in 2013 ($1M)	1.529*	1.506*	0.942**
	(0.812)	(0.813)	(0.441)
Average Distance in 2013	-0.000132***	-0.000281***	-2.82e-06
	(8.83e-06)	(0.000100)	(2.38e-05)
Observations	2,122,000	96,000	99,000
R-squared	0.247	0.342	0.370

Source: UMETRICS and LBD, author's calculations.
Note: Dependent variable is whether or not a given multi-unit firm opened a new establishment in 2014 within 50 miles of the given university. Observations are university-firm pairs. Sample includes all 2014 multi establishment firms. All regressions include university and firm fixed effects. Average distance in 2013 captures the average distance between the firm's establishments and the university in 2013. Firm NAICS is calculated using employment shares. Observation counts rounded. * denotes significance at the 10% level, ** significance at the 5% level, *** significance at the 1% level.

Table 5 presents the results of similar regressions for the subset of all firms that opened new establishments near universities in 2014. Although the coefficients are not directly comparable to those in Table 4, they do show that those firms that had a previous transaction with the university in 2013 were substantially more likely to open an establishment than those that did not. The opening of an establishment is more likely to occur the larger the size of the 2013 transaction for R&D performing businesses and for High Tech businesses. Distance is also more important (although not for High Tech businesses).

In sum, these results suggest that firms that supply inputs to a research enterprise are more likely than other firms to colocate near the university and that the decision to colocate is strongly affected by the size of the previous transactional relationship, with distance being a key factor. These results are also robust to focusing exclusively on firms with establishments that were vendors to the universities in 2013.[12]

Table 5. Colocation of New Establishments with University, New Establishment Openers

VARIABLES	All	R&D performing	High Tech
Vendor in 2013	7.429***	3.697**	4.981**
	(1.032)	(1.603)	(2.142)
Payments in 2013 ($1M)	1.446	1.476*	0.894*
	(0.882)	(0.816)	(0.477)
Average Distance in 2013	-0.00268***	-0.00301***	-0.000171
	(0.000152)	(0.000792)	(0.000473)
Observations	~138,000	~15,000	~8,000
R-squared	0.252	0.346	0.369

Source: UMETRICS and LBD, author's calculations.
Note: Dependent variable is whether or not a given multi-unit firm opened a new establishment in 2014 within 50 miles of the given university conditional on the firm opening a new establishment in 2014 in the US. Observations are university-firm pairs. Sample includes all 2014 multi-establishment firms that opened at least 1 establishment in 2014. All regressions include university and firm fixed effects. Average distance in 2013 captures the average distance between the firm's establishments and the university in 2013. Firm NAICS is calculated using employment shares. Observation counts rounded. * denotes significance at the 10% level, ** significance at the 5% level, *** significance at the 1% level.

In addition to being a catalyst for entry, a university-vendor relationship may also prevent exit. Table 6 presents the results of similar regressions focused on the relationship between

[12] Again, by focusing on only vendors we control for the unobserved selection effects that may confound our estimates. See Appendix 3 for details.

vendor transactions and establishment exit (establishment exit is defined by an establishment having positive employment in the prior period and zero employment in the current period (equation (3)). Establishments that had transactions with a university are 3.1% less likely to exit in the following year. The relationship is smaller (1.8%) for establishments in R&D performing firms and higher (6.0%) for High Tech establishments. The pattern of exits is also related to the distance between the establishment and the university: establishments that are farther away from a university have a higher probability of exit in all three samples. Again, these results are generally consistent when focusing exclusively on establishments that were vendors in the prior period, though distance plays less of a role compared to the number of funded projects a vendor supplied in that sample.[13]

Table 6. Establishment Exit

VARIABLES	All	R&D performing	High Tech
Vendor in t-1	-3.134***	-1.848***	-5.979***
	(0.110)	(0.238)	(0.214)
Distance	0.000187***	0.000191***	0.000277***
	(3.50e-06)	(1.43e-05)	(1.81e-05)
Firm Size t-1	6.94e-06***	-6.38e-07***	2.31e-05***
	(3.26e-08)	(1.29e-07)	(2.19e-07)
Firm Age t-1	-0.278***	0.0635***	-0.293***
	(0.000173)	(0.00147)	(0.000914)
Observations	173,330,000	7,916,000	7,834,000
R-squared	0.023	0.049	0.013

Source: UMETRICS and LBD, author's calculations.
Note: Dependent variable is whether or not an establishment exits in a given year (2012-2014), where exit is defined as switching from positive to zero employment. Sample includes all US establishments. All regressions include year and university fixed effects. All and R&D specifications include 3-digit NAICS fixed effects. Observations are university-establishment pairs. Observation counts rounded. * denotes significance at the 10% level, ** significance at the 5% level, *** significance at the 1% level.

[13] See Appendix 3 for details.

Much more analysis can be done, because we also have information about which researchers worked on each research grant. This level of granularity allows us to construct measures of the ties between researchers, which can then be linked to patterns in vendor purchases. We explore these relationships in the following section.

8. The source of regional ties

The literature summarized in Section 2 hypothesized that network connections are an important source of regional ties. In this section, we lay out the results from our analysis investigating how business network ties and previous transaction history determine future transactions. In this section, we abstract from the university and focus primarily on the research funded projects and persons involved in the funded projects (although physical distance is still included as a control). We construct the network distance measures by looking at the transaction history of persons one step away from the research project. As emphasized, these measures are distinct from measures of social ties (Boschma 2005), although we do not claim that they are exogenous.

Examining the Source of Persistence

In 2014 each university had, on average, about 1,100 funded projects and about 8,400 vendors, yielding more than 9 million possible university-funded project-vendor combinations per university. The chance that a given funded project purchases from a particular vendor is slight: 0.06% in 2014. However, as shown in Table 7, there is substantial persistence in purchases. In particular, if funds from a given funded project were used to make a purchase from a given vendor in 2013, the chance of a purchase being made from the same funded project in 2014 increases by approximately 35 percentage points. Prior period purchases are even more strongly related to current purchases for establishments in R&D performing firms and High Tech establishments. The network dimension created by the relationships between faculty and funded projects is also related to the probability of purchasing from a given vendor. If a purchase was made in 2013, the funded project-vendor distance in 2013 is necessarily zero; otherwise, the distance is positive and dependent on the network structure.[14] Given that there was no purchase in 2013, if the shortest network distance was 1 in 2013 (meaning that the distance between a funded research project and vendor is one funded research project away), then the probability of purchase in 2014 is higher by 1.8%. Being one step away from a different funded project that purchases from a vendor in the prior period is even more strongly related to the probability of purchase for R&D and High Tech vendors (both 3.4%).

[14] In focusing on the length of the shortest path, we ignore the number and distance of other paths between the focal grant and other grants that do business with the vendor.

Table 7: Network Distance and the Probability of Purchase in 2014

	All	All	All	R&D Performing	R&D Performing	R&D Performing	High Tech	High Tech	High Tech
Purchase in 2013	35.19***	37.25***	37.25***	41.30***	41.47***	41.46***	38.16***	38.36***	38.36***
	(0.00991)	(0.0187)	(0.0187)	(0.0424)	(0.0428)	(0.0428)	(0.0368)	(0.0372)	(0.0372)
Shortest Path=1 in 2013	1.816***	2.610***	2.246***	3.397***	3.393***	3.019***	3.368***	3.417***	3.326***
	(0.00571)	(0.0116)	(0.0159)	(0.0292)	(0.0296)	(0.0422)	(0.0225)	(0.0229)	(0.0353)
Physical Distance		-2.42e-05	5.71e-06		-1.05e-05	-1.47e-05		-1.90e-05	-2.32e-06
		(5.13e+07)	(5.13e+07)		(2.67e+08)	(2.67e+08)		(8.60e+07)	(8.60e+07)
Shortest Path=1 in 2013 * Physical Distance			0.000465***			0.000422***			9.19e-05***
			(1.39e-05)			(3.39e-05)			(2.73e-05)
Observations	102,785,000	28,090,000	28,090,000	5,246,000	5,130,000	5,130,000	7,143,000	6,996,000	6,996,000
R-squared	0.179	0.211	0.211	0.287	0.289	0.289	0.202	0.204	0.204
Univ FE	No	No	No	No	No	No	No	No	No
Univ-Grnt FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Univ-Ven FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R&D	No	No	No	Yes	Yes	Yes	No	No	No
HT	No	No	No	No	No	No	Yes	Yes	Yes

Source: UMETRICS and LBD, author's calculations.
Note: Dependent variable is whether or not a purchase in made between a funded project and vendor in 2014. The regressions include university-vendor fixed effects and university-funded project characteristics. Shortest path length is calculated as the shortest network distance between each funded project where links between funded projects are defined by shared faculty. Shortest path of 1 in 2013 is a dummy that takes a value of 1 if a funded project distance 1 away from a given funded project in 2013 purchased from a given vendor. Observations are university-vendor(establishment)-funded project triplets. Observation counts rounded. * denotes significance at the 10% level, ** significance at the 5% level, *** significance at the 1% level.

We have hypothesized that physical distance is also an important factor. To address this question, we have generated the physical distance from vendors to each university. Because we are unable to match all universities to business register data on location, we replicate the previous results for the sample of universities with distance data (in columns (2), (5), and (8)). The technical consequence of including fixed effects for individual university-vendor pairs is that it is not possible to estimate the relationship between purchases and distance. However, the interaction between distance and our dummy variable for being a project that is one step away from a vendor transaction is positive and significant. This finding suggests that network distance offsets the negative effect of distance. Intuitively, the probability of a particular project at a university doing business with a distant vendor is lower than for a proximate vendor but it appears that projects can learn about relevant distant vendors through networks. This finding is consistent across the different firm types.

9. Summary

Although descriptive rather than causal, our results directly relate geographic propinquity and regional economic activity at a detailed micro level for a research intensive sector (universities) rather than seeking to indirectly infer the role of propinquity from collocation. We also extend Boschma's measures of propinquity to include measures associated with transactions. The first measure directly captures the persistence of transactions between projects and vendors. The second measure captures indirect relationships tied to transactions with related projects. We thus shed light on the nature of the endogeneity between geography and business activity..

We show that research funds are more likely to be expended at businesses physically closer to universities and that firms that do a lot of business with a university are more likely to set up a new establishment near that university. We show that establishments that are High Tech and R&D performing are more likely to be close to a university than other establishments, consistent with the role of tacit knowledge for such businesses. We also show that these relationships are persistent: if a vendor has been a supplier to a funded project at least once before, that vendor is more likely to be a vendor on the same or related grants. Firms behave in a way that is consistent with the notion that propinquity is good for business; if a firm supplies a funded research project at a university in a given year, it is more likely than other firms to open an establishment near that university in subsequent years. Finally, we show that firms that do business with a university and are close to that university are more likely to survive. These findings are consistent with the notion that research funding stimulates regional economies, particularly in the High Tech and R&D performing sectors.

Of course, much more can be done. A key focus of future research will be to examine the persistence of vendor-faculty relationships over a faculty member's career rather than focusing on individual grants. Our results in this paper do not show causal effects, so it will be necessary to examine the impact of exogenous shocks, such as faculty moving from one university to another, new large scale grants to university-based research centers, or changes in state economic development policies.

We also believe that this paper will provide the stimulus for a new body of research on the long standing interest in understanding the economic and knowledge-transmission role of university research purchases, particularly since more universities are joining the UMETRICS program. These data have been made available for scholarly research purposes (at least in the

U.S.) by the newly-formed Institute for Research on Innovation and Science (iris.isr.umich.edu) and are also available through the Census Bureau's Federal Statistical Research Data Centers (http://www.census.gov/fsrdc).

References

Van Ark, Bart, Janet X Hao, Carol Corrado, and Charles Hulten. 2009. "Measuring Intangible Capital and Its Contribution to Economic Growth in Europe." *EIB papers* 14(1): 62–93.

Boschma, Ron. 2005. "Proximity and Innovation: A Critical Assessment." *Regional studies* 39(1): 61–74.

Chatterji, Aaron, Edward L Glaeser, and William R Kerr. 2013. *Clusters of Entrepreneurship and Innovation*. National Bureau of Economic Research.

Corrado, Carol, Jonathan Haskel, and Cecilia Jona-Lasinio. 2017. "Knowledge Spillovers, ICT and Productivity Growth." *Oxford Bulletin of Economics and Statistics*.

Corrado, Carol, Charles Hulten, and Daniel Sichel. 2009. "Intangible Capital and US Economic Growth." *Review of Income and Wealth* 55(3): 661–85.

Davis, Steven J et al. 2007. *Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Nonemployer Universes*. National Bureau of Economic Research.

Davis, Steven J, John C Haltiwanger, and Scott Schuh. 1998. "Job Creation and Destruction." *MIT Press Books* 1.

Doloreux, David, and Richard Shearmur. 2011. "Collaboration, Information and the Geography of Innovation in Knowledge Intensive Business Services." *Journal of economic geography* 12(1): 79–105.

Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies." *Handbook of regional and urban economics* 4: 2063–2117.

Fellegi, I. P., and A. B. Sunter. 1969. "A Theory for Record Linkage." *Journal of the American Statistical Association* 64: 1183–1210.

Gertler, Meric S. 2003. "Tacit Knowledge and the Economic Geography of Context, or the

Undefinable Tacitness of Being (There).” *Journal of economic geography* 3(1): 75–99.

Glaeser, Edward. 2011. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Penguin.

Glaeser, Edward L, and Giacomo A M Ponzetto. 2010. “Did the Death of Distance Hurt Detroit and Help New York?” In *Agglomeration Economics*, University of Chicago Press, 303–37.

Goldschlag, Nathan, and Javier Miranda. 2016. “Business Dynamics Statistics of High Tech Industries.”

Gordon, Ian R, and Philip McCann. 2000. “Industrial Clusters: Complexes, Agglomeration And/or Social Networks?” *Urban studies* 37(3): 513–32.

Greenia, Nicholas, Kaye Husbands Fealing, and Julia Lane. 2008. *Studying Innovation In Businesses: New Research Possibilities*.

Hausman, Naomi. 2012. *University Innovation, Local Economic Growth, and Entrepreneurship*.

Jarmin, Ron S, and Javier Miranda. 2002. *The Longitudinal Business Database*. Center for Economic Studies, U.S. Census Bureau.

Kantor, Shawn, and Alexander Whalley. 2014. “Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks.” *Review of Economics and Statistics* 96(1): 171–88.

Klepper, Steven, and Sally Sleeper. 2005. “Entry by Spinoffs.” *Management science* 51(8): 1291–1306.

Lane, Julia I., Jason Owen-Smith, Rebecca F. Rosen, and Bruce a. Weinberg. 2015. “New Linked Data on Research Investments: Scientific Workforce, Productivity, and Public Value.” *Research Policy*. http://linkinghub.elsevier.com/retrieve/pii/S0048733315000025.

Liu, Shimeng. 2015. “Spillovers from Universities: Evidence from the Land-Grant Program.”

Journal of Urban Economics 87: 25–41.

McCann, Philip. 2007. "Sketching Out a Model of Innovation, Face-to-Face Interaction and Economic Geography." *Spatial Economic Analysis* 2(2): 117–34.

Moretti, Enrico. 2012. *The New Geography of Jobs*. Houghton Mifflin.

Morgan, Kevin. 2004. "The Exaggerated Death of Geography: Learning, Proximity and Territorial Innovation Systems." *Journal of economic geography* 4(1): 3–21.

Nelson, Richard R. 2005. *Technology, Institutions, and Economic Growth*. Harvard University Press.

Owen-Smith, Jason, and Walter Powell. 2004. "Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Communit." *Organization Science*.

Saha, Subhra, Joseph Staudt, and Bruce Weinberg. 2015. "Estimating the Local Productivity Spillovers from Science." 39347(July): 1–39.

Saxenian, AnnaLee. 1996. *Regional Advantage*. Harvard University Press.

Stephan, Paula. 2014. "The Endless Frontier: Reaping What Bush Sowed?" In *The Changing Frontier: Rethinking Science and Innovation Policy*, University of Chicago Press.

Whittington, Kjersten Bunker, Jason Owen-Smith, and Walter W Powell. 2009. "Networks, Propinquity, and Innovation in Knowledge-Intensive Industries." *Administrative Science Quarterly* 54(1): 90–122.

1. Appendix 1: Network construction

Formally, we define an annual grant-to-grant network, illustrated in Appendix Figure A1, as follows. Each node in the network represents a grant in a given year. Edges are drawn between two nodes if at least one faculty member is paid from both grants in a given year. From this grant-to-grant network we can then calculate the shortest path length between each pair of grants. This measure represents a measure of how distant those grants are within the network dimension of the payroll transaction data. We hypothesize that selling to a grant in one year increases the probability that a vendor will sell to that grant in the next year. Further, we hypothesize that grants will purchase from vendors that are relatively closer to them in this grant-to-grant network because faculty will share information about positive purchasing experiences. Figure A1 illustrates this hypothesis about distance, namely that V3 is more likely to sell to G2 in the next year than to G1 (because V3 is 1 step away from G2, via F2 but 2 steps away from G1). Additionally, V3 is more likely to sell to G2 in the next year than is V4 because V4 and G3 are not connected at all. Figure A1 depicts the relationships knowable from the purchase matches. In this example, it is clear that there has been a purchase from V2 using funds from G2, but it is not known whether the purchase decision was made by F2 or F1.

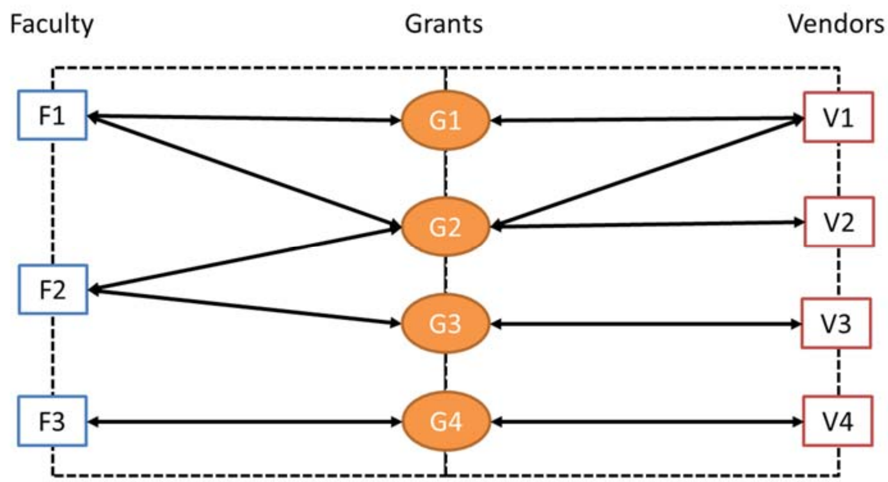


Figure A1: Illustrative mapping of faculty, grants, and vendors

Formally, let $Y_t(G_i, V_j)$ denote whether or not a purchase is made between grant *i* to vendor *j* at time *t*. Further, let $D_t(G_i, G_k)$ denote the shortest path length between grants *i* and *k* at time t. For example, in Figure 3, $D_t(G_1, G_2) = D_t(G_2, G_1) = 1$ and $D_t(G_1, G_3) = 2$. Let $V_t^j(G_i; d)$ denote whether any grants distance *d* from grant *i* purchased from vendor *j*. In this case, $V_t^j(G_i; d) = max_{k|D_t(G_i,G_k)=d} Y_t(G_k, V_j)$.

Appendix 2: Data Construction

The data are derived from the UMETRICS[15] program, an effort parallel to the federally supported STAR METRICS program. The UMETRICS vendor data used in this report are drawn

[15] UMETRICS is a university-led initiative to build a scientific framework that will inform research management, enable evidence-based decision making, and support credible advocacy. Universities participating in the UMETRICS initiative submit quarterly micro-data on university payroll, vendor, subaward/subcontract, and overhead expenditures from federal and non-federal grants and contracts. The data submitted are transactional in format and aggregated for all analyses.

from the financial transactions associated with federal research grants awarded to researchers at 14 universities. The data do not cover the universe of all research grants

The vendor data are matched to Census Business Register (BR) and Longitudinal Business Database (LBD) data. The BR consists of the universe of U.S. non-agricultural firms and their associated establishments and is the ultimate source of all other Census economic data. [16] The LBD is the longitudinally linked employer version of the BR, providing a database that allows us to track firm performance, births and deaths over time. It combines administrative records and survey-based data for all nonfarm employer business units in the United States and hence provides information about the dynamics of firm growth. Key data elements include industry classification, geographic data, employment measures, payroll, and firm age.[17]

As a benchmark, we use a subsample of the LBD that includes all establishments associated with R&D performing firms. R&D performing firms include all firms that report non-zero expenditures in R&D in any given year between 1976 and 2012. The firm identifiers and R&D expenditures are collected from two separate surveys collected over two separate time periods. The R&D data from 1976 to 2007 are collected from the Standard Industrial Research and Development Survey (SIRD), and the R&D data from 2008 until 2012 are collected from the updated version of this survey called the Business Research and Development and Innovation Survey (BRDIS)[18]. Both surveys are jointly administered by the US Census Bureau and the

---

[16] The key source data elements in the Business Register are (i) the SS-4, by which a new business tells the IRS whether it is beginning as a sole proprietorship, partnership, corporation, or personal service corporation; the state or foreign country in which it is incorporated; and whether it is applying because it is a new entity, has hired employees, has purchased a going business, or has changed type of organization (specifying the type) and (ii) the 1120S K-1 series, which provides information on corporate shareholders (Greenia, Husbands Fealing, and Lane 2008).

[17] Non-employer businesses, which constitute the majority of businesses in the United States (although only 4% of sales and receipts), have no paid employees. Our ability to track businesses from the non-employer to the employer stage allows us to identify startups that may not succeed as well as their transition path.

[18] https://bhs.econ.census.gov/bhs/brdis/about.html

National Science Foundation and represent a national sample of firms beginning in 1992. All firms that report conducting R&D in one year are retained to the next year, with additional firms sampled (based on survey weights). R&D performing firms make up a small share of all firms in the United States. Of the 5M+ firms in existence in the United States in 2012, fewer than 12,500 report conducting R&D (unweighted). These firms are also known to significantly differ from typical U.S. firms in a number of dimensions, including being much larger and much more likely to engage in international trade and multinational activity (Davis et al. 2007).

To match the university vendor and the Business Register (BR) data, we use fuzzy name and location matching and rule-based block matching techniques (Fellegi and Sunter 1969). Once BR matches are identified, we create a longitudinal university-vendor-year panel using the LBD. The LBD embodies a number of cleaning algorithms, including the retiming of births and deaths around Economic Census years (Jarmin and Miranda 2002).

The university vendor data are annualized by calculating the sum and the mean of the vendor payment amounts for each vendor for each transaction year. These data are merged with the LDB to add latitude and longitude fields for distance calculations between vendors and their associated universities. The LBD also provides firm age and cleaned versions of BR attributes.

Appendix 3: Colocation and Transaction detail

Table A3.1 presents a regression analysis of the probability that a firm with multiple establishments in 2013 opens a new establishment within 50 miles of a university in 2014 conditional on being a vendor in 2013 to at least one university. By focusing only on firms with vendor establishments in 2013 we are better able to control for the unobserved selection effects that may otherwise contaminate our estimates. Here we find results consistent with those for the full sample.

Table A3.1. Colocation of New Establishments with University, 2013 Vendors

VARIABLES	All	R&D performing	High Tech
Payments in 2013 ($1M)	2.657***	2.488***	1.577***
	(0.599)	(0.665)	(0.231)
Average Distance in 2013	-0.000907*	-0.00278**	-0.000399
	(0.000490)	(0.00136)	(0.000541)
Firm Size in 2013	0.000197***	0.000251***	0.000344***
	(2.12e-05)	(3.96e-05)	(3.24e-05)
Firm Age in 2013	0.0696**	0.0453	-0.0435
	(0.0291)	(0.0527)	(0.0327)
Grant Count in 2013	0.00570	0.0204	0.00809
	(0.0134)	(0.0153)	(0.0377)
Grant Employment in 2013	2.09e-05	-7.65e-05	2.22e-06
	(8.75e-05)	(0.000101)	(0.000258)
Observations	4,700	1,800	1,100
R-squared	0.230	0.393	0.430

Source: UMETRICS and LBD, author's calculations.
Note: Dependent variable is whether or not a given multi-unit firm opened a new establishment in 2014 within 50 miles of the given university. Observations are university-firm pairs. Sample includes all 2014 multi establishment firms that had at least one vendor establishment in 2013. All regressions include university fixed effects. All and R&D specifications include 3-digit NAICS fixed effects. Average distance in 2013 captures the average distance between the firm's establishments and the university in 2013. Grant count is the number of unique grants the firm supplied and grant employment is the number of unique grants the firm supplied weighted by the number of individuals working on each grant. Firm NAICS is calculated using employment shares. Observation counts rounded. * denotes significance at the 10% level, ** significance at the 5% level, *** significance at the 1% level.

Table A3.2 presents a regression analysis of the probability that an establishment exits in a given year for a sample of establishments that were vendors in t-1. Again, by focusing only on establishments that were vendors in the prior year we are better able to control for the unobserved selection effects that may otherwise contaminate our estimates. Here we find results broadly consistent with samples that include establishments that were not vendors in the prior year. However, distance plays less of a role compared to the number of grants in the prior period. Vendors with more grants in t-1 are less likely to exit in t.

Table A3.2. Establishment Exit, *t-1* Vendors

VARIABLES	All	R&D performing	High Tech
Payments in t-1 ($1M)	3.49e-08	-2.48e-08	-3.12e-08
	(1.98e-07)	(1.50e-07)	(2.32e-07)
Average Distance in t-1	0.000257*	-0.000305	-0.000109
	(0.000134)	(0.000304)	(0.000285)
Firm Size in t-1	2.79e-05***	2.42e-05**	4.34e-05***
	(6.00e-06)	(1.14e-05)	(1.22e-05)
Firm Age in t-1	-0.116***	-0.0274	-0.102***
	(0.0105)	(0.0327)	(0.0195)
Grant Count in t-1	-0.00787**	-0.00950**	-0.00201
	(0.00356)	(0.00420)	(0.0177)
Grant Employment in t-1	-1.17e-05	1.06e-05	-3.55e-05
	(2.23e-05)	(2.47e-05)	(0.000133)
Observations	30,000	5,000	7,000
R-squared	0.016	0.048	0.011

Source: UMETRICS and LBD, author's calculations.
Note: Dependent variable is whether or not an establishment exits in a given year (2012-2014), where exit is defined as switching from positive to zero employment. Sample includes all US establishments that were vendors to at least one university in t-1. All regressions include year and university fixed effects. All and R&D specifications include 3-digit NAICS fixed effects. Observations are university-establishment pairs. Observation counts rounded. * denotes significance at the 10% level, ** significance at the 5% level, *** significance at the 1% level.