NET Institute*

www.NETinst.org

Working Paper #09-20

September 2009

Heterogeneity and Information Spillovers in Web Service Sourcing

George Deltas and Seung-Hyun Hong

University of Illinois

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

Heterogeneity and Information Spillovers in Web Service Sourcing*

George Deltas
University of Illinois
deltas@illinois.edu

Seung-Hyun Hong
University of Illinois
hyunhong@ad.uiuc.edu

September 2009

Abstract

This paper studies the role of firm heterogeneity and information spillovers in the sourcing decision to provide web services. To this end, we develop a theoretical model that relates these factors to firms' decisions to outsource or use in-house resources (insource). Based on this theoretical framework, we further construct an econometric model. Using our estimated model, we investigate why insourcing of web services is much more prevalent than outsourcing. We find that insourcing is likely to generate the higher value for most firms, and lack of information on the efficacy of outsourcing is unlikely to account for the dominance of insourcing over outsourcing. Therefore, differences in firm choices with regard to web services provision is primarily driven by firm heterogeneity, rather than by information spillovers.

Keywords: Heterogeneity; Information spillovers; Outsourcing; Switching regression

JEL classification: L22, L23, L86

*We thank the NET Institute for financial support. All errors are our responsibility.

Correspondence can be addressed to either of the authors at Department of Economics, University of Illinois at Urbana-Champaign, 1407 West Gregory Drive, Urbana, IL 61821.

1 Introduction

The advance and diffusion of information technology (IT), often together with organizational changes, has been shown to be important sources of productivity growth (e.g., Bresnahan, Brynjolfsson, and Hitt 2002; Brynjolfsson and Hitt 2003; Stiroh 2002). Consequently, significant benefits are expected for those adopting new IT. Nevertheless, many firms might still delay their adoption of new IT because the adoption of new technology tends to entail high fixed costs (Stoneman 2002). However, a distinct feature of IT is that its recent advance has also facilitated the outsourcing of IT services. Firms can thus avoid high fixed costs by outsourcing, while still benefiting from new information technologies.

In this paper, we study both IT adoption and IT outsourcing, and investigate what are important determinants in firms' decisions with regard to developing an IT service in-house or outsourcing its development to an outsider vendor. Several empirical papers have studied either IT adoption or IT outsourcing, but few empirical studies have directly examined both insourcing and outsourcing of IT.[1] Because we consider both sourcing decisions, our study is also related to the large literature on make versus buy (e.g., Baker and Hubbard 2003; Hubbard 2000). However, IT outsourcing is a relatively new business practice, so that it can be considered as a new technology, rather than a buy decision. Moreover, if firms do not have sufficient information on IT outsourcing, they would not consider IT outsourcing at all, suggesting that there is no make versus buy problem for these firms. We consider this aspect as well, hence partially contributing to the empirical literature on the boundary of firms.

We focus on IT adoption and IT outsourcing to provide web services. The provision of websites or other advanced web services such as business-to-business (B2B) and business-to-consumer (B2C) exchanges requires several inputs including web servers, web application software, and programmers or web developers. Firms may insource these inputs by adopting web servers and employing

[1] Forman, Goldfarb, and Greenstein (2005, 2008) examine IT adoption and geographic location, where location indirectly captures the extent of local markets for IT outsourcing. Arora and Forman (2007) study the relationship between IT outsourcing and local supply of outsourcing, but do not directly consider insourcing or IT adoption.

web developers, but they can alternatively outsource web hosting services. To study these sourcing decisions, we use the data from *Computer Intelligence Technology Database* (CITDB). The CITDB contains detailed information on establishment characteristics and whether establishments provide web services such as websites, B2B, B2C, or web-based customer services. The CITDB is particularly useful for our purpose, since it collects establishment-level information on web hosting outsourcing as well as variables related to in-house provision, including web server adoption and the number of programmers and web developers.[2]

Our data show that approximately 60% of firms provided web services using either in-house resources or outsourcing. Among them, about two thirds of firms relied only on in-house resources, and more firms have chosen to insource over time. This observation thus raises our key question: Why do more firms provide web services using in-house resources and do so at a progressively increasing rate, even though outsourcing might seem more attractive, because it does not entail high fixed costs associated with adopting web servers and employing web developers.

To study this question, we examine potential determinants in the sourcing decisions, particularly focusing on the following two factors. The first is heterogeneity or firm-specific needs in the provision of web services. Web hosting companies, for instance, may provide only standardized services or limited customized services, in which case firms with specific needs may not outsource web hosting services because their value of customized product, developed in-house, could outweigh any cost saving benefit from using standardized web hosting services. The second factor is related to information spillovers. Because IT outsourcing is a relatively new business practice, sufficient information on web hosting services might not have been diffused widely. As a result, it is possible that if all firms had sufficient information about the efficacy of outsourcing, most of them could outsource web hosting services.

We focus on these two factors because of their potential implications on efficiency. For example,

[2]Note that our main empirical analysis focuses on the decisions of single-establishment firms, so that we avoid any firm-wide considerations which we should take into account if we examine multi-establishment corporations. As a result, we use "firms" and "establishments" interchangeably in this paper.

if most firms have specific needs and prefer their own customization, disseminating more information on outsourcing would not necessarily lead more firms to outsource. Hence, the observed tendency toward insourcing is less likely to be inefficient. In contrast, if firm heterogeneity is not an important factor and most firms are not aware of outsourcing opportunities, then the dissemination of relevant information and experience would lead to more outsourcing, suggesting that the observed insourcing trend is unlikely to be efficient.

To investigate the role of these factors in the sourcing decision, we first develop our theoretical model and study the optimal sourcing decisions. Based on the theoretical framework, we then construct an econometric model, in which we separate two regimes – one where outsourcing is a potentially viable option, and the other without the possibility for outsourcing. In the first regime, firms choose between outsourcing, in-house provision, and no web service, which leads to a multinomial logit model for the choice probability. In the second regime, outsourcing is not available, hence resulting in a binary choice model. Because our theoretical model also implies a threshold crossing model for regime switching, we combine the regime switching probability and the choice probability to construct a switching regression model.

The estimated model shows that the mean of the regime switching probability is about 0.7, implying that outsourcing was not viable only for 30% of our sample. Hence, for at least 70% of firms in our sample, information spillover is less likely to explain why more firms provided web services using in-house resources, instead of outsourcing.[3] We use the estimated parameters of our model to perform a number of counter-factual experiments. In the first experiment, we consider the possibility that outsourcing is viable for all firms. We find that if that were the case, the number of firms outsourcing would increase by about 10%, which suggests that information spillover might be an important factor to some extent. However, in our second experiment, where we assume an exogenous increase in the number of other firms that outsourced in the same region or industry,

[3]We note that even for firms in the regime where outsourcing is viable, information spillovers *can* increase the probability that outsourcing becomes the actual *choice*. However, the impact of information spillovers on this set of firms is likely small, compared to the firms for which outsourcing is not initially perceived as being a viable choice.

we do not find any significant increase in the outsourcing propensity. Because the number of other firms outsourcing is likely to be positively associated with the degree of information spillovers, this second result suggests that information spillovers may be relatively unimportant, and that firm heterogeneity is more likely to account for the dominance of in-house provision over outsourcing.

In the next section, we develop our theoretical framework. Section 3 describes our data and key variables, and presents descriptive statistics. Section 4 constructs our empirical model and discusses estimation approach. Section 5 presents the results. Section 6 concludes the paper.

2 Theoretical Framework

In this section, we develop our theoretical framework to study firms' sourcing decisions in the provision of web services. We first describe our model environments, in which we introduce several parameters that capture the extent of heterogeneity and information spillover. We then examine the optimal sourcing decisions and investigate the role of heterogeneity and information spillover in these decisions, which leads to our empirical framework developed in section 4. Finally, we relate the parameters in our theoretical model to variables observed in our data.

2.1 Model Environment

We consider a firm's decision on how to provide advanced web services such as websites or business-to-business exchanges, which require nontrivial investment in several inputs including web servers, web application software, and web application developers.[4] Firms can develop their web services using in-house resources, or outsource their development to a second party. They can also choose not to develop any web service.

Firms have needs for web services of particular types. Web service types are characterized by their location on a unit (Salop) circle. Firm i desires a web service of location or attribute x_i, where distance is measured clockwise from the top of the circle. A set of independent providers/vendors

[4] We do not consider basic web services such as mere Internet connection or e-mail which do not require significant investment in information technology.

can provide web services with attributes given by locations x_j, with $j = 1, \ldots, J$. A vendor can potentially have developed multiple web services at different locations; ownership is only relevant for price determination, but as will become clear below, not for whether a firm outsources or not (given our assumption of efficient bargaining). A firm can develop a web service of any type by incurring a fixed cost of F, or forgo the use of a web service.

The transaction price between a firm and a vendor is determined via negotiations, given that many purchases require minor customization and involve firm specific transaction costs. We do not model the price determination process, except to assume that it is efficient: if there are gains from trade, the two parties will transact; otherwise, a firm will produce in-house or not have the web service. We understand that the assumption of efficiency in transactions is not realistic, but we believe any market imperfections are essentially orthogonal to the issues we want to understand and thus their omission does not detract from the other aspects of the model.

The choice of a web service with an attribute x_p is related to the firm's gross (expected) profit through the function[5]

$$\pi = R(1 - r|x_j - x_p|) \tag{1}$$

where R is a measure of the scale of the firm and r is a measure of the sensitivity of the firm's profits to the fit of the web service with its business model and lies in the $[0, 1]$ interval. We adopt such a parsimonious reduced form representation of the profit function because our data does not provide any information on the deeper constituent elements of the profit function. To derive the optimal choice of the firms, we need to compute the payoff for each of the three possible choices.

Let the nearest available by a vendor web service have location x_a, and assume without loss of generality that $x_a > x_j$. Then, the firm's net payoff function from outsourcing and obtaining a web service with characteristics x_a is

$$\Pi_{\mathsf{out}} = R(1 - r(x_a - x_j)) - P - C \tag{2}$$

[5]We recognize that many other factors, unknown at the time of the web service development, will affect the firm's profit. Throughout, "profit" will refer to "expected profit" with the expectation taken over all these other factors.

where P is the price payed to the vendor, and C is a transaction cost which includes the costs associated with identifying the vendor, negotiating any firm-specific adaptation of the web service, and concluding the negotiations.

If the firm uses in-house resources, it will choose to build the desired web service, with attribute x_j and its net profit will be

$$\Pi_{\mathsf{inh}} = R - F \tag{3}$$

where F is the in-house development cost. Finally, if it chooses to not have a web service, we assume that its profits are given by

$$\Pi_{\mathsf{no}} = R(1 - r) \tag{4}$$

or, equivalently, that the (gross of costs) value of the least suitable web service relative to no web service is equal to the value of the most suitable web service relative to the least suitable web service. We could readily relax this assumption at the cost of adding one more parameter to the model.

A firm will make the choice that maximizes its net profits. Given our assumption of efficient bargaining and that the cost of the web services that are already developed is zero, the price P that is relevant to determine whether the firm will outsource or not is zero (the transactions price will be higher, but the firm that is indifferent between outsourcing and not outsourcing will pay a price of zero). Thus, the firm's net profits are given by

$$\Pi^* = \max\{R(1 - r(x_a - x_j)) - C,\ R - F,\ R(1 - r)\} \tag{5}$$

Note that the payoff increases with firm size, R, but does so fastest with the house provision, slowest with no web service, and of intermediate rate with a web service purchased from a vendor. This is illustrated in Figure 1, which plots the profit for each of the three possible choices as a function of firm size, as well as the profit at the optimal web service sourcing decision (which is the upper envelope of the above three functions).

2.2 Optimal Firm Decision

To determine the profit maximizing firm decision, we make pair-wise comparisons of the payoffs of each of the three possible choices. A firm prefers outsourcing to in-house production if

$$\begin{aligned} R(1 - r(x_a - x_j)) - C &> R - F \Rightarrow \\ R - rR\Delta x &> R - \Delta c \Rightarrow \\ rR &< \frac{\Delta c}{\Delta x} \end{aligned} \tag{6}$$

where $\Delta x = |x_a - x_j| > 0$ and $\Delta c = F - C$. Note that Δc is likely to be positive because developing a web service from scratch is certainly more expensive than any transaction costs from obtaining this web service from a vender (recall these costs do not include the price). The product on the left side of the above inequality, $r \times R$, which will appear often in the theoretical model, is a measure of the demand for the web service, which increases with the size of the firm and with the parameter r denoting the degree of the firm's dependence on the web service for its business (low values of r indicate that a firm does not lose much profit by having a less appropriate web service or no web service at all).

A firm prefers outsourcing to not having a web service if

$$\begin{aligned} R(1 - r(x_a - x_j)) - C &> R(1 - r) \Rightarrow \\ R - rR\Delta x - C &> R - rR \Rightarrow \\ rR(1 - \Delta x) &> C \Rightarrow \\ rR &> \frac{C}{1 - \Delta x} \end{aligned} \tag{7}$$

A firm prefers in-house development to no web service if

$$\begin{aligned} R - F &> R(1 - r) \Rightarrow \\ rR &> F \end{aligned} \tag{8}$$

From the above inequalities, it is clear that if the demand for web services rR is sufficiently low, the firm will choose to have no web service (this choice dominates in all pair-wise comparisons as rR goes to zero), whereas if rR is sufficiently high the firm will prefer in-house production (this choice dominates in all pair-wise comparisons as rR becomes larger than all right-hand-side expressions). This can also be readily seen by a quick inspection of Figure 1. The question is for which parameter values outsourcing is the profit maximizing choice for firms with intermediate demand for a web service, i.e., for firms of intermediate "effective" size. Given the two profit function slope restrictions, this boils down to a single inequality. It must be that a firm that is just indifferent between no web service and outsourcing is smaller than a firm that is just indifferent between outsourcing and producing in-house. In the notation of Figure 1, it must be that $R_1 < R_3$. Equivalently, it must be that the right-hand-side of the inequality (7) is smaller than the right-hand-sided of (6), i.e.,

$$\begin{aligned} \frac{C}{1-\Delta x} &< \frac{\Delta c}{\Delta x} \quad \Rightarrow \\ C\Delta x &< (F-C)(1-\Delta x) \quad \Rightarrow \\ C\Delta x &< F + C\Delta x - F\Delta x - C \quad \Rightarrow \\ C &< F(1-\Delta x) \quad \Rightarrow \\ \frac{C}{1-\Delta x} &< F \end{aligned} \tag{9}$$

This inequality means that for outsourcing to be a potentially viable option for a firm, the cost of outsourcing, normalized by how close is the best available web services to the firm's needs, must be lower than the firm's cost of producing (the ideal) web services in-house.[6] On the other hand, if this inequality is reversed, then the normalized cost of outsourcing dominates the cost of in-house production, so that outsourcing is not a viable option anymore, as illustrated by Figure 2. The above can be summarized in the proposition below.

[6]Also notice that this condition also implies, after some algebra, that a firm that is just indifferent between outsourcing and in-house production is larger than the firm that is indifferent between in-house production and doing nothing ($R_3 > R_2$), and that this firm is turn bigger than the firm that is bigger than the firm that is indifferent between outsourcing and doing nothing ($R_2 > R_1$).

Proposition 1 *Suppose that* $\frac{C}{1-\Delta x} < F$*, that is, the normalized cost of outsourcing is lower than the cost of in-house production. Then,*

(a1) *Firms with a low demand for web services (*$rR < \frac{C}{1-\Delta x}$*) will not have a web service.*

(b1) *Firms with an intermediate demand for services (*$\frac{C}{1-\Delta x} < rR < \frac{\Delta c}{\Delta x}$*) will outsource.*

(c1) *Firms with a high demand for web services (*$rR > \frac{\Delta c}{\Delta x}$*) will design a web service in-house.*

Conversely, suppose that $\frac{C}{1-\Delta x} > F$*, i.e., the normalized cost of outsourcing is larger than the cost of in-house production. Then,*

(a2) *Firms with a low demand for web services (*$rR < F$*) will not have a web service.*

(b2) *Firms with a high demand for web services (*$rR > F$*) will design a web service in-house.*

Figure 3 illustrates case 1, in which outsourcing is a potential middle ground between no provision and in-house provision, while Figure 4 illustrates case 2, under which outsourcing is not viable. Holding the distribution of firm size R constant, we immediately obtain the following comparative statics on the set of firms that choose each of the three options.

Corollary 1 The set of firms that do not have a web service increases with outsourcing transaction costs C, the in-house development costs F, and the unsuitability of vendor services Δx; it decreases with the importance of an appropriate web service r.

Corollary 2 The set of firms that outsources the web service provision increases with the suitability of vendor services $\frac{1}{\Delta x}$ and the in-house development costs F; it decreases with the outsourcing transaction costs C. An increase in the importance of an appropriate web service r has an ambiguous effect on the number of firms that outsource.

Corollary 3 The set of firms that develop their web services in-house increases with outsourcing transaction costs C, the importance of an appropriate web service r, and the unsuitability of vendor services Δx; it decreases with the in-house development costs F.

These comparative statics are related to the formulation of our empirical analysis, as we describe in section 2.4. They naturally lead to a single equation discrete choice estimation framework, and we examine them in section 3.3.

Notice the Corollaries 1 to 3 pertain to the choice of a firm unconditional to whether case (a) or case (b) are relevant for that firm. That is, we did not condition on values of the cost parameters and the vendor web service suitability such that the outsourcing is potentially viable (or that it is always non-viable). A distinction between the comparative statics of the firms in the two cases leads naturally to a switching regimes model. It can be useful in evaluating how reasonable the above stylized framework is, as the partitioning into the two regimes imposes some restrictions in the data. In particular, the comparative statics in regime (a) are given as follows:

Corollary 1a Suppose that outsourcing is a potentially viable option, that is, $\frac{C}{1-\Delta x} < F$. Then, the set of firms that do not have a web service increases with outsourcing transaction costs C , and the unsuitability of vendor services Δx; it decreases with the importance of an appropriate web service r. A (marginal) change in the in-house development costs F has no effect on the set of firms that have no web service.

Corollary 2a Suppose outsourcing is a potentially viable option, i.e., $\frac{C}{1-\Delta x} < F$. Then, the set of firms that outsources the web service provision increases with the suitability of vendor services $\frac{1}{\Delta x}$ and the in-house development costs F; it decreases with the outsourcing transaction costs C. An increase in the importance of an appropriate web service r has an ambiguous effect on the number of firms that outsource.

Corollary 3a Suppose outsourcing is a potentially viable option, i.e., $\frac{C}{1-\Delta x} < F$. Then, the set of firms that develop their web services in-house increases with outsourcing transaction costs C, the importance of an appropriate web service r, and the unsuitability of vendor services Δx; it decreases with the in-house development costs F.

Note that Corollaries 2a and 3a exhibit the same comparative statics as Corollaries 2 and 3, but that Corollary 1 differs. The comparative statics in regime (b), where only two options are potentially viable, are given below:

Corollary 1b Suppose outsourcing is not a potentially viable option, i.e., $\frac{C}{1-\Delta x} > F$. Then, the set of firms that do not have a web service increases with in-house development costs F, and it decreases with the importance of an appropriate web service r. A (marginal) change in the outsourcing transaction costs C and in the unsuitability of vendor services Δx has no effect on the set of firms that have no web service.

Corollary 3b Suppose outsourcing is not a potentially viable option, i.e., $\frac{C}{1-\Delta x} > F$. Then, the set of firms that develop their web services in-house increases with the importance of an appropriate web service r and it decreases with the in-house development costs F. A (marginal) change in the outsourcing transaction costs C and in the unsuitability of vendor services Δx has no effect on the set of firms that have no web service.

Notice that both of the above corollaries differ in terms of comparative statics from their unconditional counterparts. Even though these conditional comparative statics are very sharp in their predictions, they could well have some qualitative validity; they capture the notion that for some firms (defined by the parameter condition) outsourcing may not be a realistic option, and for these firms that choice between the two remaining possibilities should not depend on any factor that is payoff relevant to the outsourcing decision.

2.3 Dynamics

The preceding analysis is static in nature. The fundamentals that affect a firm decision for web service provision can vary over time. How should this affect the optimal choice of web service provision? Once a web service has been developed in-house, the costs of maintaining it are likely small. Thus, changes in the costs of, say, transacting with the market should not affect the firm decision to not-outsource, even though outsourcing may have been profit maximizing if the decision were to be made today. Moreover, forward looking firms should take into consideration the possibility that fundamentals may change and when making their web service provision decision (e.g. by postponing web service development if they expect outsourcing transaction costs to go down). However, such dynamic “option value” considerations are likely of second order importance in the decision to obtain a web-site and on whether to outsource it or build it in-house. Current value considerations likely dominate largely because of obsolescence of underlying technologies and because of high turnover of involved personnel. Moreover, a careful treatment of dynamics would likely not materially affect the discussion below at it relates to our empirical strategy.

Therefore, dynamics can be modeled through the presence of obsolescence of currently adopted web services, and their replacement by successor web services. A firm that has made a particular decision regarding the provision of a web service may not make the same decision in the next instance due to changes in their desired web service type, the vendor offerings, the size of the firm, the importance of a web service in its business model, and costs associated with the development and outsourcing of web services.

In particular, we assume that firms with no web service make an annual decision as to whether to outsource or develop one on the basis on the trade-offs described in the above section. Existing web services become obsolete at an exogenous rate ρ and would need to be re-developed. Reasons for obsolescence may involve changes in operating systems, hardware capabilities, changes in the design state-of-the-art, etc. A firm that currently outsources may choose to produce a web-house in-house even if the current web service is not obsolete if its preferred web service characteristics, x_j changes sufficiently from its prior value to make the existing match very ineffective in meeting the firm’s needs.

Without putting structure on the processes governing the evolution of r, F, C, R, and x_j, it is not possible to say which transitions are more likely than others. Clearly, however, changes in from one state to another will be related to the changes in the distributions in the underlying parameters on the basis of the comparative statics discussed in section 2.2, and they will be distributed over time. For example, we would expect that a reduction in the costs of contracting with vendors will lead to increases in outsourcing and a reduction of in-house production and in the number of firms that do not have a web service, and these changes will be distributed over time.

2.4 Links to Empirical Analysis

Though we do not directly observe the values of the parameters defined in the preceding sections, we do observe (i) proxies and (ii) variables that are causally related with higher or lower values of these parameters. The parameter R, indicating the scale of the firm, can be proxied by the number of employees or firm revenue. The latter of the two variables may be partially endogenous to the web service provision decision (and to a far lesser extent, so could the former).[7]

The degree of the firm's dependence on the web service for its business, denoted by r in our theory, is likely higher for firms that deal primarily with consumers, especially those with a mail-order business. It is likely to be lower for firms that sell undifferentiated products to businesses. Classification on the relative value of a suitable web service is made on the basis of the NAICS industry group that a firm belongs to. Given that much of our analysis relies on fixed effects, much of this systematic effect is absorbed by them, though there likely remains some additional heterogeneity at the firm level. The importance of web services may also increase with the degree of competition in an industry, which could be proxied by concentration and market share volatility (indicating that the customer base of a firm is contestable). The fraction of competing firms who already have a web service should also increase the value of a particular firm of having a web service itself, and more so in competitive industries.

The costs of in-house web service development, F, are expected to decline as the firm becomes more sophisticated in their use of information technology. This can be proxied by the number of programmers (or more generally the number of IT employees) employed by the firm (excluding

[7] In principle, both employees and revenue could be instrumented by the average number of employees and revenue of a firm's competitors which should proxy for firm size driven by technological industry-specific considerations. Given the very short nature of our panel, lagging these variable is not a feasible option unless we treat the data as a cross-section.

those who are directly engaged in web-development). Programming technology also changes over time, and thus F is likely to change over time.

The costs of transacting with a vendor, C, depend on a number of unobserved factors. However, we expect them to decrease with the familiarity that a firm has with interacting with such vendors. Firms that have used vendors in the past are likely to have lower transaction costs. These past transactions may not be limited to web service outsourcing, but also include the outsourcing of other IT services, such as help desk, LAN services, or programming. Similarly, if firms have observed outsourcing of web service services by other firms and have been able to learn from it, they are also likely to experience lower transaction costs (including lower search costs of identifying an appropriate vender and suitable product).

Such information spillovers would be positively associated with (i) the number/percentage of firms who have outsourced among those located in the same city/region, regardless of the industry that they are in (spatial spillovers), and (ii) the number/percentage of firms who have outsourced among those in the same industry, regardless of where they are located (industry spillovers). The strongest information spillovers should take place from firms that have outsourced and are both in the same industry and in the same geographical location. The number/percentage of such firms may decrease transaction costs even after controlling for spatial and industry spillovers as characterized above.[8] The transaction costs may also depend on unobserved and nearly permanent location characteristics (e.g., the quality of the internet backbone in a city). Changes in the IT industry and technology (e.g., communications technology, product standardization, etc.) suggest that C will also vary secularly from year to year.[9]

Finally, the suitability of vendor products, measured by the distance in the characteristics space between the attribute desired by the firm and the closest attribute available by the industry, is likely to depend on the thickness of the industry offerings. This can be proxied by the number of firms

[8]Quantitatively smaller spillovers may occur by observing the outsourcing of other services, such as programming, LAN services, etc.

[9]Incidentally, the fact that both C and F are changing, and likely trending down, over time implies that there is no expected trend in the fraction of firms that outsource.

that outsource their web services, as the more such firms purchase from the industry, the greater the variation in web service design that can be supported (clearly, there is a feedback loop between vendor offerings and number of customers: we draw no causal inference here, but only use this measure as a proxy). The association between suitable product availability for a particular firm and number of other firms that have outsourced their web service design is likely to be stronger when these firms belong in the same (or similar) industries, than if they are in very different industries.

These variables directly or indirectly reflect the extent of heterogeneity and information spillover in firms' sourcing decisions, and section 4 develops our empirical framework to examine them. In the next section, we describe specific variables observed in our data that correspond to the variables discussed above.

3 Data and Descriptive Statistics

3.1 Data Description

We use data from the 2000-2004 CITDB collected by the Harte-Hanks Market Intelligence. The CITDB is a yearly survey of over 100,000 establishments in the United States. It contains detailed establishment-level data on the use of a variety of information and communication technologies. This dataset has been used in several papers (e.g., Bresnahan, Brynjolfsson, and Hitt 2002; Bresnahan and Greenstein 1996; Forman, Goldfarb, and Greenstein 2005, 2008; Hong and Rezende 2008). The CITDB is useful for our purpose because it contains information on both insourcing and outsourcing. The unit of observation is an establishment in a year. The Harte-Hanks has attempted to survey the same establishment each year, so that the dataset contains panel information of many establishments. Because of the voluntary nature of the survey, however, some establishments did not respond to survey requests, and the CITDB has added new establishments each year. As a result, while the total number of observations remains similar each year, many establishments were not surveyed every year. Nevertheless, we will make use of this incomplete panel information to improve identification on heterogeneity and information spillover.

We focus on the sourcing decisions to provide web services, but not other IT services, for two reasons. First, the recent growth of the Internet stems from the development and provision of advanced web services such as websites or web-based consumer services. Thus, understanding the adoption pattern of web services is important. Second, though the CITDB contains detailed information on the adoption of various information technologies as well as of outsourcing of several IT services, our data allow us to construct variables for both insourcing and outsourcing for the provision of web services, but not for other IT services such as help desk or systems integration. In other words, we have no reliable way to determine whether a firm which does not outsource, for example, the help desk function, has developed a similar function in-house.

We restrict our sample as follows. First, we only consider the decisions of single-establishment firms, in order to help isolate the sourcing decisions as described in our stylized theoretical framework, and avoid dealing with any firm-wide considerations.[10] Therefore, our empirical analysis does not examine the decisions of multi-establishment firms, and we use firms and establishments interchangeably in this paper. Second, we do not use observations in our data if they contain outdated information on computing technology. The CITDB does not survey all establishments each year, so that for some observations, information collected in the previous year is re-used. However, the CITDB also reports the date when the information was collected, and we use only observations with up-to-date information.

3.2 Variable Definition

To examine the sourcing decisions to provide web services, we define two key variables based on the information contained in our data. First, we define web.outsourcing to be a dummy indicator variable for whether a firm outsources hosting web servers, web site management, or web application

[10]The investigation of firm-wide versus partial establishment-only adoption issues is a promising topic for future research. For example, for some lines of business, a single establishement of a firm may develop in-house resources to be used by all establishments. In other lines of business, an establishment may develop in-house resources, while other establishments have no need for the use of any such services. Finally, in other firms, an establishment may develop in-house resources for its own use, while other establishments of the same firm may outsource the development of web services for their own particular needs.

development. Though the CITDB separates these outsourcing activities, we combine them because all these activities are directly related to the provision of web services using outside vendors, and they are highly positively correlated, that is, firms outsourcing, say, hosting web servers tend to outsource web site management and web application development as well. Because potential complementarity between different outsourcing activities is not our focus in this paper, we do not consider them separately.[11]

Second, we define web.inhouse to be an indicator variable for whether a firm uses in-house resources to provide web services. Though the CITDB does not provide the exact variable for this sourcing decision, it still contains various information indicating insourcing. Specifically, we consider the following information: whether the firm has homepage for its Internet application; whether it owns any Internet server; whether it has employees web developers; or whether software related to web services is installed in any server, in which such software includes web server software such as Apache, web database software, homepage software, or web development application software.[12] If the firm reports to have any of these in-house resources for web services, it indicates insourcing web services. However, if web.outsourcing is equal to 1, then regardless of the values of the insourcing proxy variables, the firm does not rely solely on in-house resources for the provision of web services. Given that these in-house resources can be complements to the utilization of outsourced web services, we consider such firms to be outsourcing rather than insourcing. Therefore, we define web.inhouse to be one if the firm has any of in-house resources listed above and does not outsource web services. Hence, web.outsourcing and web.inhouse are mutually exclusive, which is also consistent with our theoretical framework.[13]

For the variables discussed in section 2.4, we use four groups of variables in our data. The first

[11] The CITDB also contains information on outsourcing routers or firewalls in the web outsourcing category, but we do not consider this information because it is not directly related to the provision of web services.

[12] Firms can still have a website homepage even though they do not report having a homepage for their Internet application. Moreover, advanced web services are not limited to providing web sites. Therefore, we also consider multiple sources of information that is indicative of whether the firm provides web services using in-house resources.

[13] Multi-establishment firms, which we do not include in our data, may in fact have substantial in-house web services capabilities while also outsourcing important aspects of such services to vendors. This is not very likely for single establishment firms in our sample. Any in house capabilities of firms that outsource is likely to be of auxiliary nature.

captures the scale of a firm, and includes revenue (in millions of dollar), the number of #employees, and the number of #internet.users within the firm. The second group of variables reflects the degree of business dependence on web services. This includes e-business, an indicator variable for whether Internet application includes B2B or B2C exchanges, and customer.svc, a dummy variable equal to 1 if the Internet application includes customer services. We also use #web.county (or ratio.web.county) which is the number (or fraction) of other establishments within the same county that have web services in the previous year (i.e., those with either web.outsourcing or web.inhouse equal to 1). Note that we compute them by using establishments' decisions in the previous year and we exclude one's own decision. We compute similar variables for the NAICS three-digit industry codes, and they are #web.naics3 and ratio.web.naics3.

The third set of variables is related to transaction costs associated with outsourcing web services and the distance between the attribute desired by the firm and the closest attribute available by outsourcing vendors. These variables include pop.gt.250k, a dummy equal to 1 if the firm is located in the area with the population more than 250,000, and total.router which is total number of routers used in the establishment, potentially capturing the communication costs between the establishment and outside vendors. The CITDB provides information on outsourcing of other services such as programming, application design, helpdesk, system tuning, network managment, WAN design, etc., which is not related to outsourcing web services. We combine them and define other.outsourcing to be one if the firm outsources any of these services. Using this variable for other establishments in the previous year, we also compute #other.out.county (or ratio.other.out.county, #other.out.naics3, ratio.other.out.naics3). Similarly, we obtain #web.out.county (or ratio.web.out.county, #web.out.naics3, ratio.web.out.naics3) for outsourcing web services by other establishments in the same county in the previous year.

The final group of variables is related to the costs associated with developing web services in-house. We use #programmers and #it.employees for the number of programmers and IT employees. We also use total.server and other hardware capabilities. Additionally, we consider several indicator variables for whether the firm installed related software (not for web services), and these variables include helpdesk for technical support, publishing for publishing software, development for application development, database for database, and data.mgmt for data management software.

3.3 Descriptive Statistics

Table 1 presents summary statistics of variables defined in the previous section.[14] In the table, the first group of variables is the key variables on the sourcing decision. The table shows that web.inhouse has increased, while web.outsourcing has decreased over time. It also shows that more firms used in-house resources to provide web services, rather than outsourcing the provision of these services. Though outsourcing would appear to be more attractive because it does not entail high fixed costs associated with the in-house development of web services, it seems to be becoming progressively less attractive in our data. As discussed in our theoretical framework, this dominance of insourcing over outsourcing can be explained by two main factors.

The first factor is firm heterogeneity in various determinants discussed in our theory section, including those related to the demand for web services and fixed costs of in-house development. For example, firms might have adopted more IT and hired programmers for other uses not directly related to the provision of web services, but such investment could indirectly lower the costs associated with developing web services in-house. Moreover, the development of new software may have reduced the inputs required for the development of web services by a firm. Table 1 shows that the variables related to the parameters R and r in our theoretical framework (i.e., the second set of variables such as #internet.users) and those related to the parameter F (i.e., the third group of variables including #it.employees) tend to increase over time, suggesting that more firms are likely to experience decreasing costs of in-house development of web services.

The second factor is related to information spillovers, as discussed in our theoretical framework. As more firms outsource web services development, more firms learn about outsourcing, which could

[14] Note that we use the information from the previous year to compute the variables for the number/fraction of establishments within the same region or industry. For this reason, our empirical analysis examines firms' decisions from 2001 to 2004, and the table does not report summary statistics for 2000.

lower potential search costs or transaction/contracting costs. However, the opposite is also possible if less firms are outsourcing, which could instead increase transaction costs, so that outsourcing might not be viable for most firms. In terms of Proposition 1, the normalized cost of outsourcing could have been higher than the cost of in-house production for most firms. As a result, most firms may not outsource because it is not a viable option for them. In Table 1, the fourth group of variables is related to potential information spillover and transaction costs of outsourcing. The changes in these variables seem to be consistent with the decline in web services outsourcing.

To further investigate the relationship between the sourcing decisions and various variables reflecting different factors, we next present the results from probit regressions of the sourcing decisions on all regressors in Table 1. The first three columns in Table 2 report the results for three choice variables – not using web services, outsourcing, and using in-house resources. Note that these regressions include three groups of regressors as in Table 1, and higher values of the first variables are associated with higher values of R and r, while higher values for the second group (or the third group) are related to lower values of F (or C). Therefore, firms with higher values for these variables are likely to provide web services either in-house or by outsourcing, as discussed in section 2.2. This is consistent with the negative coefficients for most variables in the first column for no.web, and the positive coefficients for many variables in the second (for web.outsource) and the third (for web.inhouse) columns. Also, if F is low, then firms are more likely to use in-house resources, while low C is more likely to lead firms to outsource, which is somewhat consistent with the coefficients in the second and the third columns. Because firms can also choose not to use any web service, however, not all coefficients seem to correspond to our theoretical model. One simple alternative to check is to compare only firms that provide web services either in-house or by outsourcing. Hence, we drop those with no.web $= 1$, and estimate a probit regression of web.inhouse. The results are reported in the fourth column. Overall, many coefficients seem to be consistent with our theoretical discussion in section 2.2.

Nevertheless, these results provide only suggestive evidence on the importance of different factors and do not explain why more firms use in-house resources than outsourcing. To better understand underlying mechanisms that led to the dominance of in-house development over outsourcing, we develop a somewhat more structural empirical model in the next section.

4 Empirical Framework

4.1 Econometric Model

Our theoretical framework provides basic structure for our econometric model. In particular, inequalities in Proposition 1 in section 2.2 lead to threshold crossing models for the choice probabilities of the sourcing decisions. To begin, let us consider three groups of variables: X_{it} is a vector of variables that reflect R, the scale of firm i at period t, and r, the degree of its dependence on web services for its business; Z_{it} is a vector of variables that capture C and Δx, that is, transaction costs associated with outsourcing web services, and the extent to which firms' needs are met by outsourcing vendors; W_{it} is a vector of variables that reflect F, the costs associated with developing web services in-house.

We then consider the inequality that determines two regimes. Let A_{it} denote the indicator for the availability of outsourcing option for firm i at period t. Then, $A_{it} = 1$ if $\frac{C}{1-\Delta x} < F$. Notice that this inequality depends on C, Δx, and F, which are represented by Z_{it} and W_{it}. As a result, we posit the following threshold crossing model for regime switching:

$$\begin{aligned} A_{it} = 1 &\quad \text{if } W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai} + \epsilon_{Ait} > 0, \\ A_{it} = 0 &\quad \text{otherwise,} \end{aligned}$$

where γ_A and δ_A are vectors of parameters to be estimated, η_{Ai} represents fixed effects capturing potential unobserved heterogeneity, and ϵ_{Ai} is assumed to follow the extreme value distribution. Therefore, the probability for regime switching can be written as

$$\Pr(A_{it} = 1 | W_{it}, Z_{it}, \eta_{Ai}) = \frac{\exp(W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai})}{1 + \exp(W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai})}. \tag{10}$$

To rewrite the several inequalities that determine firm's choice, note that the net profits for no web services, outsourcing, and in-house provision are respectively given by

$$\Pi_{\mathsf{no}} = R(1-r), \qquad \Pi_{\mathsf{out}} = R(1-r\Delta x) - C, \qquad \Pi_{\mathsf{inh}} = R - F.$$

These profits can be normalized by subtracting $R(1-r)$, so that Π^*_{no}, the normalized profit for no web service is equal to zero. The normalized profits for other choices are then written as

$$\begin{aligned} \Pi^*_{\mathsf{out}} &= rR(1-\Delta x) - C, \\ \Pi^*_{\mathsf{inh}} &= rR - F. \end{aligned}$$

Using the variables above, we can rewrite these profits as

$$\begin{aligned} \Pi^*_{\mathsf{out}} &= X_{it}\beta_1 + W_{it}\gamma_1 + \eta_{1i} + \epsilon_{1i}, \\ \Pi^*_{\mathsf{inh}} &= X_{it}\beta_2 + Z_{it}\delta_2 + \eta_{2i} + \epsilon_{2i}, \end{aligned}$$

where β_1, γ_1, β_2, and δ_2 are parameters to be estimated; η_{1i} and η_{2i} are fixed effects capturing potential unobserved heterogeneity such as market unobservables that affect firms' decisions. We assume that ϵ_{1i} and ϵ_{2i} follow the extreme value distribution.

Given these profits, the inequalities in Proposition 1 naturally lead to the following discrete choice model. To derive the choice probabilities, let y_{0it} denote indicator dummy for no web services, y_{1it} denote the indicator for outsourcing, and y_{2it} denote the indicator for in-house provision. If $A_{it} = 1$, firm i chooses between outsourcing, in-house provision, and no web service. Noting that these choices are mutually exclusive, the extreme value distribution assumption then implies the standard multinomial logit model, where the choice probability for outsourcing is written as

$$\begin{aligned} \Pr(\Pi^*_{\mathsf{out}} > 0 \text{ and } \Pi^*_{\mathsf{out}} > \Pi^*_{\mathsf{inh}}) &= \Pr(y_{1it} = 1 | A_{it} = 1, X_{it}, Z_{it}, W_{it}, \eta_{1i}, \eta_{2i}) \\ &= \frac{\exp(X_{it}\beta_1 + W_{it}\gamma_1 + \eta_{1i})}{1 + \exp(X_{it}\beta_1 + W_{it}\gamma_1 + \eta_{1i}) + \exp(X_{it}\beta_2 + Z_{it}\delta_2 + \eta_{2i})}. \quad (11) \end{aligned}$$

Similarly, we obtain the choice probability for in-house provision as

$$\begin{aligned}\Pr(\Pi^*_{\mathsf{inh}} > 0 \text{ and } \Pi^*_{\mathsf{inh}} > \Pi^*_{\mathsf{out}}) &= \Pr(y_{2it} = 1 | A_{it} = 1, X_{it}, Z_{it}, W_{it}, \eta_{1i}, \eta_{2i}) \\ &= \frac{\exp(X_{it}\beta_2 + Z_{it}\delta_2 + \eta_{2i})}{1 + \exp(X_{it}\beta_1 + W_{it}\gamma_1 + \eta_{1i}) + \exp(X_{it}\beta_2 + Z_{it}\delta_2 + \eta_{2i})}. \quad (12)\end{aligned}$$

However, if $A_{it} = 0$, outsourcing option is not available, so that firm i makes a binary choice. To allow for flexibility in our estimation, we assume that firm i uses in-house provision if $\Pi^{**}_{\mathsf{inh}} > 0$, where $\Pi^{**}_{\mathsf{inh}} = X_i\beta_3 + W_i\gamma_3 + \eta_{3i} + \epsilon_{3i}$; β_3 and γ_3 are additional parameters to estimate; η_{3i} represents fixed effects; ϵ_{3i} is assumed to follow the extreme value distribution. As a result, the choice probability under this regime is given by

$$\Pr(y_{2it} = 1 | A_{it} = 0, X_{it}, W_{it}, \eta_{3i}) = \frac{\exp(X_{it}\beta_3 + W_{it}\gamma_3 + \eta_{3i})}{1 + \exp(X_{it}\beta_3 + W_{it}\gamma_3 + \eta_{3i})}. \quad (13)$$

The models in (11), (12), and (13), however, cannot be directly estimated because the regime indicator A_{it} is not observed. For this reason, we consider the choice probability unconditional on A_{it}. This probability for in-house provision is then the mixture of (10), (12), and (13) as follows.

$$\begin{aligned}\Pr(y_{2it} = 1 | H_{it}, \eta_i) &= \frac{\exp(X_{it}\beta_3 + W_{it}\gamma_3 + \eta_{3i})}{1 + \exp(X_{it}\beta_3 + W_{it}\gamma_3 + \eta_{3i})} \times \frac{1}{1 + \exp(W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai})} \\ &+ \frac{\exp(X_{it}\beta_2 + Z_{it}\delta_2 + \eta_{2i})}{1 + \exp(X_{it}\beta_1 + W_{it}\gamma_1 + \eta_{1i}) + \exp(X_{it}\beta_2 + Z_{it}\delta_2 + \eta_{2i})} \\ &\quad \times \frac{\exp(W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai})}{1 + \exp(W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai})}, \quad (14)\end{aligned}$$

where $H_{it} = (X_{it}, W_{it}, Z_{it})$ and $\eta_i = (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{Ai})$. The choice probability of no web service can be obtained similarly. As for outsourcing, note that it is not available if $A_{it} = 0$. That is, $\Pr(y_{1it} = 1 | A_{it} = 0, H_{it}, \eta_i) = 0$. Hence, the choice probability of outsourcing is given by

$$\begin{aligned}\Pr(y_{1it} = 1 | H_{it}, \eta_i) &= \frac{\exp(X_{it}\beta_1 + W_{it}\gamma_1 + \eta_{1i})}{1 + \exp(X_{it}\beta_1 + W_{it}\gamma_1 + \eta_{1i}) + \exp(X_{it}\beta_2 + Z_{it}\delta_2 + \eta_{2i})} \\ &\quad \times \frac{\exp(W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai})}{1 + \exp(W_{it}\gamma_A + Z_{it}\delta_A + \eta_{Ai})}. \quad (15)\end{aligned}$$

The model developed above is a version of switching regression models. The difference from conventional approach as in Lee and Porter (1984) and Porter (1983) is that the probability of regime switching is not a constant but a function of variables. This type of switching regression model is not commonly used, but there are a few related models where latent classes or segments are assumed and the probability of segment membership depends on observed characteristics (see Formann (1992) and Gupta and Chintagunta (1994)).

4.2 Estimation

In the absence of η_i, all the parameters of the model in (14) and (15) are identified because of the exclusion restrictions from our theoretical model (e.g. X_{it} does not enter (10)), as well as the logit structure. Moreover, if we ignore η_i, we can estimate the model by the maximum likelihood estimation using the following likelihood function.

$$L = \prod_{t=1}^{T}\prod_{i=1}^{N_t}\left[\Pr(y_{0it}=1|H_{it})^{y_{0it}} \times \Pr(y_{1it}=1|H_{it})^{y_{1it}} \times \Pr(y_{2it}=1|H_{it})^{y_{2it}}\right],$$

where T is total years observed, and N_t is the number of observations in year t.

The presence of η_i, however, complicates the estimation of our model. One approach to address this issue is to impose parametric distribution assumptions on η_i and integrate out η_i. The use of this random effect approach is not uncommon, but it is computationally intensive, thus requiring simulated likelihood approach as in Hong and Wolak (2008). Alternatively, we may not impose any parametric assumptions on η, that is, treat them as fixed effects. One related approach is to estimate the model including dummy variables for each region and each industry. This will account for market-level unobservables which are likely to affect firms' decisions. Given the mixture structure of our model, however, this seemingly simple approach is not practically feasible, because there are too many parameters to estimate.

Consequently, we consider an iterative approach that can be easily implemented, while still allowing us to account for market-level unobservables. Our estimation approach consists of five steps. In the first step, we compute the initial values for $P_A \equiv \Pr(A_{it} = 1|W_{it}, Z_{it}, D_{it})$, where D_{it} is a vector of dummy variables for region and industry. Because we do not observe A_{it}, we replace y_{1it} for A_{it} and estimate a logit model. Using the estimated parameters, we impute $\hat{P}_A^0$ for all observations. In the second step, we assign $A_{it}^0 = 1$ if $y_{1it} = 1$ or $\hat{P}_A^0 \geq q$-th percentile, where $q = 1 - E(\hat{P}_A^0)$; and $A_{it}^0 = 0$ otherwise. We then use only the observations with $A_{it} = 1$ and run the multinomial logit model based on (11) and (12). Likewise, we use only the observations with $A_{it} = 0$ and estimate the binary choice logit model in (13). Even though D_{it} includes many dummy variables, the estimation of these logit models is straightforward.

For the third step, we consider (14) and (15), and rewrite them as

$$\begin{aligned} P_2 &= P_{20} \times \frac{1}{1 + \exp(W_{it}\gamma_A + Z_{it}\delta_A)} + P_{21} \times \frac{\exp(W_{it}\gamma_A + Z_{it}\delta_A)}{1 + \exp(W_{it}\gamma_A + Z_{it}\delta_A)}, \\ P_1 &= P_{11} \times \frac{\exp(W_{it}\gamma_A + Z_{it}\delta_A)}{1 + \exp(W_{it}\gamma_A + Z_{it}\delta_A)}, \end{aligned}$$

where $P_2 \equiv \Pr(y_{2it} = 1|H_{it}, D_{it})$ unconditional on A_{it}, estimated by the conventional discrete choice model, and similarly, $P_1 \equiv \Pr(y_{1it} = 1|H_{it}, D_{it})$; $P_{20} \equiv \Pr(y_{2it}|A_{it} = 0, H_{it}, D_{it})$, estimated from the binary choice logit model, and $P_{21} \equiv \Pr(y_{2it}|A_{it} = 1, H_{it}, D_{it})$, estimated from the multinomial logit model in the second step; and $P_{11} \equiv \Pr(y_{1it}|A_{it} = 1, H_{it}, D_{it})$ from the multinomial logit model. Because we can impute $\hat{P}_2$, $\hat{P}_1$, $\hat{P}_{20}$, $\hat{P}_{21}$, and $\hat{P}_{11}$ for all observations, we can estimate the coefficients γ_A and δ_A by using nonlinear least squares of the equations above.

In the fourth step, we use the estimates for γ_A and δ_A from the previous step, and then impute the regime switching probability $\hat{P}_A^1 = \exp(W_{it}\hat{\gamma}_A + Z_{it}\hat{\delta}_A)/(1 + \exp(W_{it}\hat{\gamma}_A + Z_{it}\hat{\delta}_A))$. Using $\hat{P}_A^1$, we follow the same procedure as in the second step by assigning $A_{it}^1 = 1$ if $y_{1it} = 1$ or $\hat{P}_A^1 \geq q$-th percentile, where $q = 1 - E(\hat{P}_A^1)$, and estimate the multinomial logit model for those with $A_{it}^1 = 1$ and the binary choice logit model for those with $A_{it}^1 = 0$. In the final step, we repeat the third and the fourth steps until the coefficient estimates for our model in (14) and (15) are converged. The proposed iterative approach uses only straightforward estimation methods, while still allowing us to control for market-level unobservables.

5 Results

5.1 Model Estimates

We estimate the econometric model in (14) and (15), using the iterative estimation approach in section 4.2. The estimation results are presented in Tables 3-5. The coefficient estimates are converged after 20 iterations. The convergence of our estimates is partly shown in Table 6, where we report the mean of the regime switching probability in each iteration.

Using the imputed regime switching probability and the cutoff rule discussed in the previous section, we separate our sample into two groups – the first group is likely to be in the regime where outsourcing is viable, and the second is not. Table 3 presents the results from the multinomial logit estimation using the first group, and Table 4 reports the estimation results for the binary choice logit model using the second group. Most coefficients are estimated precisely. A few coefficient estimates related to C and Δx do not seem to be in accordance with our theoretical model. Ratio.other.out.county and ratio.web.out.naics3 are likely to be negatively correlated with transaction costs, so that the coefficients for these variables are expected to be positive, but our estimates are negative.[15] Nevertheless, most coefficient estimates for variables related to R, r, and F are consistent with our theoretical framework.

Table 5 presents the results from nonlinear least squares for the regime switching probability. All the coefficients are estimated precisely. Section 2.2 shows that for outsourcing to be a viable option, it must be that $\frac{C}{1-\Delta x} < F$. That is, $A_{it} = 1$ if $F - \frac{C}{1-\Delta x} > 0$. Therefore, the coefficient estimates for the variables that reduce F should be negative, while those for the variables that decrease C should be positive. This is in accordance with the estimates for #programmer, #it.employees, total.server, pop.gt.250k, total.router, other.outsourcing, ratio.other.out.county, and ratio.web.out.naics3. However, the coefficients for the variables associated with related software such as publishing software and database software have the opposite sign, suggesting that either

[15]Because the coefficient estimates for these variables in the regime switching equation have expected signs in Table 5, while those for ratio.web.out.county and ratio.other.out.naics3 do not, there seems to be some issues related to identification. We will investigate this issue in more detail in the next version of this paper.

these variables do not necessarily reduce F or we need further investigation into potential issues.

Table 6 presents the summary statistics of the regime switching probabilities P_A^k which we impute for all observations in each iteration k. Table 6 also shows that the mean of the converged regime switching probability is approximately 0.7. In other words, outsourcing is a viable option for 70% of our sample, while it is not practically available for 30% of our sample. Therefore, for at least 70% of firms in our sample, firm heterogeneity, rather than information spillover, is likely to account for the dominance of in-house provision over outsourcing, suggesting the potential importance of firm heterogeneity to explain the sourcing decision in the provision of web services. In the next section, we perform counterfactual exercises to further examine related implications of our model estimates.

5.2 Counterfactual Results

We begin by considering an extreme scenario for our counterfactual exercise, in which we assume that outsourcing is a viable option for all firms. That is, we set $A_{it} = 1$ for all observations, and use the estimated multinomial logit model to compute the counterfactual choice probabilities $\hat{P}_{j1}$, where $P_{j1} = \Pr(y_j = 1|A = 1, H)$ for $j = 0, 1, 2$, repectively denoting no web service, outsourcing, and in-house provision. We calculate these choice probabilities for all observations, and Table 7 reports their mean values. For comparison, we also report the actual fraction of firms for each decision, and the mean values of the predicted choice probabilities $\hat{P}_j = \hat{\Pr}(y_j = 1|H)$.

Column 1 in the table presents the choice probabilities for all years. We first note that the predicted $\hat{P}_j$ is somewhat lower than the actual fraction of firms for no web service and in-house provision, whereas it is rather higher than the actual fraction for outsourcing.[16] Because our counterfactual is based on the estimated model, we use the predicted choice probability as the benchmark. The counterfactual results indicate that if outsourcing is viable for all firms, then 13% more firms (= 0.411 - 0.281) will outsource – 6.8% firms that used to use no web service, and 6.2% firms that used to rely on in-house resources. This implies that firms do not use any web service

[16] This issue suggests that we need further specification check. We will address it in the next version of this paper.

or outsourcing partly because outsourcing is not practically available for them.

Though the preceding result appears to show the importance of information spillover, there are two issues. First, information spillover may be an important factor, but the viability of outsourcing also depends on other factors, as $A_{it} = 1$ if $F - \frac{C}{1-\Delta x} > 0$. For example, relatively high F can also make outsourcing viable, and likely lead firms to outsource rather than use in-house resources. As a result, we perform further counterfactuals focusing on potential spillover below and present the results in Table 8. The second issue is that if information spillover is indeed the main factor and lack of information causes the bottleneck in using outsourcing, then disseminating more information on outsourcing would lead more firms to outsource. Because information is likely to be diffused more over time, we can check this possibility by examining changes in the counterfactual choice probability over time. Columns 2-5 in Table 7 present the results. The counterfactual choice probability of outsourcing $\hat{P}_{11}$, however, is declining from 2001 to 2004. Although there is a decreasing trend in outsourcing even in the benchmark $\hat{P}_1$, the decrease in $\hat{P}_{11}$ is steeper than that in $\hat{P}_1$. This is inconsistent with the information spillover story, suggesting that other factors, particularly changes in the degree of firm heterogeneity, might be more important in explaining the increasing dominance of in-house provision.

To further investigate the effect of information spillover, we consider the variables that are likely to reflect information spillover. Because the fraction of firms outsourcing in the same region (or industry) in the previous year would be positively correlated with potential information spillover, we focus on the fractions of firms using any web service (ratio.web.county and ratio.web.naics3), web service outsourcing (ratio.web.out.county and ratio.web.out.naics3), and outsourcing other IT services (ratio.other.out.county and ratio.other.out.naics3) for each county and industry. Our counterfactual is to increase each of these variables by 10% and examine whether it would increase the counterfactual $\hat{P}_1$. Rows 1-6 in Table 8 present the results. Except for ratio.web.out.county and ratio.other.out.naics3, however, the increase in the fraction of firms outsourcing actually decreases the counterfactual $\hat{P}_1$. Though the increase in ratio.web.out.county and ratio.other.out.naics3 increases $\hat{P}_1$, it reduces the regime switching probability.[17] The results in Table 8 therefore imply that information spillover or lack of information is less likely to explain why more firms use in-house resources for their web services and less firms have outsourced web services over time.

Outsourcing might seem appealing because it does not entail significant fixed costs associated with in-house provision, and this attractive feature might lead firms to outsource, rather than insource, particularly at the early stage of web service development. Nevertheless, information on outsourcing is likely to be unknown to most firms at the early stage. Hence, even though many firms could have been better off by outsourcing, they might not outsource because of lack of information, thus implying potential inefficiency due to the information bottlenecks. The overall results from our counterfactual exercises, however, suggest that firm heterogeneity, as opposed to information spillover, is more likely to account for the dominance of in-house provision over outsourcing in the web service sourcing decision. Firms may not outsource, but that is not because of lack of information on outsourcing, but rather because of other factors such as their strong preference for customization.

6 Conclusion

In this paper, we study the role of firm heterogeneity and information spillover in the sourcing decision to provide web services. To this end, we develop our theoretical model that relates these factors to firms' decisions to insource or outsource. We then construct and estimate our econometric model to examine why more firms insource than outsource. We find that firm heterogeneity, rather than information spillover, is more likely to account for the dominance of insourcing over outsourcing, which suggests that potential inefficiency due to the information bottlenecks is unlikely. Nevertheless, more investigation into the effect of information spillover would be useful, and we will further examine the underlying mechanisms behind firm heterogeneity in our future research.

[17] This pattern seems to be related to the issue discussed earlier – the opposite signs of the coefficients for these variables in the multinomial logit equation and the regime switching equation.

References

Arora, A. and Forman, C. "Proximity and Information Technology Outsourcing: How Local Are IT Services Markets?" *Journal of Management Information Systems*, 2007, pp. 73-102.

Augereau, A., Greenstein, S., and Rysman, M. "Coordination versus Differentiation in a Standard War: 56K Modems." *RAND Journal of Economics*, 2006, pp. 887-909.

Baker, G. and Hubbard, T. "Make versus Buy in Trucking: Asset Ownership, Job Design, and Information." *American Economic Review*, 2003, pp. 551-572.

Bresnahan, T., Brynjolfsson, E., and Hitt, L. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics*, 2002, *117*, pp. 339-376.

Bresnahan, T. and Greenstein, S. "Technical Progress and Co-invention in Computing and in the Uses of Computers." *Brookings Papers on Economic Activity: Microeconomics*, 1996, pp. 1-77.

Forman, C., Goldfarb, A., and Greenstein, S. "How Did Location Affect Adoption of the Commercial Internet? Global Village vs. Urban Leadership." *Journal of Urban Economics*, 2005, pp. 389-420.

Forman, C., Goldfarb, A., and Greenstein, S. "Understanding the Inputs into Innovation: Do Cities Substitute for Internal Firm Resources?" *Journal of Economics & Management Strategy*, 2008, pp. 295-316.

Formann, A. "Linear Logistic Latent Class Analysis for Polytomous Data." *Journal of the American Statistical Association*, 1992, pp. 476-486.

Gupta, S., and Chintagunta, P. "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models." *Journal of Marketing Research*, 1994, pp. 128-136.

Hong, S. and Rezende, L. "Lock-in and Unobserved Preferences in Server Operating Systems Adoption: A Case of Linux vs. Windows." 2008, manuscript, University of Illinois.

Hong, S. and Wolak, F. "Relative Prices and Electronic Substitution: Changes in Household-level Demand for Postal Delivery Services from 1986 to 2004." *Journal of Econometrics*, 2008, pp. 226-242.

Hubbard, T. "Contractual Form and Market Thickness in Trucking." *RAND Journal of Economics*, 2001, pp. 369-386.

Lee, L. and Porter, R. "Switching Regression Models with Imperfect Sample Separation Information – with an Application on Cartel Stability." *Econometrica*, 1984, pp. 391-418.

Porter, R. "A Study of Cartel Stability: The Joint Executive Committee 1880-1886." *Bell Journal of Economics*, 1983, pp. 301-314.

Stiroh, K. "Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say?" *American Economic Review*, 2002, pp. 1559-1576.

Stoneman, P. *The Economics of Technological Diffusion*, Wiley-Blackwell, 2002.

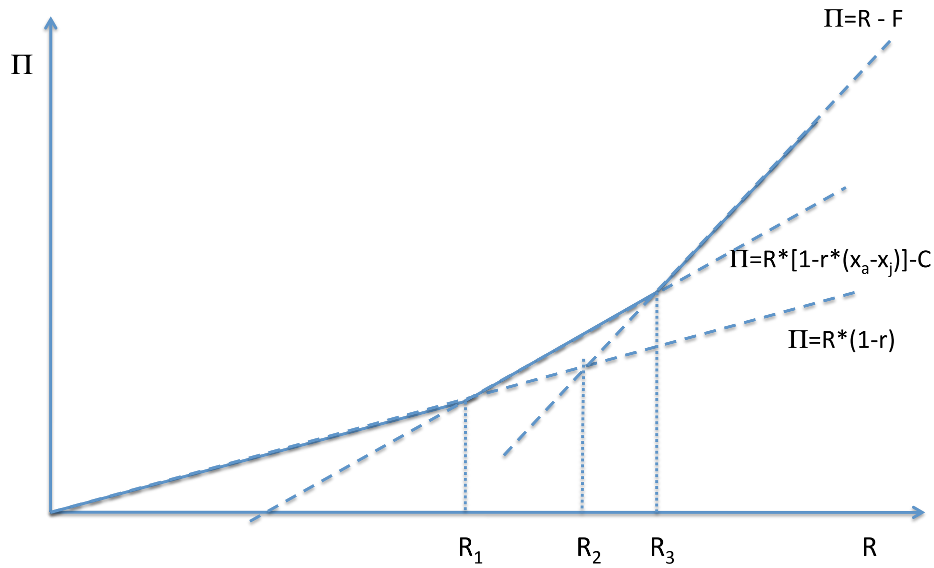

Figure 1: Profit functions for each web service sourcing decision.

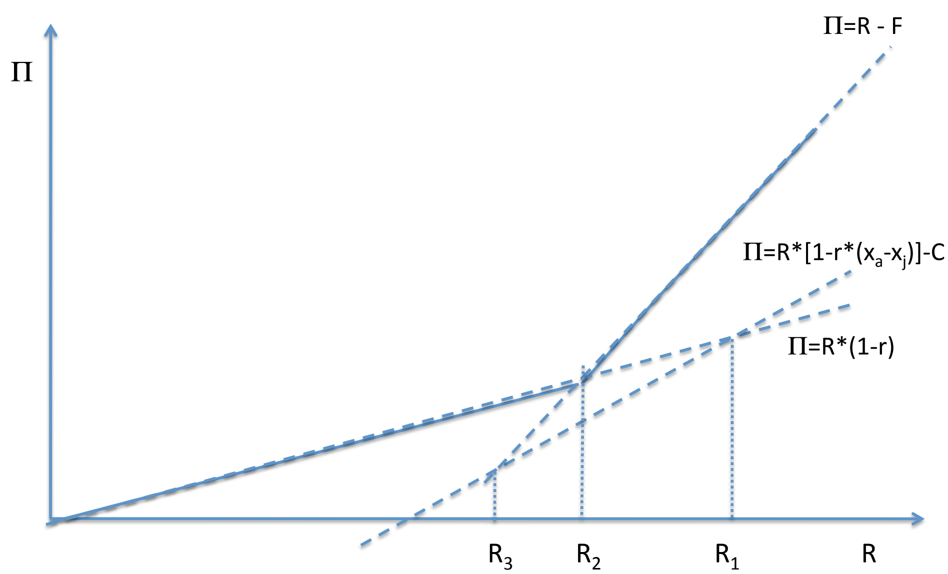

Figure 2: Profit functions for each decision when outsourcing is not viable.

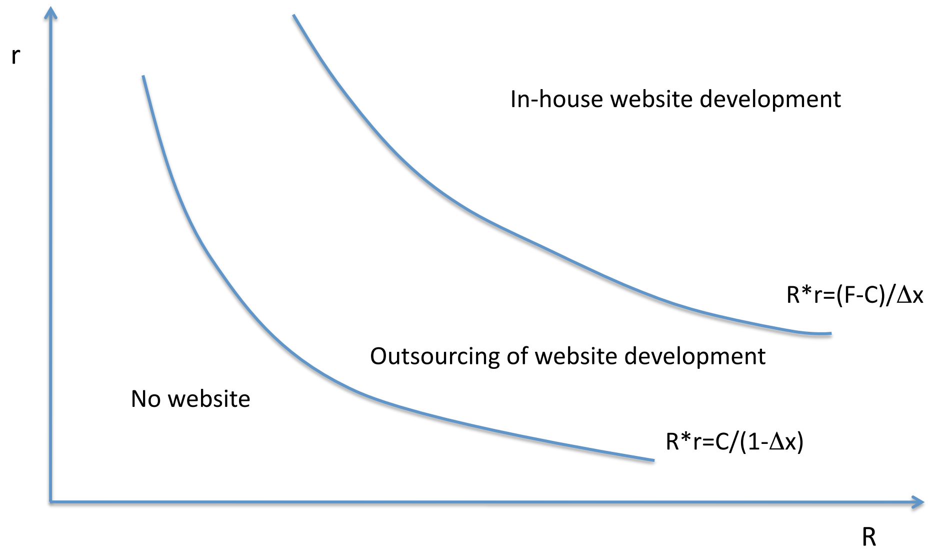

Figure 3: Outsourcing is a viable option (case 1).

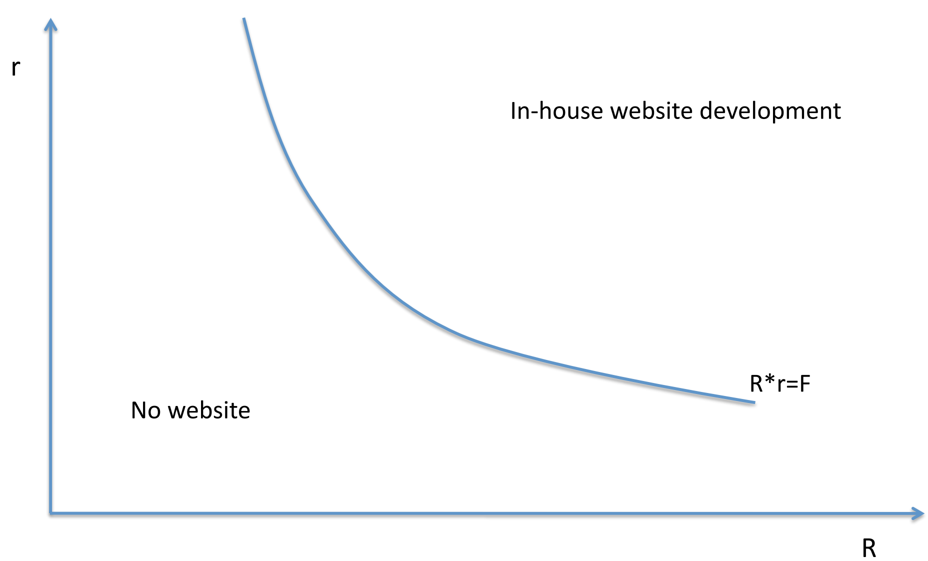

Figure 4: Outsourcing is not a viable option (case 2).

Table 1: Summary Statistics of Key Variables[a]

Year	2001	2002	2003	2004
	(1)	(2)	(3)	(4)
web.inhouse	0.39	0.41	0.46	0.50
web.outsourcing	0.24	0.23	0.18	0.12
no.web	0.38	0.35	0.36	0.39
revenue (in $million)	33.09	35.78	33.99	28.38
#employees	213.73	229.17	237.31	244.35
#internet.users	62.50	67.10	73.06	84.32
e-business	0.11	0.12	0.12	0.10
customer.svc	0.05	0.08	0.11	0.12
#web.county	227.8	260.4	261.9	262.1
ratio.web.county	0.45	0.49	0.53	0.53
#web.naics3	1,970	2,241	2,411	2,680
ratio.web.naics3	0.47	0.52	0.56	0.57
#programmers	2.07	1.92	1.78	2.01
#it.employees	5.10	5.22	5.41	7.75
total.server	4.20	4.59	5.08	5.25
total.lan.server	3.64	3.82	4.40	4.69
helpdesk	0.11	0.14	0.16	0.17
publishing	0.02	0.03	0.04	0.05
development	0.04	0.03	0.03	0.02
database	0.01	0.02	0.03	0.02
data.mgmt	0.16	0.18	0.18	0.17
firewall	0.02	0.02	0.02	0.02
pop.gt.250k	0.71	0.72	0.71	0.71
total.router	1.93	1.99	2.17	2.22
other.outsourcing	0.30	0.28	0.27	0.25
#web.out.county	68.1	86.9	85.9	74.6
ratio.web.out.county	0.13	0.16	0.17	0.15
#other.out.county	130.9	132.8	119.7	112.6
ratio.other.out.county	0.25	0.24	0.23	0.22
#web.out.naics3	468.2	589.7	621.2	571.1
ratio.web.out.naics3	0.14	0.16	0.18	0.15
#other.out.naics3	1,009	1,005	982	1,056
ratio.other.out.naics3	0.27	0.25	0.25	0.24
#observations	34,715	41,287	34,453	29,726

[a]The table reports the mean of each variable. All variables are defined in section 3.2, except for no.web which is equal to 1 if both web.inhouse and web.outsourcing are zero.

Table 2: Probit Regression Results[a]

	Full Sample (dependent variable below)						Exclude observations w/no.web = 1	
	no.web		web.outsource		web.inhouse		web.inhouse	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
Variable	(1)		(2)		(3)		(4)	
revenue (in $million)	-0.001 (0.000)	0.00	0.000 (0.000)	0.11	0.000 (0.000)	0.00	0.000 (0.000)	0.00
#internet.users	-0.002 (0.000)	0.00	0.000 (0.000)	0.00	0.001 (0.000)	0.00	0.001 (0.000)	0.00
e-business	-0.563 (0.014)	0.00	0.263 (0.012)	0.00	0.208 (0.011)	0.00	-0.082 (0.013)	0.00
customer.svc	-0.253 (0.016)	0.00	0.057 (0.015)	0.00	0.151 (0.014)	0.00	0.014 (0.016)	0.39
ratio.web.county	-0.401 (0.046)	0.00	0.079 (0.053)	0.14	0.385 (0.045)	0.00	0.198 (0.063)	0.00
ratio.web.naics3	2.056 (0.388)	0.00	-0.964 (0.424)	0.02	-0.714 (0.375)	0.06	0.061 (0.492)	0.90
#programmers	-0.033 (0.001)	0.00	-0.002 (0.000)	0.00	0.009 (0.001)	0.00	0.004 (0.001)	0.00
#it.employees	-0.001 (0.000)	0.00	-0.001 (0.000)	0.00	0.002 (0.000)	0.00	0.001 (0.000)	0.00
total.server	-0.006 (0.000)	0.00	-0.001 (0.000)	0.00	0.003 (0.000)	0.00	0.002 (0.000)	0.00
helpdesk	-0.288 (0.012)	0.00	0.093 (0.012)	0.00	0.180 (0.011)	0.00	0.011 (0.013)	0.40
publishing	-0.578 (0.028)	0.00	-0.046 (0.024)	0.05	0.382 (0.021)	0.00	0.157 (0.025)	0.00
development	-0.384 (0.027)	0.00	-0.044 (0.023)	0.06	0.314 (0.021)	0.00	0.166 (0.024)	0.00
database	-0.462 (0.031)	0.00	0.123 (0.027)	0.00	0.249 (0.025)	0.00	0.004 (0.029)	0.88
data.mgmt	-0.325 (0.011)	0.00	0.033 (0.011)	0.00	0.259 (0.009)	0.00	0.095 (0.012)	0.00
firewall	-0.516 (0.032)	0.00	0.058 (0.027)	0.04	0.316 (0.025)	0.00	0.084 (0.029)	0.00
pop.gt.250k	-0.020 (0.010)	0.04	0.034 (0.011)	0.00	0.003 (0.009)	0.76	-0.019 (0.012)	0.12
total.router	-0.002 (0.000)	0.00	0.000 (0.000)	0.83	0.001 (0.000)	0.00	0.000 (0.000)	0.35
other.outsourcing	-0.365 (0.009)	0.00	0.595 (0.009)	0.00	-0.141 (0.008)	0.00	-0.534 (0.010)	0.00
ratio.web.out.county	-0.018 (0.076)	0.82	0.299 (0.086)	0.00	-0.195 (0.075)	0.01	-0.407 (0.102)	0.00
ratio.other.out.county	-0.098 (0.059)	0.10	0.070 (0.067)	0.29	0.067 (0.058)	0.25	0.002 (0.079)	0.98
ratio.web.out.naics3	-0.762 (0.421)	0.07	-1.165 (0.465)	0.01	2.127 (0.408)	0.00	2.327 (0.540)	0.00
ratio.other.out.naics3	-0.813 (0.361)	0.02	1.417 (0.408)	0.00	-0.771 (0.352)	0.03	-0.721 (0.476)	0.13
#observations	140,181		140,181		140,181		88,525	

[a]The value in the parenthesis is standard error. The table reports the results from Probit estimation using different dependent variables. All estimations include dummies for year, region, and industry, but their estimates are suppressed. The full sample for these estimations does not include outdated observations and multi-establishment firms.

Table 3: Iterative Estimation Results – Multinomial Logit for $A_{it} = 1$[a]

	web.outsource			web.inhouse		
	Estimate	S.E.	p-value	Estimate	S.E.	p-value
Variable	(1)			(2)		
revenue (in $million)	0.0014	0.0002	0.00	0.0014	0.0002	0.00
#internet.users	0.0036	0.0002	0.00	0.0043	0.0002	0.00
e-business	1.1409	0.0327	0.00	0.8759	0.0330	0.00
customer.svc	0.4839	0.0361	0.00	0.4857	0.0362	0.00
ratio.web.county	-0.5565	0.1092	0.00	0.5515	0.0988	0.00
ratio.web.naics3	-0.7029	0.0975	0.00	2.4195	0.0969	0.00
#programmers				0.0131	0.0016	0.00
#it.employees				0.0021	0.0006	0.00
total.server				0.0087	0.0008	0.00
helpdesk				0.3278	0.0211	0.00
publishing				0.6816	0.0414	0.00
development				0.7753	0.0379	0.00
database				0.5813	0.0449	0.00
data.mgmt				0.4777	0.0188	0.00
firewall				0.5483	0.0468	0.00
pop.gt.250k	0.0297	0.0181	0.10			
total.router	0.0003	0.0004	0.44			
other.outsourcing	-0.4934	0.0171	0.00			
ratio.web.out.county	1.3367	0.1562	0.00			
ratio.other.out.county	-1.1787	0.1207	0.00			
ratio.web.out.naics3	-7.4532	0.1805	0.00			
ratio.other.out.naics3	3.4851	0.1889	0.00			

[a]Our estimation converged after twenty iterations. The table reports the estimates from the multinomial logit estimation in the converged iteration.

Table 4: Iterative Estimation Results – Logit for $A_{it} = 0$[a]

	web.inhouse		
Variable	Estimate	S.E.	p-value
revenue (in $million)	0.0007	0.0002	0.00
#internet.users	0.0033	0.0001	0.00
e-business	1.0181	0.0393	0.00
customer.svc	0.5033	0.0423	0.00
ratio.web.county	0.6757	0.0803	0.00
ratio.web.naics3	2.9776	0.0754	0.00
#programmers	0.1415	0.0056	0.00
#it.employees	0.0021	0.0007	0.00
total.server	0.0318	0.0019	0.00
helpdesk	0.5846	0.0349	0.00
publishing	1.0639	0.0830	0.00
development	0.5759	0.1114	0.00
database	0.6973	0.1117	0.00
data.mgmt	0.6231	0.0307	0.00
firewall	1.0868	0.1116	0.00

[a]The table reports the estimates from the logit estimation in the converged iteration.

Table 5: Iterative Estimation Results – Regime Switching Probability[a]

Variable	Estimate	S.E.	p-value
#programmers	-0.0033	0.0002	0.00
#it.employees	-0.0003	0.0001	0.00
total.server	-0.0028	0.0001	0.00
helpdesk	0.0868	0.0031	0.00
publishing	0.0566	0.0083	0.00
development	0.1813	0.0097	0.00
database	0.1649	0.0100	0.00
data.mgmt	0.0969	0.0029	0.00
firewall	0.1073	0.0103	0.00
pop.gt.250k	0.0110	0.0018	0.00
total.router	0.0002	0.0001	0.02
other.outsourcing	1.2357	0.0065	0.00
ratio.web.out.county	-0.0291	0.0125	0.02
ratio.other.out.county	0.6846	0.0106	0.00
ratio.web.out.naics3	3.2459	0.0202	0.00
ratio.other.out.naics3	-0.6378	0.0176	0.00

[a]The table reports the estimates from the nonlinear LS in the converged iteration.

Table 6: Summary Statistics of Imputed P_A

P_A^k	Mean	S.D.	Min	Max
P_A^0	0.1968	0.0975	0.0000	0.6733
P_A^1	0.5605	0.0435	0.4035	0.6914
P_A^2	0.6859	0.1133	0.4348	0.9465
P_A^3	0.6866	0.1134	0.4208	0.9475
$\vdots$			$\vdots$	
P_A^{15}	0.6998	0.1073	0.0013	0.9415
P_A^{16}	0.7025	0.1050	0.0019	0.9398
P_A^{17}	0.7027	0.1049	0.0018	0.9398
P_A^{18}	0.7029	0.1050	0.0017	0.9396
P_A^{19}	0.7029	0.1050	0.0017	0.9399
P_A^{20}	0.7029	0.1050	0.0017	0.9399

Table 7: Counterfactual Choice Probabilities Under $A_{it} = 1$ for All Observations[a]

		By Year			
	All years	2001	2002	2003	2004
	(1)	(2)	(3)	(4)	(5)
		A. No Web Services			
no.web	0.368	0.376	0.355	0.359	0.389
predicted $\hat{P}_0$	0.319	0.337	0.326	0.310	0.300
counterfactual $\hat{P}_{01}$	0.251	0.247	0.259	0.255	0.239
		B. Outsourcing			
web.outsourcing	0.197	0.237	0.235	0.181	0.116
predicted $\hat{P}_1$	0.281	0.327	0.282	0.251	0.261
counterfactual $\hat{P}_{11}$	0.411	0.482	0.408	0.362	0.388
		C. In-house Provision			
web.inhouse	0.435	0.387	0.410	0.460	0.495
predicted $\hat{P}_2$	0.400	0.336	0.392	0.439	0.439
counterfactual $\hat{P}_{21}$	0.338	0.271	0.333	0.382	0.372

[a] The table reports the mean of the choice probabilities. Each panel contains three rows that report actual fractions of firms, the predicted choice probability $\hat{P}_j$ and the counterfactual choice probability $\hat{P}_{j1}$, where $\hat{P}_j = \hat{\Pr}(y_j = 1|H)$ and $\hat{P}_{j1} = \hat{\Pr}(y_j = 1|A = 1, H)$ for $j = 0, 1, 2$ respectively denoting no web service, outsourcing, and in-house.

Table 8: Counterfactual Choice Probabilities Under the Increased Spillover [a]

	no.web	web.outsourcing	web.inhouse	Switching Prob.
	(1)	(2)	(3)	(4)
	A. Benchmark Probability			
	0.319	0.281	0.400	0.703
	B. Counterfactual Probability			
(1) ratio.web.county + 0.10	0.316	0.268	0.416	0.703
(2) ratio.web.naics3 + 0.10	0.292	0.249	0.459	0.703
(3) ratio.web.out.county + 0.10	0.310	0.302	0.388	0.702
(4) ratio.other.out.county + 0.10	0.325	0.268	0.407	0.716
(5) ratio.web.out.naics3 + 0.10	0.358	0.192	0.450	0.763
(6) ratio.other.out.naics3 + 0.10	0.298	0.330	0.372	0.690
(7) all ratios increased by 0.10	0.300	0.201	0.499	0.763

[a] The table reports the mean of the probability $\hat{P}_j = \hat{\Pr}(y_j = 1|H)$ for $j = 0, 1, 2$ respectively denoting no web service, outsourcing, and in-house. The benchmark probability is the choice probability predicted from the estimated model. The counterfactual probability is the predicted choice probability when the ratio variable is increased by 10%.