Understanding the Inputs into Innovation: Do Cities Substitute for Internal Firm Resources?

Chris Forman

College of Management
Georgia Institute of Technology
Atlanta GA 30308
chris.forman@mgt.gatech.edu

Avi Goldfarb

Rotman School of Management
University of Toronto
Toronto ON M5S 3E6, Canada
agoldfarb@rotman.utoronto.ca

Shane Greenstein

Kellogg School of Management
Northwestern University
Evanston IL 60208
greenstein@kellogg.northwestern.edu

We examine whether there is a trade-off between employing internal (firm) resources and purchased external (local) resources in process innovation. We draw on a rich dataset of Internet investments by 86,879 US establishments to examine decisions to invest in advanced Internet technology. We show that the marginal contribution of internal resources is greater outside of a major urban area than inside one. Agglomeration is less important for firms with highly capable IT workers. When firms invest in innovative processes they act as if resources available in cities are partial substitutes for both establishment-level and firm-level internal resources.

1. INTRODUCTION

Investment in innovation does not happen by accident. Firms choose to develop resources and processes to facilitate innovation (e.g., Cohen and Levinthal, 1990; Åstebro, 2002). Geographic location also plays an important role (e.g., Griliches, 1957). Prior work has demonstrated evidence of localization in innovation (Jaffe et al., 1993), suggesting that a propitious location may lower the costs of innovative output (Furman et al., 2005). It is widely assumed that such concerns have motivated firms in information technology (IT) hardware, software, and pharmaceuticals to cluster together (e.g., Saxenian, 1996; Bresnahan and Gambardella, 2004).

There is less understanding of the trade-off between location and internal resources. In particular, there is little empirical evidence on the extent of localization of substitution between internal (firm) resources and purchased external (local) inputs into innovation. This is a surprising gap in evidence, because it is a foundation for understanding firm heterogeneity in valuing agglomeration. If firms can innovate in their operations by substituting away from internal resources when they locate in cities, where purchased services are more readily available, then agglomeration will be most important for smaller firms with fewer internal resources.[1] Alternatively, if firms with rich internal resources use them to invest in process innovations, then locating in lower-density areas may have little effect on their ability to innovate.

We examine this trade-off in the context of information technology investments. In particular, we focus on a set of frontier Internet technologies that facilitate communication within the establishment, which we label Within-Establishment Internet (WEI). We analyze a survey (conducted by Harte Hanks through December 2000) of use of WEI and other information technology at 86,879 establishments that had over 100 employees.[2] The sample consists of established firms rather than start-ups, which allows us to treat establishment location as determined prior to the decision to invest in Internet technologies. Most of the organizations in the sample have *some* experience with basic information technologies, such as personal computers (PCs), but they differ tremendously in their capacity to manage large IT projects. Only a fraction of these establishments have extensive experience with advanced IT projects. The data contain detailed information about

---

1. Vernon (1963) first articulated this idea in a case study of the New York City area. He argued that agglomeration economies are especially important for small (start-up) firms that lack scale economies in their own organizations.

2. We use the terms *firm* and *organization* interchangeably. *Multi-establishment firms* can have establishments in more than one location, more than one establishment in a single location, or both.

establishment-level IT personnel (programmers). Because 45,948 establishments come from 7,035 different multi-establishment organizations, they also vary in their potential to move assets between establishments. Furthermore, establishments come from all over the United States, both major urban areas and isolated rural locations, so they vary in their potential to hire from local labor and service markets.

We first show that establishments that are part of firms with more computer programmers invest in an Internet-based process innovation more frequently. Establishments in large cities also invest more frequently. Our main result is that *establishments act as if these inputs into innovation are partial substitutes for each other*. Being in an urban location is most closely associated with innovation for firms with few programmers. When firms have many programmers, even at other establishments, the relationship between urban location and investment is relatively weak. We infer that the marginal contribution of internal resources to the probability of innovation differs across locations. We also infer that firms that can draw on rich internal resources can innovate outside of cities. We find no evidence that internal and external resources are complements as inputs into process innovation.

Our paper advances the existing literature on the role of resources in innovation. First, we provide empirical evidence on the determinants of adopting a process innovation related directly to operations. Almost all prior research about the inputs into innovation has instead employed different measures of innovation inputs or outputs, such as patent citations or patent output (e.g., Singh, 2004). Second, we test our hypotheses on a large cross section of industries and locations. Prior empirical studies of the role of resources in process innovation focus on case studies of a narrower set of industries and locations (e.g., Kelley and Helper, 1999; Henderson, 2003).

The emphasis in our research and our conclusion contrasts with much of the existing literature on the role of internal expertise in research and development activities (e.g., Arora and Gambardella, 1994; Cassiman and Veugelers, 2006).[3] This literature argues that internal expertise complements the external stock of knowledge by facilitating knowledge flows. In our context, we examine a process innovation in regular operations. Implementing this type of innovation requires investing in both labor time and experience, and we expect firms to treat these as substitutable inputs on the margin. That is, external (local) expertise

---

3. Arora and Gambardella (1994) examine whether internal and external knowledge are complements in biotechnology innovation. Cassiman and Veugelers (2006) find that internal R&D and external knowledge acquisition are complements in a sample of 269 Belgian manufacturing firms. In both of these cases, external knowledge is not defined relative to location.

substitutes for internal expertise during innovative investments. We find evidence consistent with this premise.

Although we do not examine the decision of firms to relocate in response to the diffusion of new technology, we do draw conclusions that contrast with prior research on localization of innovation (e.g., Porter, 1998). We conclude that highly capable firms, that is, firms who employ many skilled IT workers, do not necessarily need to relocate to agglomerated areas or clusters to successfully innovate. Capable firms may be located in more isolated regions and still, nonetheless, innovate at comparatively low cost. Firms without such skilled IT workers, however, will face different concerns when they innovate, and, for purposes of innovating, may benefit from relocating near other firms that also innovate.

We next develop the conceptual framework and hypotheses. Sections 3 and 4 present the data and empirical framework. Section 5 presents our results. In the conclusion, we develop implications for the literature on the geography of innovation and the literature on outsourcing.

2. Conceptual Framework and Hypotheses

We examine how internal and external factors affect establishment-level decisions to adopt WEI, a process innovation. WEI is a group of technologies that reduce the cost of communications internal to the establishment but not outside the establishment. Although our empirical work focuses on this technology only, the motivation for our hypotheses is more general: in what ways might firms draw on internal and external resources when they innovate?

We divide the types of resources available into two broad types: (1) "external" resources available locally outside the firm and (2) "internal" resources available within the firm. Resources within the firm are further divided into resources available within the establishment and resources available outside the establishment within the same firm. We measure external resources by the local population.[4] Internal resources are measured by the number of software programmers. From this point forward, we will also refer to the internal resources as "programmers."

Our main questions are (1) What is the *marginal* contribution of each of these resources? and (2) How do these resources interact to affect these marginal contributions? To examine these, we develop and test

---

4. We focus on population because the external resources used for IT projects may involve much more than simply IT workers, for example knowledge spillovers and physical infrastructure. We show results are robust to using the number of local IT workers instead.

hypotheses related to being in a city, employing many programmers in an establishment or firm, and the interactions between these measures.

> *Hypothesis 1 (from Forman et al., 2005a): Investment in WEI will be increasing as location size increases.*

Large cities may have thicker labor markets for complementary services or for specialized skills such as programming in new computer languages. Thicker markets lower the (quality-adjusted) price of obtaining IT services such as contract programming and of hiring workers to perform development activities in-house. Increases in location size also may increase the presence of non-market-mediated knowledge spillovers that reduce adoption costs (e.g., Goolsbee and Klenow 2002). Such locations may also have greater availability of complementary information technology infrastructure, such as broadband services. Increases in each of these factors may decrease the costs of adopting complex Internet technologies in cities, other things being equal.[5]

> *Hypothesis 2a:* *Establishments with more programmers, ceteris paribus, will be more likely to invest in WEI than establishments with fewer programmers.*
>
> *Hypothesis 2b:* *Establishments in firms with more programmers in their other establishments, ceteris paribus, will be more likely to invest in WEI than establishments in firms with fewer programmers.*

Having more resources elsewhere in the firm means that lower cost resources can be tapped or loaned between divisions of the same firm. Also, the internal firm resources that arise as a result of prior investments in other IT projects may lower adoption costs. The number of programmers measures a mobile direct input into the investment: programmer time. That is, resources and other investments in the organization are already employed in some IT task, and the new technical opportunity leads them to be redeployed for use in advanced Internet applications. The number of programmers also provides a proxy for the experience of the firm or establishment with IT projects. Prior IT projects may reduce

---

5. Note minor similarities and differences with prior work. Forman et al. (2005a) inferred that the geographic variation in WEI investment was consistent with the "urban leadership hypothesis." However, that prior inference did not control for internal capabilities, as we do here. This is consistent with prior theory work arguing that firms locate administrative and support functions strategically. Duranton and Puga (2002) argue that a firm may find it advantageous to locate administrative and support services in large areas because of better availability and a larger variety of complementary services.

development costs if programmers are able to transfer lessons learned from one project to another.[6]

> *Hypothesis 3a:* *The sensitivity of WEI investment to increases in location size will be declining as the number of programmers in the establishment increases.*
>
> *Hypothesis 3b:* *The sensitivity of WEI investment to increases in location size will be declining as the number of programmers found in other establishments within the same firm increases.*

Hypothesis 3 examines whether internal resources are less crucial in cities, that is, whether the marginal contribution of (internal) skilled IT employees is higher outside cities than inside.[7] If an establishment has better internal resources, there are fewer advantages of external resources. Similarly, if an establishment has better external resources, the advantages of internal resources are reduced. Our evidence supporting Hypothesis 3, that is, internal resources are less crucial in cities, constitutes the main contribution of this paper.

   A comparison with existing work that examines the relationship between internal and external inputs to innovation is informative. Much prior work has emphasized complementarities between internal and external expertise. Both Cohen and Levinthal (1989) and Arora and Gambardella (1994) argue that internal R&D helps firms understand and use externally available information. In other words, "a basic research capability is often indispensable in order to monitor and evaluate research being conducted elsewhere" (Rosenberg, 1990). Cassiman and Veugelers (2006) provide empirical evidence of this phenomenon by analyzing the effect of the make and buy decisions for R&D sourcing on the percentage sales of new products for 269 firms who do at least some R&D.

   In contrast to the focus on R&D activities, we examine a setting in which firms evaluate the use of internal and external resources to aid in implementing an IT-based process innovation. Although internal firm resources such as programmers may help identify and evaluate the performance of external IT resources, our setting is particularly

   6. For example, software developers may be able to share common tools for design, development, and testing (Banker and Slaughter, 1997), or they may be able to reuse code (Barnes and Bollinger, 1991). Software development may also have learning economies (Attewell, 1992) that through experience reduce the unit costs of new IT projects. Much prior research in the costs of innovative activity has also long presumed that experience with prior related projects can lower the costs of innovation (Cohen and Levinthal, 1990).
   7. As mentioned earlier, Vernon (1963) first articulated this idea in a case study of the New York City area. He argued that agglomeration economies are especially important for small (start-up) firms that lack scale economies in their own organizations. Agglomeration economies have also been cited as a reason for the success of nascent software firms in countries such India, Ireland, and Israel (Arora et al., 2004).

conducive to measuring substitution effects, if there are any, because we will observe whether firms undertake projects related to implementing a process innovation or not. As noted above, prior work on other IT projects may create learning economies and spillovers that decrease the costs of adapting general purpose IT to organizational needs, reducing the importance of external consultants and local spillovers. Moreover, when IT labor forces are mobile, shared human capital at other establishments decreases the value of consultants and thicker labor markets in large cities.

Our setting also differs from prior work in that we focus on external resources that are geographically local. Development of new IT systems involves extensive coordination to learn user needs for new IT systems, and often involves difficult and time-consuming changes to internal business processes. Both require a local presence from external resources. Thus, whereas the R&D literature focused on the role of internal resources on facilitating communication with external resources, in our setting the communication demands are less complicated. Instead, we examine inputs into innovation in regular operations where both internal and external inputs require labor time onsite.

In summary, although the existing literature emphasizes the possibility that local resources complement internal resources by allowing establishments to better apply local general expertise to firm-specific problems, we believe it will not be the dominant effect in our context. Yet, because it is an open question in the literature, we allow for both substitution and complementarity in our empirical work and allow the data to assess the validity of our hypotheses.

3. Data

The data used in this study come from the Harte Hanks Market Intelligence Computer Intelligence Technology database (hereafter CI database).[8] This database contains establishment- and firm-level data on characteristics such as number of employees, number of programmers, and use of Internet applications. Harte Hanks collects this information to resell as a tool for the marketing divisions of technology companies. Interview teams survey establishments throughout the calendar year; our sample contains the most current information as of December 2000.

Harte Hanks tracks over 300,000 establishments in the United States. Because we focus on commercial Internet use, we exclude government, military, and nonprofit establishments (mostly in higher education). Our sample from the CI database contains all commercial

---

8. This section provides an overview. For more detail, see Forman et al. (2002).

establishments with over 100 employees—in total 115,671 establishments. Harte Hanks provides one observation per establishment. We use the 86,879 clean observations with complete data generated between June 1998 and December 2000.[9] Harte Hanks also tracks whether an establishment is affiliated with a larger organization. In total, there are 47,966 distinct organizations, and 7,035 of these have more than one establishment.

Our dependent variable is a measure of investment in advanced Internet technology that either changes existing internal operations or implements new services involving communication within the establishment. We label this investment Within-Establishment Internet, or WEI. We look for indications that an establishment has made investments that involved frontier technologies or substantial co-invention. The threshold for defining *substantial* is necessarily arbitrary within a range.[10] It usually arises as part of other intermediate goods, such as software, computing, or networking equipment. Investment in WEI involves the use of Internet protocols in the input and output of data to and from business applications software. Examples include (1) intranet applications that enable Web access to information stored in business applications software, such as inventory or accounting data and (2) applications that have functionality involving integration with back-end databases (e.g., Web access to a data warehouse).[11]

Our measure of location size is a dummy variable that equals one when the establishment is located in a metropolitan statistical area (MSA) with a population over 500,000, which we term *sizeable MSA*. This is the simplest way to represent differences between large cities

9. Our regressions control for the time of survey. Consequently we dropped establishments that did not indicate when they were surveyed. We also dropped establishments that were not surveyed on information technology. There is a small bias in the dropped observations toward locations where information technology investment is high. We weight observations to match Census County Business Patterns data by NAICS and location to control for any location and industry bias in the sample.

10. We tested a number of slightly different definitions and did not find any significant changes to our findings. Results are also robust to other measures of advanced technology use including whether the establishment uses an Internet programming language or has installed a PC server.

11. To be specific, an establishment is counted as investing in WEI if it invests in one of the following applications that utilize Internet protocols in the input or output of data: (1) business application software that involves intensive use of database management systems, such as accounting, sales and marketing, payroll, ERP and MRP, inventory, order processing, and data warehousing; (2) science and research applications used for financial analysis and modeling, CAD/CAM/CAE, data analysis, and engineering; or (3) office applications, such as personnel management, project management, and groupware. See Forman, Goldfarb, and Greenstein (2005a) for more details.

and small cities and rural locations, and keeps the results stark and easy to interpret.[12]

We use the number of programmers in the establishment or organization to measure the internal resources that can be deployed to build new Internet applications. For the establishment measure, the data contain information on the number of programmers at each establishment.[13] For the organization measure, we examine only the multi-establishment firms within our sample. We compute the total number of programmers from other establishments within the same firm.[14]

In Table I, we provide descriptive statistics. Although there are differences between sizeable MSAs and other areas in the number of programmers per establishment, the differences are small enough not to affect our interpretation of the marginal effects across major cities and other areas. Our regressions also include controls for establishment employment, controls for whether the establishment is part of a multi-establishment firm, three-digit NAICS dummies, and dummies for the month the survey was conducted.[15] These variables control for many other unmeasured determinants of demand and supply.

4. EMPIRICAL STRATEGY

We estimate a probit model of investment in the process innovation, WEI. Our latent endogenous variable is the net value to establishment *i* of investing in co-invention activity related to use of WEI. We observe only discrete choices, namely, whether or not the establishment chooses to invest. The observed decision takes on a value of one or zero. We specify the net benefit function as a linear function of all its parameters

---

12. Results are robust to other definitions of what defines a city as well as continuous measures of city size and the number of IT workers in the city.

13. In our database, the *programmers* variable is constructed using the following cells: 1–4, 5–9, 10–24, 25–49, 50–99, 100–249, 250–499, and 500 or more. To convert this measure into a continuous variable, we take the midpoint of each interval and use 500 as the value for the right-censored observations. In our sample, less than 1% (85) of the establishments have a right-censored value for programmers. Qualitative results do not change if a dummy for 500 or more is included.

14. These measures quantify the total number of programmers instead of the total quality or cost. If urban areas have thicker labor markets for higher-quality programmers at the same wage rates as rural areas then that would bias our estimates away from hypothesis 3a and 3b. If urban areas have higher wage rates for higher- or same-quality programmers, then the bias in our estimate is ambiguous.

15. Establishments were interviewed over a 2-year period. Those interviewed toward the end of the period are more likely to have invested. Therefore, we control for the month surveyed.

TABLE I.
DESCRIPTIVE STATISTICS

	M	SD	Minimum	Maximum	Number of observations
Full Dataset					
Ln(Programmers in organization+1)	1.7382	2.2898	0	8.5535	86,879
Ln(Programmers in establishment +1)	0.5100	1.0189	0	6.2166	86,879
MSA population over 500,000 dummy	0.7371	0.4402	0	1	86,879
Multi-establishment firm dummy	0.4479	0.4973	0	1	86,879
Ln(establishment employment)	5.3376	0.7248	4.605	10.933	86,879
WEI (the dependent variable)	0.1192	0.3240	0	1	86,879
Ln(Local IT workers+1)	7.4480	3.8118	0	11.4904	86,879
MSA population over 500,000					
Ln(programmers in organization+1)	1.7865	2.3393	0	8.5535	64,038
Ln(programmers in establishment +1)	0.5606	1.0899	0	6.2166	64,038
Other areas					
Ln(programmers in organization+1)	1.6029	2.1393	0	8.5535	22,841
Ln(programmers in establishment +1)	0.3682	0.7691	0	6.2166	22,841

plus interactions for the number of programmers and whether the establishment is in a city.[16] Our base specification for the net value from investing in WEI technology is

$$Y_i = \alpha_0 + \alpha CITY_i + \beta OrganizationProgrammers_i$$

$$+ \gamma EstablishmentProgrammers_i$$

$$+ \delta OrganizationProgrammers_i CITY_i$$

$$+ \phi EstablishmentProgrammers_i CITY_i$$

$$+ \theta_1 MultiEstablishment_i + \theta_2 Establishment\ Employees_i + \theta_3 x_i + u_i,$$

where Y_i is the latent value to establishment i of investing in WEI and x_i is a vector of controls including three-digit NAICS industries and time of survey. The signs of the coefficients in a probit do not necessarily indicate the sign of the marginal effect (Ai and Norton, 2003). Thus, to identify each of our hypotheses, we focus directly on the signs and magnitudes of the marginal effects, calculated at mean values and using the proper formulas as in Ai and Norton's study. We verified through visual inspection that these marginal effects did not qualitatively vary much across the relevant range of values leading to positive investment. In all cases, the signs of the coefficients hold for the marginal effects.

Our base specification treats our variables measuring external and internal resources (i.e., city and number of programmers, respectively) as statistically exogenous, and then we will later test the sensitivity of results to modeling for endogeneity. We assume that u is distributed as a normal i.i.d. variable. We weight models by the actual geographic distribution of establishments for industry and size, according to Census County Business Patterns data. If our data undersample a given two-digit NAICS at a location relative to the Census, then each observation in that NAICS-location is given more importance (for details, see Forman et al., 2002).

5. Results

In this section, we first show the impact of cross-sectional changes in location size and the number of programmers on WEI investment. We

---

16. We also experimented with adding quadratic and other higher-order terms for establishment and organizational capabilities. The marginal effects for a quadratic specification are shown below as a robustness check.

then examine the interaction between the roles of internal IT workers and of cities. Table II presents the main results. Later, Table III examines the robustness of the results.

Column (1) of Table II shows evidence in favor of Hypotheses 1 and 2, without considering interaction effects. In particular, an increase in location size has a significantly positive effect on investment in WEI technology (supporting Hypothesis 1). Furthermore establishments with more programmers are more likely in invest (supporting Hypothesis 2a) and establishments in firms with more programmers are more likely to invest (supporting Hypothesis 2b).

Columns (2) and (3) present the coefficient estimates and marginal effects of the main results including interactions between internal IT workers and being in a sizable MSA.[17] The marginal impact of Hypothesis 1, being in a sizeable MSA, increases the likelihood of investing in WEI by 3.26 percentage points at mean values. This is a large amount, given that the percentage of firms investing is just 11.92%. The marginal impact of Hypothesis 2, having more programmers in your establishment or organization, is also substantial. To demonstrate the change in the likelihood of investment associated with a change in IT programmers, we compute the marginal effect of a one standard deviation change. These marginal effects are computed by multiplying the probability derivative, defined in footnote 18, with a one standard deviation change (defined from column 2 of Table I). An increase in the log number of programmers at the establishment by one standard deviation will increase the average probability of adoption by 7.99 percentage points. Increasing the log number of programmers by other establishments in the same firm has a similar, but smaller, impact. One standard deviation increase in (logged) organization-level

17. We include formulas for these marginal effects below. For brevity, we use *"EP"* and *"OP"* as shorthand for *EstablishmentProgrammers* and *OrganizationProgrammers*. The formulas are:

$$\frac{\partial Y_i}{\partial City_i} = (\alpha + \delta OP_i + \phi EP_i)\Phi'(.), \quad \frac{\partial Y_i}{\partial OP_i} = (\beta + \delta City_i)\Phi'(.), \quad \frac{\partial Y_i}{\partial EP_i} = (\gamma + \phi City_i)\Phi'(.),$$

$$\frac{\partial^2 Y_i}{\partial City_i \partial OP_i} = \phi\Phi'(.) + (\alpha + \delta OP_i + \phi EP_i)(\beta + \delta City_i)\Phi''(.),$$

and

$$\frac{\partial^2 Y_i}{\partial City_i \partial EP_i} = \delta\Phi'(.) + (\alpha + \delta OP_i + \phi EP_i)(\gamma + \phi City_i)\Phi''(.).$$

It is not straightforward to interpret the standard errors on these marginal effects. They do not combine to give an overall test statistic. The test statistics need to be calculated observation-by-observation. We present the marginal effects and standard errors at mean values for the sample.

TABLE II.
MAIN RESULTS

	(1) Direct Effect Only Coefficient	Includes Interactions of Programmers with "City"		Includes Interaction of Organization and Establishment Programmers	
		(2) Coefficient	(3) Marginal Effect	(4) Coefficient	(5) Marginal Effect
Ln(programmers in the organization)	0.0152 (0.0057)**	0.0399 (0.0082)**	0.0053 (0.0016)**	0.0624 (0.0068)**	0.0109 (0.0012)**
Ln(programmers in the establishment)	0.2670 (0.0085)**	0.3395 (0.0193)**	0.0784 (0.0024)**	0.3764 (0.0141)**	0.0814 (0.0017)**
City dummy (MSA Population>500,000)	0.0995 (0.0180)**	0.2070 (0.0257)**	0.0326 (0.0051)**	0.1451 (0.0224)**	0.0238 (0.0045)**
Ln(programmers in the organization) * City		-0.0285 (0.0078)**	**-0.0075 (0.0022)**	-0.0162 (0.0064)*	-0.0038 (0.0018)*
Ln(programmers in the establishment) * City		-0.0804 (0.0206)**	**-0.0148 (0.0057)**	-0.0639 (0.0143)**	-0.0123 (0.0039)**
Ln(programmers in the organization) * Ln(programmers in the establishment)				-0.0231 (0.0018)**	**-0.0039 (0.0005)**
Ln(establishment employment)	0.2318 (0.0132)**	0.2299 (0.0132)**	0.0644 (0.0037)**	0.2518 (0.0089)**	0.0709 (0.0025)**
Multi-establishment firm dummy	0.1126 (0.0274)**	0.1100 (0.0273)**	0.0308 (0.0077)**	0.0686 (0.0199)**	0.0193 (0.0056)**
Observations	86,872	86,871		86,872	
LL	−24550.40	−24,528.56		−24,423.46	

Notes: Standard errors are in parentheses. For the marginal effects, these are calculated at mean values. All regressions are weighted to reflect the actual geographic distribution of establishments from County Business Patterns and include dummy variables for three-digit NAICS and month of survey. Significance levels do not change if standard errors are clustered by firm. Key results in **bold.**
+Indicates significance at 90% confidence level.
*Indicates significance at 95% confidence level.
**Indicates significance at 99% confidence level.

TABLE III.

ROBUSTNESS AND FURTHER ANALYSIS (MARGINAL EFFECTS ONLY)

	(1) Establishment-Level Decisions Only	(2) Instruments 1[a]	(3) Instruments 2[a]	(4) Local Resources Measured by # IT Workers	(5) Full Translog-Type Specification[b]	(6) Manufacturing Only[c]	(7) Services Only[c]
Ln(programmers in organization)	0.0040 (0.0043)	0.0096 (0.0045)*	0.0045 (0.0025)+	0.0056** (0.0016)	0.0155** (0.0027)	0.0159 (0.0023)**	0.0026 (0.0020)
Ln(programmers in establishment)	0.0647 (0.0074)**	0.0743 (0.0115)**	0.0972 (0.0301)**	0.0805** (0.0026)	0.1353** (0.0042)	0.0856 (0.0038)**	0.0764 (0.0030)**
City dummy (MSA Population>500,000)	0.0141 (0.0205)	0.0246 (0.0061)**	0.0072 (0.0049)		0.0310** (0.0054)	0.0260 (0.0078)**	0.0397 (0.0076)**
Ln(programmers in the organization) * City	−0.0183 (0.0089)*	−0.0118 (0.0072)	−0.0046 (0.0016)**		−0.0042+ (0.0022)	−0.0038 (0.0032)*	−0.0038 (0.0034)
Ln(programmers in the establishment) * City	−0.0252 (0.0197)	−0.0084 (0.0355)	−0.0117 (0.0158)		−0.0099 (0.0075)	−0.0140 (0.0077)+	−0.0133 (0.0081)+
Ln(establishment employment)	0.0637 (0.0122)**	0.0708 (0.0055)	0.0116 (0.0156)	0.0659** (0.0038)	0.0703** (0.00367)	0.0831 (0.0052)**	0.0626 (0.0047)**
Multi-establishment dummy		0.0135 (0.0153)	0.0145 (0.0083)+	0.0316** (0.0079)	0.0251** (0.00895)	0.0404 (0.0106)**	0.0250 (0.0100)*
Ln(Local IT workers)				0.0080** (0.0014)			
Ln(programmers in the organization) * Ln(local IT workers)				−0.00089** (0.00027)			
Ln(programmers in the establishment) * Ln(local IT workers)				−0.00078 (0.00075)			
Observations	6,708	86,792	86,792	86,872	86,872	24,240	58,767
LL	−3,640.56	−25,551.23	−26,355.95	−24,515.34	−24,370.63	−89,45.55	−15,377.88

Notes: Values represent marginal effects at means. Standard errors (calculated at mean values) are in parentheses. All regressions except Columns (2) and (3) are weighted to reflect the actual geographic distribution of establishments from County Business Patterns and include dummy variables for three-digit NAICS and month of survey. Significance levels do not change if standard errors are clustered by firm.

[a] Column (2) includes instruments for the number of programmers in the establishment and in the organization. Column (3) also includes instruments for the interactions of city with both the number of programmers in the establishment and in the organization.

[b] The regression estimates coefficients for the quadratic terms for the number of programmers in the establishment and in the organization and a term for their interaction. These are used in the marginal effect calculation.

[c] Manufacturing establishments are defined by two-digit NAICS 31, 32, and 33. Services are NAICS 42, 44, 45, 48, 49, 51, 52, 53, 54, 55, 56, 61, 62, 71, 72, and 81.

+Indicates significance at 90% confidence level; * indicates significance at 95% confidence level; ** indicates significance at 99% confidence level.

programmers increases the average probability of WEI by 1.21 percentage points.

We next examine the extent to which internal IT workers and cities are substitutes. The results in columns (2) and (3) of Table III present the main results of our paper. There is considerable evidence that internal resources substitute for the benefits of locating in a city, supporting Hypotheses 3a and 3b. The key effects are all in the expected direction. The interaction coefficients are significant at least at the 5% level and the marginal effects are significant for well over half of the observations. The results in column (3) show that establishments outside sizeable MSAs benefit 1.51 percentage points more from a one-standard-deviation increase in the log of programmers in the establishment (supporting Hypothesis 3a). Similarly, establishments outside sizeable MSAs benefit 1.72 percentage points more than establishments in sizable MSAs from a one-standard-deviation increase in the log of programmers at other establishments in the same firm (supporting Hypothesis 3b). Interestingly, although establishment programmers have a much stronger direct effect than organizational programmers (marginal effects of 7.99 percentage points vs. 1.21 percentage points), the extent of substitution between cities and establishment-level programmers is almost equal to that of cities and organization-level programmers.[18] This suggests that although only a fraction of organization-level resources are mobile, IT workers at other establishments do similar activities to those already employed at establishments doing investment, which means organizational resources can partly substitute similarly for cities. The marginal contribution of internal resources to innovation appears to be lower in cities than in other areas.

Columns (4) and (5) show that the main results are robust to the inclusion of an interaction between programmers in the establishment and programmers in other establishments in the same firm. Interestingly, the marginal effect of this interaction is negative: as the number of programmers in the establishment rises, having more programmers at other establishments in the same firm has less correlation to WEI investment. This provides further evidence of the reasoning behind Hypothesis 3b: programmers can be deployed throughout an organization and learning economies and knowledge can be transmitted across IT

18. We also examined the robustness of our marginal effects to changes in where the marginal effects are evaluated. We examined the distribution of marginal effects for nonzero capabilities, because capabilities are expected to increase the likelihood of investment only when they are nonzero. Of the 8,739 establishments with positive organization-level and establishment-level programmers, over 99% of organization-level interactions are negative and significant at the 5% level and over 90% of establishment-level interactions are negative and significant at the 5% level.

projects at different establishments.[19] Some prior research using patent citations has also provided evidence of intra-firm knowledge spillovers in other settings (e.g., Frost, 2001; Furman et al., 2005).

Figure 1 presents another view of the main results. It presents the predicted probabilities of investing in WEI using the results in column (2) of Table III under different combinations of location size and internal resources. Figure 1(A) presents results for the number of programmers in other establishments in the same firm. Figure 1(B) presents results for the number of programmers in the establishment. Both Figures 1(A) and (B) show that establishments located in sizeable MSAs have a greater likelihood of WEI investment (Hypothesis 1). For example, Figure 1(A) shows that when the organization has no programmers outside the establishment, location in a sizeable MSA increases the probability of investing considerably, from 11.6% to 16.1%. Moreover, Figure 1 provides support for Hypothesis 2: the upward sloping lines show that the probability of action increases as the number of programmers increases, whether or not the establishment is in a sizeable MSA. Most important, Figure 1 demonstrates how the prediction of Hypothesis 3 shapes behavior: The curve depicting establishments in sizeable MSAs is flatter than that for other establishments. The marginal impact of increasing the number of programmers is lower for establishments in sizeable MSAs.

Our interpretations of Table II rely on several assumptions. First, we assume that the location of an establishment is predetermined. This assumption is supported by the unexpectedly rapid diffusion of the Internet. Also, the establishments in our sample are large and from firms with long histories who had many reasons to locate an establishment in their chosen place, so, they did not suddenly relocate when the Internet became available. Second, the probit model limits inference. As in any probit model, we do not observe the variance of u_i. Therefore, at most we can infer whether the estimated direction of the net benefit function with respect to variables is consistent with predictions from theory. We can also infer whether the estimated direction is consistent with substitution/complementarity among inputs under the null. Though such findings are necessary but not sufficient for inference (see Arora and Gambardella, 1990; Athey and Stern, 2003), given the novelty of the question and setting, we believe these partial inferences still are interesting.

---

19. We do not emphasize this interaction result because the decision to hire at the establishment will be correlated with the decision to hire at the firm. Nevertheless, we view it as an important robustness check on the mobility of resources and knowledge within organizations.

A

B
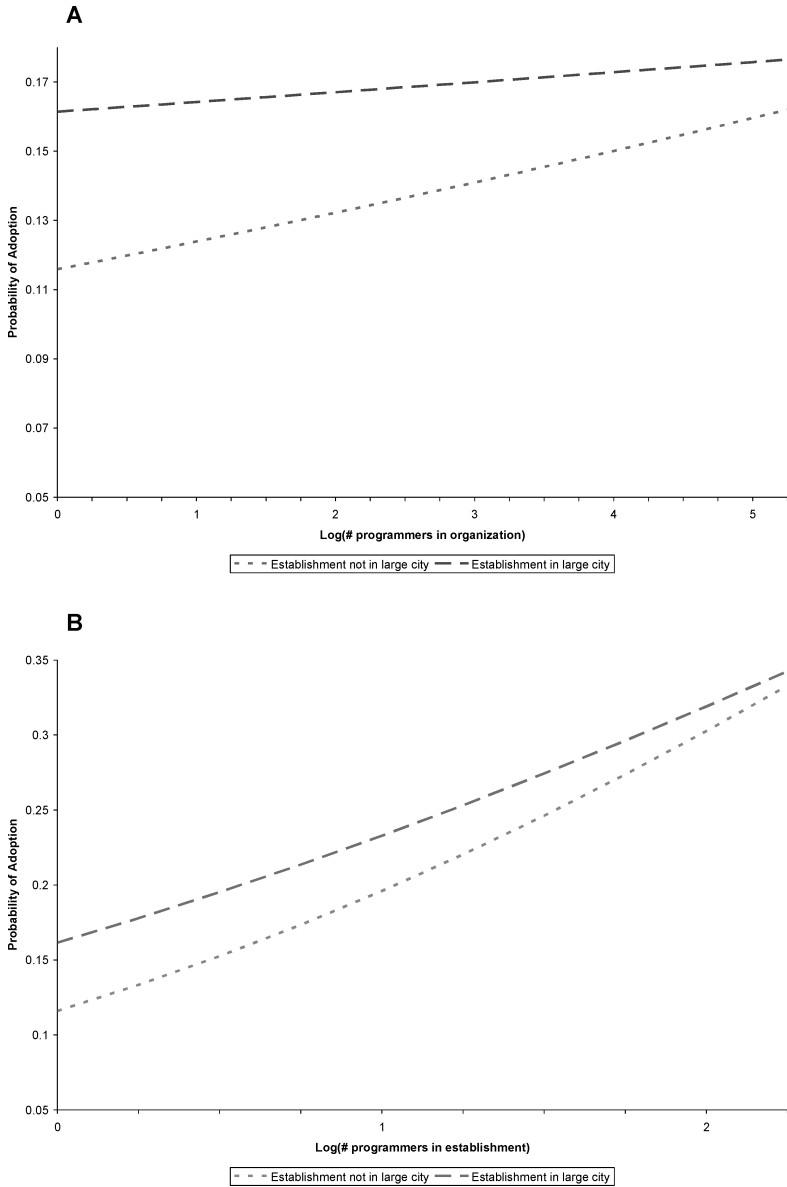
Predictions are based on a representative firm in the second half of 2000 with mean values of employment, industry effects, and multi-establishment status. Figure 1A assumes establishment capabilities are zero. Figure 1B assumes organizational capabilities are zero.

FIGURE 1. (A): PROBABILITY OF ADOPTION BY ORGANIZATIONAL CAPABILITIES WITH ESTABLISHMENT CAPABILITIES AS ZERO (B): PROBABILITY OF ADOPTION BY ESTABLISHMENT CAPABILITIES WITH ORGANIZATIONAL CAPABILITIES AS ZERO

We can test the sensitivity of our results to several crucial iden-
tifying assumptions. Our third assumption is that the errors are i.i.d.
Though this is a routine assumption, it implies that we assume that
the error in measuring the investment decision of one establishment
is independent of the error in every other establishment's decision,
including other establishments in the same firm. This assumption is
questionable for multi-establishment firms in which a central executive
decisionmaker (e.g., Chief Information Officer) possibly coordinates the
choice for each establishment under his domain and allocates mobile
internal resources (i.e., mobile IT workers) across establishments within
the firm. If IT investment decisions are centralized and these firms have
greater resources, then the coefficient estimates for resources for multi-
establishment firms will be biased. To look for such bias, Table III Col-
umn (1) estimates the coefficients with a subsample of establishments
with autonomy to make their own decisions.[20] Qualitative results do
not change.

Fourth, our econometric model assumes that the number of
programmers in the establishment and/or organization is statistically
exogenous. In support of this assumption, many of the establishments
in our sample maintain large Information Systems groups that support
many internal IT services, so that the WEI technologies will be only one
of many projects undertaken by such groups.[21] However, to examine
the robustness of our results to this assumption in Table III columns
(2) and (3) we present the results of instrumental variables regres-
sions. We use the number of programmers of other establishments and
organizations in the same industry as instruments for establishment-
level and organization-level programmers.[22] In column (2), we use
instruments for the establishment and organizational programmers
variables. In column (3), we use all five instruments for the four po-
tentially endogenous variables, namely, establishment programmers,
organizational programmers, and their interactions with being in a

---

20. Furthermore, although we present results without clustering of standard errors,
all of our significance results are robust to clustering the standard errors by firm.

21. We believe these endogeneity concerns are more likely to matter for program-
mers in the establishment than for programmers at other establishments in the same
firm because establishment-level investments are less likely to influence hiring at other
establishments even within the same firm. Therefore, we view our results or programmers
in the organization as particularly compelling evidence for substitution between external
and internal resources. Nevertheless, we show our results are robust to instrumenting for
both variables.

22. In particular, instrumental variables probit regressions were used. Following Mad-
dala (1983, p. 247–252), we used Amemiya Generalized Least Squares. In the first stage, the
endogenous variables are treated as a linear function of the instruments and the exogenous
variables. The second-stage probit uses the predicted values for the endogenous variables
from the first stage.

high-density area.[23] Although the significance on the interaction terms is sometimes lost, the signs do not change.

Table III also presents a handful of other robustness checks. Column (4) shows that results are robust to measuring external resources by the number of IT workers in the MSA rather than the population.[24] Interestingly, an increase in the number of external IT workers is associated with an increase in the likelihood of investment in WEI; the coefficient on external IT workers is positive and significant at the 1% level. This provides further evidence of the reasoning behind Hypothesis 1: the likelihood of investment in WEI is greater in large cities in part because of richer stocks of human capital in these locations. Column (5) shows that results are robust to a full translog-type specification. Columns (6) and (7) examine whether our results are relevant in all areas of the economy or if they are limited to services or manufacturing. The results suggest that, although we cannot reject our substitution Hypothesis 3 in services, the evidence for substitution between internal and external resources is strongest in manufacturing. Our working paper version also shows robustness to other measures of internal resources, to controls for the level of competition, to different definitions of being in a city, and to investment in other information technologies besides WEI (Forman et al., 2005b).

---

23. We define five instruments. First, we instrument for a firm's establishment capabilities with the establishment capabilities of other establishments in other firms in the same two-digit NAICS industry in the other locations that the firm has an establishment. Second, similarly, we instrument for a firm's organizational capabilities with the organizational capabilities of other establishments in other firms in the same two-digit NAICS industry in other locations where the firm has an establishment. These instruments should be correlated with the capabilities of an establishment but not with the propensity of the establishment to invest in WEI, conditional on its industry. Third and fourth, we use two instruments for the interaction of establishment capability and sizeable MSA. We interact the previous instrument for establishment capabilities with a dummy for sizeable MSA. We also use establishment capabilities at other establishments in other industries in the same location. These capabilities will be affected by the same local supply conditions but will not be directly correlated with the decision to invest. We construct our fifth instrument, the interaction of organizational capabilities and sizeable MSA, by interacting the above instrument for organizational capabilities (i.e., instrument 2) with a sizeable MSA dummy. (We do not use the organizational capabilities equivalent of the second establishment capabilities instrument, because it is not clear how organizational capabilities of establishments in a city will be correlated.) In sum we have five main instruments for four potentially endogenous variables.

24. To measure IT workers, we use data from the Bureau of Labor Statistics Occupational Employment and Wage Statistics program from 2000. The total number of IT workers is equal to the number of workers in the following categories: Computer and Information Scientists, Research (15–1011); Computer Programmers (15–1021); Computer Software Engineers, Applications (15–1031); Computer Support Specialists (15–1041); and Computer Systems Analysts (15–1051). These data are only available for MSAs—for establishments located outside of MSAs we added a missing data variable that is equal to one when this series was not available.

6. Discussion and Conclusions

During the early diffusion of the Internet, commentators speculated about the many ways in which new information technologies would rearrange the spatial distribution of economic activity (e.g., see Cairncross, 1997). In practice, such speculation described a long-run vision, not the short-run co-inventive activity of firms facing newly available information technology. In the short run, many establishments were in fixed locations while making innovative investments. In those locations they substituted between technically skilled IT workers employed at the establishment or in their parent organization and others hired from local labor markets.

In this study, we find extensive statistical evidence of localization of substitution between internal and external inputs into innovation. We show that establishments located in large urban areas innovate as if they face fewer constraints and have lower costs. We also find a symmetric role for internal resources: establishments that are in organizations with a greater number of computer programmers invested in WEI technology more frequently. Overall, we conclude that establishments engaged in co-inventive activity draw upon a variety of resources: internal establishment programmers, internal organizational programmers, and external purchased services. In contrast to prior work, we find that all of these channels are substitutes for one another as inputs into innovative investment.

These results have implications for understanding the sources of co-inventive activity required for process innovation, as well as managerial implications for the optimal location of innovative activity. In particular, these findings suggest that the advantages of agglomeration will be most important for single-establishment firms that have been unable to develop internal resources for innovative activity. The findings are consistent with those of researchers who have argued that agglomeration of firms with similar input demands can provide benefits through the provision of complementary third-party services. These benefits will be most valuable among small firms and for firms in young or infant industries, where the firm-specific human capital of IT staff and business processes are still being developed.[25] More generally, the findings are consistent with Saxenian's (1996) observation that managers at firms that anticipate innovating will be better off locating near other firms that are innovating.

25. For example, our results are consistent with the global distribution of firms engaged in software development. Although small independent firms engaged in software development in countries such as India and Ireland cluster in a relatively small number of areas, the location of large US firms that produce software products or services is distributed throughout the United States and worldwide.

Our findings are also consistent, albeit more speculatively, with those of researchers who have argued that as industries mature and average firm size increases, there is less need for the complementary resources and knowledge transfer found in cities. As a result, firms may relocate to shape their innovative activities (Furman et al., 2005), or to economize on transportation costs or save on wages (Duranton and Puga, 2001). Nevertheless, caution is warranted in following this line of reasoning. We have examined only one reason why firms would desire urban locations. Firms may agglomerate in the same location for a variety of reasons: knowledge transfer, labor market pooling, knowledge spillovers, transportation costs, and others.

Finally, our results have implications for ongoing research about outsourcing. It is a comparatively unexplored theme in outsourcing research whether the location of an establishment shapes the propensity of establishments to use market-mediated external channels. Our evidence about investment in WEI suggests location is a determinant of the outsourcing decision. Furthermore, our results suggest that location will matter more when the firm has fewer internal resources. Using direct measures of outsourcing to better understand the roles of location and internal resources in this context is an interesting subject for future research.

References

Ai, C. and E.C. Norton, 2003, "Interaction Terms in Logit and Probit Models," *Economics Letters*, 80, 123–129.

Arora, A. and A. Gambardella, 1990, "Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology," *Journal of Industrial Economics*, 38(4), 361–379.

—— and ——, 1994, "Evaluating Technological Information and Utilizing it: Scientific Knowledge, Technological Capability and External Linkages in Biotechnology," *Journal of Economic Behavior and Organization*, 24(1), 91–114.

Arora, A., A. Gambardella, and S. Torrisi, 2004, "In the Footsteps of Silicon Valley? Indian and Irish Software in the International Division of Labor," in T. Bresnahan and A. Gambardella, eds., *Building High-Tech Clusters: Silicon Valley and Beyond*, Cambridge, MA: Cambridge University Press.

Astebro, T., 2002, "Noncapital Investment Costs and the Adoption of CAD and CNC in U.S. Metalworking Industries," *RAND Journal of Economics*, 3(4), 672–688.

Athey, S. and S. Stern, 2003, "An Empirical Framework for Testing Theories about Complementarities in Organizational Design," Working Paper, Northwestern University.

Attewell, P., 1992, "Technology Diffusion and Organizational Learning: The Case of Business Computing," *Organizational Science*, 3, 1–19.

Banker, R. and S. Slaughter. 1997. "A Field Study of Scale Economies in Software Maintenance," *Management Science* 43: 1709–1725.

Barnes, B.H. and T.B. Bollinger. 1991. "Making reuse cost-effective." *IEEE* Software 8(1): 13–24.

Bresnahan, T. and A. Gambardella. 2004. "The Division of Inventive Labor and the Extent of the Market," in *General Purpose Technologies and Economic Growth*, E. Helpman, ed. Cambridge, MA: The MIT Press.

Cairncross, F., 1997, *The Death of Distance*, Cambridge, MA: Harvard University Press.

Cassiman, B. and R. Veugelers, 2006, "In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition," *Management Science*, 52(1), 68–82.

Cohen, W.M. and D.A. Levinthal, 1989, "Innovation and Learning: The Two Faces of R&D," *The Economic Journal*, 99, 569–596.

—— and ——, 1990, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, 35, 128–152.

Duranton, G. and D. Puga, 2001, "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economic Review*, 91(5), 1454–1477.

—— and ——, 2002, "From Sectoral to Functional Specialization," NBER Working Paper No. 9112.

Forman, C., A. Goldfarb, and S. Greenstein, 2002, "Digital Dispersion: An Industrial and Geographic Census of Commercial Internet Use," NBER Working Paper No. 9287.

——, ——, and ——, 2005a, "How Did Location Affect Adoption of the Commercial Internet? Global Village versus Urban Leadership," *Journal of Urban Economics*, 58(3), 389–420.

——, ——, and ——, 2005b, "Technology Adoption In and Out of Major Urban Areas: When Do Internal Firm Resources Matter Most," NBER Working Paper No. 11642.

Frost, T. 2001. "The Geographic Sources of Foreign Subsidiaries Innovations." *Strategic Management Journal* 22: 101–123.

Furman, Jeffrey L., Margaret K. Kyle, I. Cockburn, and R. Henderson, 2005, "Public & Private Spillovers, Location, and the Productivity of Pharmaceutical Research," Working Paper, School of Management, Boston University.

Goolsbee, A. and P. Klenow. 2002. "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." *Journal of Law and Economics* 45: 317–344.

Griliches, Z., 1957, "Hybrid Corn: An Exploration in the Economics of Technical Change," *Econometrica*, 25, 501–522.

Henderson, J.V., 2003, "Marshall's Scale Economies," *Journal of Urban Economics*, 53(1), 1–28.

Jaffe, Adam B., M. Trajtenberg, and R. Henderson, 1993, "Geographic Localization of Knowledge Spillovers as Evidenced in Patent Citations," *Quarterly Journal of Economics*, 108(3), 577–598.

Kelley, M. and S. Helper, 1999, "Firm Size and Capabilities, Regional Agglomeration, and the Adoption of New Technology," *Economics of Innovation and New Technology*, 8, 79–103.

Maddala, G.S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, UK: Cambridge University Press.

Porter, Michael E., 1998, "Clusters and Competition: New Agendas for Companies, Governments, and Institutions," in Michael E. Porter, ed., *On Competition*, Boston: Harvard Business School Press.

Rosenberg, N., 1990, "Why do Firms do Basic Research (with Their Own Money)?" *Research Policy*, 19, 165–174.

Saxenian, A., 1996, *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, Cambridge, MA: Harvard University Press.

Singh, J., 2004, "Multinational Firms and Knowledge Diffusion: Evidence Using Patent Citation Data," Working Paper, INSEAD, Singapore.

Vernon, R., 1963, *Metropolis 1985*, New York: Doubleday.