United Nations University

WIDER

World Institute for Development Economics Research

Discussion Paper No. 2002/89

The Corporate Digital Divide

Determinants of Internet Adoption

Christopher Forman *

October 2002

Abstract

This paper shows how organizational, technical, and environmental factors affected firm decisions to adopt Internet technologies during the early years of the commercialization of the Internet. Organizations that had made prior investments in client/server networks had a higher likelihood of Internet adoption, however investments in proprietary or platform-specific client/server technologies raised the cost of switching from legacy systems. Small firms and those that were geographically concentrated were less likely to adopt. The study shows that organizations commonly adopted access and intranet technologies together, and suggests that low adaptation costs characterized the rapid diffusion of these early Internet technologies.

Keywords: technological change, diffusion, Internet, adoption

JEL classification: O30, O31, O33

* Kellogg School of Management, Northwestern University; email c-forman@kellogg.nwu.edu

This is a revised version of the paper originally prepared for the UNU/WIDER Conference on the New Economy in Development, 10-11 May 2002, Helsinki.

Acknowledgements

I am very grateful to Shane Greenstein for his guidance and support. I thank David Besanko, Ranjay Gulati, Mike Mazzeo, and seminar participants at Northwestern University for many helpful comments and suggestions. I would also like to thank the General Motors Strategy Center and the Kellogg School of Management for generous financial support. All errors are my own.

Tables and figures appear at the end of the paper.

UNU World Institute for Development Economics Research (UNU/WIDER) was established by the United Nations University as its first research and training centre and started work in Helsinki, Finland in 1985. The purpose of the Institute is to undertake applied research and policy analysis on structural changes affecting the developing and transitional economies, to provide a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and to promote capacity strengthening and training in the field of economic and social policy making. Its work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

UNU World Institute for Development Economics Research (UNU/WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Camera-ready typescript prepared by Liisa Roponen at UNU/WIDER
Printed at UNU/WIDER, Helsinki

ISSN 1609-5774
ISBN 92-9190-302-7 (printed publication)
ISBN 92-9190-303-5 (internet publication)

1 Introduction

This study examines organization decisions to adopt the Internet using a large sample concentrated in the finance and services sectors. I find that 75 per cent of the organizations had adopted basic Internet access and 28 per cent had adopted e-commerce in 1998. The large percentages of non-adopters are surprising, given all the promise and press associated with the Internet. Why did so many firms choose not to adopt this new technology? This paper is a first step toward answering this question. Organization responses to the Internet are studied by examining the adoption of three different technologies: basic access, intranet, and a group of more advanced e-commerce applications that enables communication between a firm, its suppliers, and its customers. The analysis shows how organizational, technical, and environmental factors affected firm decisions on whether or not to adopt the Internet.

To predict whether organizations adopt the Internet, I develop a set of theories detailing the causes of variation in the costs and benefits of adoption. I estimate a trivariate probit model of organizational decisions to adopt access, intranet, and e-commerce technologies, and use the results to identify whether the empirical evidence is consistent with each hypothesis. Count data models (Cameron and Trivedi 1997) are employed to identify the factors determining the number of Internet applications adopted. The analysis is conducted at the organization level on a sample of over 8000 organizations surveyed by Harte Hanks Market Intelligence, concentrated primarily in the FIRE (finance, insurance, and real estate) and services sectors.

In the first set of results, I analyse the competing effects of prior investments in client/server (C/S) technologies on the costs and benefits of adoption. Many Internet technologies are an extension of the C/S computing platform, implying organizations that had previously invested in C/S will have lower technical costs of adoption. Prior investment in C/S may also signal high marginal benefit to decentralized computing within the organization, and that the organization may have made complementary investments in organizational design. As a result of these effects, I find that investments in C/S increase the probability of basic access and intranet adoption.

The impact of prior investments in C/S on the probability of access and intranet adoption is unambiguously positive. However, the effects of marginal investments in C/S technologies that are proprietary or platform-specific are not unambiguous. Such investments can enable adoption because they suggest an increase in technical sophistication and even greater familiarity with C/S technologies within the organization. They can also be a barrier to adoption if users have made complementary investments in the installed base that are not transferable to an Internet-based platform. On net, I find that the combined impact of these two effects was mixed, and depended on the complexity of the installed base of C/S technologies. For many organizations with complex C/S networks, the lock-in effects outweighed the effects of technical sophistication, decreasing the likelihood of adoption.

The geographic concentration of employees also affected the probability of adoption. I argue that smaller, less geographically dispersed organizations benefited less from reductions in variable communications costs created by the adoption of early Internet technologies. The analysis shows that organizations that were larger and had multiple establishments were more likely to adopt. In contrast, many organizations that were small and had only one establishment did not even adopt basic access. Conditional on firm size and the existence of multiple establishments, organizations that were more geographically concentrated or located in urban areas also had a lower likelihood of adoption.

Last, I study the manner with which organizations adopt Internet technologies. I find that organizations that did adopt the Internet tended to adopt groups of multiple Internet technologies, rather than adopting individual technologies in isolation. In particular, I find strong evidence that organizations tended to adopt access and intranet together: over 70 per cent of access adopters had also adopted intranet in 1998. However, many adopters of access and intranet chose not to adopt e-commerce. This reinforces the view that in the early years of its diffusion, the Internet was used primarily as a basic communications technology. Most organizations were using the Internet for basic research on the World Wide Web or for e-mail, rather than for more advanced commerce activities.

I also find evidence that organizations adopting multiple technologies were more likely to adopt them simultaneously rather than sequentially. I find that in 1998 a typical organization that had previously adopted basic access will adopt on average 16 per cent fewer applications than an otherwise identical organization that was adopting the Internet for the first time. I argue that this pattern is due in part to the low costs of adapting early Internet technologies—such as e-mail or World Wide Web access—to organizational needs.

This study views the set of communications protocols and technologies that defined the Internet as a general purpose technology (GPT).[1] Some organizations were well positioned to take advantage of this new GPT; others were not. The rate of diffusion of GPTs is often driven by the user costs of adapting new technologies to organizational needs, termed co-invention by Bresnahan and Trajtenberg (1995). Bresnahan and Greenstein (1997) show that high co-invention costs were responsible for the slow diffusion of C/S networking technologies. In contrast, in this study I will argue that the reverse occurred in the early stages of Internet diffusion: low co-invention costs were responsible for rapid diffusion.

Recent papers by Brynjolfsson and Hitt have found a complementary relationship between IT investments and decentralization in organizational design (Hitt and Brynjolfsson 1997; Brynjolfsson and Hitt 1998). Brynjolfsson and Hitt (1998) find that decentralized organizations have stronger demand for IT, and argue such decentralized, IT-intensive organizations will experience the greatest benefits from technological improvements in IT. In contrast, in the present study I find an ambiguous relationship between investments in decentralization and the demand for the Internet. Organizations

1 Bresnahan and Trajtenberg (1995) define GPTs as key technologies that are characterized by their ‘persuasiveness, inherent potential for technical improvements, and innovational complementarities’.

that had already invested in decentralized C/S computing systems were more likely to be early adopters of the Internet. However, organizations that made heavy proprietary or platform-specific investments in C/S were poorly positioned to take advantage of new Internet technologies.

There is a vast literature on the diffusion and adoption of new technologies (e.g., Rogers 1995). This paper fits within the diffusion literature that examines the roles of heterogeneous firm incentives and environmental conditions in the adoption of new innovations. As in those papers, I distinguish between the roles played by market conditions (e.g., Hannan and McDowell 1984), geographic factors (Griliches 1957), and hypotheses related to internal firm features.

Finally, this study builds on recent attempts to understand variation in supplier strategies in on-line markets. Prior work has examined the incentive effects of vertical integration and channel conflict on firm decisions to make products available on-line (Carlton and Chevalier 2001; Gertner and Stillman 2000). Several papers have also examined on-line pricing strategies (e.g., Brynjolfsson and Smith 2000; Scott Morton, Zettelmeyer, and Risso 2000; Brown and Goolsbe 2002). This paper extends this prior work by examining how internal organization characteristics influenced the adoption of enabling Internet technologies.

2 Early patterns of Internet adoption

Because of its non-commercial origins, many Internet technologies were already quite mature by the time of the commercialization of the Internet. Technologies for access and intranet had been perfected by years of academic and governmental use, and could be applied immediately to organizational needs. In contrast, many complementary technologies needed to run commercial transactions successfully over the Internet remained to be developed. Organizations that wished to conduct commercial transactions faced considerable costs of adapting existing Internet technologies for use in these new applications. In this section, I examine the diffusion patterns of three classes of Internet technologies: basic access, intranet, and e-commerce. The sample is drawn from the Harte Hanks CI Technology database (hereafter CI database), a survey of establishment technology infrastructure conducted by consultants Harte Hanks Market Intelligence.[2] The sample includes all establishments over 100 employees surveyed by Harte Hanks in SIC codes 60-67, 73, 87, and 27 over the period 1996-98.[3]

[2] The CI is an abbreviation for computer intelligence, as Computer Intelligence Infocorp originally maintained this database.

[3] These are the industrial classifications for printing and publishing (27); finance, insurance, and real estate (60-67); business services (73); and engineering, accounting, and other management research and consulting firms (87). Further details on the sampling methodology are provided in Section 5 and in the data appendix.

2.1 Technology definitions

2.1.1 Access

Basic access is the means through which employees obtain content and/or send messages over the broader Internet. Access is the most basic type of Internet service, and is a necessary condition for the adoption of most other Internet applications. At its most basic level, access can involve nothing more than users obtaining dial-up service to the local Internet Service Provider (ISP), however, it can also involve the use of high-speed Internet connections such as a T-1 or T-3 line.[4] The technology for access was already well developed by the start of my sample, and because most early applications of access involved the retrieval of information from static webpages, the organizational costs of using access for basic research purposes were probably low.

2.1.2 Intranet

An intranet is a network based on TCP/IP protocols that is available internally to an organization's employees.[5] By allowing data from the Internet to be distributed within the organization, intranets enable an organization to maintain a single gateway to the public Internet. Intranets reduce the costs of high-speed access by enabling the costs of a T-1 or T-3 line to be spread over many users. They also improve network security by reducing the number of openings an intruder can use to break into the network. In addition to facilitating communication with the Internet, intranets are also used for communication solely internal to the organization. An example of this is when a human resources department publishes webpages for internal use by its employees.

Intranet adoption involves higher technical and organizational costs than access. Technical costs will be higher because intranet adoption may involve substantial retooling of the network if it does not support the Internet's TCP/IP networking protocols. The technical costs of migrating from host-based (i.e., mainframe) systems to an intranet will be particularly high. Organizational costs will also be higher. For instance, network administrators may have made specialized investments in the maintenance and management of proprietary network systems and may be resistant to adoption of a technology that renders these investments obsolete.

2.1.3 E-commerce

The last technology that I examine is a collection of Internet applications that enable communication between the firm, its customers, and suppliers. These applications are

---

4 T-1 and T-3 lines are dedicated connections supporting data rates or 1.544 and 43 megabits per second, respectively. Firms commonly leased all or part of these lines from local phone companies to obtain Internet access.

5 TCP/IP is the suite of major communication protocols that are used in the Internet.

grouped together because they perform similar functions and are expected to have similar costs of adoption.[6]

The purpose of these applications is to dynamically process idiosyncratic requests from customers and suppliers. They require communication between Internet applications and pre-existing firm databases, and may also require the firm to provide security and/or privacy protection to external users. The adoption of e-commerce often involves the creation of an entirely new distribution channel; a distribution channel that requires considerable organizational change in firm sales and distribution techniques. Thus, e-commerce applications are regarded as having the highest technical and organizational costs of adoption.

2.2 Technology diffusion

In this section I present aggregate patterns of Internet diffusion. The unit of observation in the CI database is the establishment. However, establishment-level analysis of the data may be inappropriate because adoption of networked applications likely depends on observable and unobservable characteristics of other establishments within the same organization. Because of these potential problems, I aggregate establishments up to the organization level and conduct all analyses using the organization as the unit of observation. For the purposes of this paper, an organization is defined as the set of all establishments in the sample from the same firm.[7]

Table 1 presents diffusion rates of the three technologies over the period 1996-98.[8] The most surprising fact in this table is that many organizations did not adopt any form of the Internet at all. Even in 1998, several years into the diffusion of the Internet, 25 per cent of organizations still had not adopted basic access. Given the low costs of basic access adoption, this is very surprising.

Second, there was considerable heterogeneity in the rate of diffusion across technologies. Although there were many non-adopters, adoption of access was rapid, achieving a 32.5 per cent penetration rate among organizations as early as 1996 and achieving 75.0 per cent diffusion by 1998. In contrast, adoption of e-commerce was substantially slower. Diffusion in 1996 was only 6.9 per cent, and grew to 27.8 per cent by the end of the sample. The diffusion rates for intranet were somewhere in between. These results

6 Establishments are said to have adopted e-commerce if they use any of the following applications: business-to-business e-commerce, business-to-consumer e-commerce, customer service, education, extranet, publishing, purchasing, and technical support.

7 A fuller analysis of the problems with establishment-level analysis, as well as a discussion of the methodology for constructing an organization-level data set, is included in the Data Appendix.

8 The size of the Harte Hanks sample increased significantly over the sample period, from 5389 observations in 1996 to 8388 observations in 1998. To remove concerns about changing sample composition, Tables 1 through 3 include only those observations that were in the sample in 1996. Analysis using all observations gives qualitatively similar results.

support the hypothesis that the fixed costs of adoption for access were far lower than were the costs of adopting intranet and e-commerce.

Table 2 shows the set of technologies adopted by organizations in each year. Several facts are worth noting. First, there was substantial heterogeneity in the set of applications adopted by organizations. Organizations differed in their costs and benefits to adopting the Internet, and responded by adopting very different combinations of technologies. Second, most organizations adopted access and intranet together (column 3), rather than adopting access alone. This is true in every year, and is noteworthy because it suggests that most organizations did not choose the simplest form of Internet adoption, even in the early years of Internet commercialization. It also suggests the potential presence of complementarities between access and intranet adoption. Of course, other factors may be driving this pattern of multiple technology adoption.

Organizations may choose to adopt multiple Internet technologies all at once. They may also adopt them incrementally, adopting progressively more complicated technologies after gaining familiarity with simpler ones. Adoption should occur over time if co-invention costs are high or if there are large incremental costs to adopting new technologies, and should occur simultaneously if the fixed costs of making corporate networks 'Internet-compatible' are the major barrier to adoption. Using another feature of the dataset, it is possible to examine the timing of Internet adoption by examining counts of the number of applications adopted.[9] Table 3 details the number of applications adopted by organizations in 1998 and over the entire sample period. Analysis of the number of new applications adopted in 1998 (not included in the table) reveals that first-time adopters of the Internet adopt on average 1.42 applications, while prior adopters obtain an average of 0.81 new applications. This difference is significant at the 1 per cent level, potentially suggesting that organizations front-loaded their adoption of Internet technologies. In section 6.2, I will use count data models to more carefully examine the timing of Internet adoption and use the results to make inferences about the size of co-invention costs for Internet technologies.

Overall, Tables 1 through 3 show substantial heterogeneity both in the rate and manner with which organizations adopted Internet technologies. Many organizations did not adopt the Internet at all. Those who did often adopted access and intranet together. It also appears that organizations often front-loaded their adoption of Internet applications. These results suggest there may exist substantial heterogeneity in the organizational and technical costs and benefits of adoption across firms. What it does not identify, however, are the particular factors driving this heterogeneity in adoption behaviour. Why are some firms adopting more quickly than others? Are geographical, competitive, or firm-specific organizational or technical factors driving the patterns observed? To answer these questions, one needs to more carefully identify the potential factors driving organization

9 Using my terminology, there may be multiple Internet applications for a particular Internet technology. Access, intranet, and e-commerce are defined as Internet technologies. Table A.2 shows some of the applications observed for intranet. The total number of applications adopted by an organization is calculated by summing applications adopted for all three technologies. The construction of this variable is detailed more fully in the data appendix.

behaviour. In the next section several theories are presented that may explain the variation in adoption behaviour across organizations. An econometric framework is then developed to examine which are consistent with the empirical evidence.

3 Theories of Internet adoption

I examine the ability of several competing hypotheses to explain organization-level decisions to adopt Internet technologies. Section 3.1 describes several hypotheses that identify potential differences among firms in the costs and benefits of adoption. These hypotheses will describe how prior investments in installed base can have potentially competing effects on the probability of adoption, as well as how geographic decentralization can influence the likelihood of adopting the Internet. In Section 3.2, I discuss the role of complementarities and adoption costs on organization decisions to adopt multiple Internet technologies.

3.1 Theories on variation in the costs and benefits of adoption

In this section several theories are presented, detailing how the costs and benefits of Internet adoption vary among firms. Later I will examine organizational adoption behaviour using an econometric model. A set of proxies for each theory will be developed and used to investigate the usefulness of each theory as a predictor of organizational demand for the Internet.

3.1.1 Lead user theory

The first hypothesis comes from the classic diffusion literature (e.g., Rogers 1995). It says that innovative organizations, or those that are traditionally closest to the technical frontier, will be among the first to adopt Internet technologies. There are two major reasons for this. First, technically sophisticated organizations may have the internal skills necessary to adapt TCP/IP-based technologies to user needs. As a result, they will be more willing to bring such technologies into the organization.

Second, organizations that are on the technical frontier will be more likely to have made the transition from host-based to decentralized C/S computing. As noted above, the costs of Internet adoption for such organizations will be lower because most Internet technologies are a natural extension of the C/S paradigm, and can be run over C/S networks.[10] Several authors (e.g., Hitt and Brynjolfsson 1997) have also argued that firms that have adopted C/S are more likely to have made complementary investments in decentralized authority and decision structures. These investments are likely to lower the organizational costs of adopting Internet technologies.

---

[10] See Bernard (1998) and Orfali, Harkey, and Edwards (1999) on the costs of converting from proprietary C/S to an intranet.

3.1.2 Competing effects of installed base

The lead user theory said that prior investments in C/S increase the probability of adopting Internet technologies. However, although the average effect of C/S investments may be positive, marginal investments in platform-specific technologies may actually slow the rate of Internet adoption. The competing effects of installed base theory emphasize how platform-specific investments may hinder Internet adoption. Legacy investments can slow adoption if users have developed competencies or made complementary investments in the installed base that are incompatible with new technologies. In such cases, tangible and intangible investments in the installed base may raise the costs of switching to the new technology, effectively locking in users.[11]

Lock-in can manifest itself in several ways in the adoption of Internet technologies. First, pre-existing investments in C/S software that is customized to current systems may be difficult to transfer to new platforms. Second, investments in proprietary vendor technologies such as Novell's NetWare may be incompatible with the Internet's protocols. Last, a large installed base in mainframe systems will likely slow migration to the Internet because of the considerable costs of migrating from host-based to C/S-type Internet platforms.

3.1.3 Geographic concentration

The geographic concentration theory says that the decreases in communications costs caused by the adoption of Internet technologies will be less valuable for geographically concentrated organizations and those located in urban areas. Organizations with these characteristics will be less likely to adopt access, intranet, and e-commerce technologies.

Organizations that have adopted the Internet at multiple locations and have geographically dispersed establishments can send data communications over the Internet backbone. This technology, known as virtual private network (VPN), may represent a significant cost savings over alternatives such as face-to-face communication, traditional mail, or even data communications over private line services. Geographic dispersion will also increase the benefits of e-commerce adoption if Internet technologies can lower coordination costs between multiple organization departments, suppliers, and customers. Organizations that are concentrated geographically will not receive these benefits from adopting the Internet, and so may be less likely to adopt.

For similar reasons, the marginal benefits to Internet adoption will be lower for organizations primarily located in urban areas. Urban organizations will, on average, be closer to customers and suppliers. Electronic coordination with external parties is likely to be less important for these organizations than for those located in rural areas.

The link between organizational design and the demand for IT is not new. Neither is the link between coordination costs and IT demand. Prior papers have argued that organizations with decentralized authority structures or high coordination costs benefit

[11] Klemperer (1995) offers a survey of switching costs and lock-in.

most from the lower communication costs caused by technological change in IT.[12] However, to my knowledge this paper is the first to examine the link between IT demand and the geographic location of human capital within the firm.

3.2 Theories on multiple technology adoption

Tables 2 and 3 showed that most adopters of the Internet adopted multiple technologies rather than a single technology in isolation. In this section, two theories are provided on the forces driving multiple technology adoption. The data requirements necessary to test these explanations are stringent, and I will be unable to do so in this paper. Rather, I adopt the strategy of discussing whether the statistical results are consistent with each hypothesis and discuss qualitative features of the technologies that would tend to bolster or refute the empirical findings.

3.2.1 Complementarity

Often the benefits from adopting a technology are increased if a second, complementary, technology is adopted concurrently. Milgrom and Roberts (1990) show how the presence of complementarities can induce firms to adopt clusters of practices. The complementarity theory postulates that complementarities exist between Internet technologies and that organizations will tend to practice clustering in their adoption of the Internet. The data requirements necessary to formally test for complementarities are stringent, however, because of the difficulty of identifying between complementarities and unobservable shocks impacting multiple organizational decisions. Because of lack of necessary data, I will be unable to formally test for complementarities here.[13]

3.2.2 Learning and experimentation

While the complementarity theory states that organizations will tend to adopt clusters of Internet applications, the learning and experimentation theory looks at how organizations adopt multiple Internet technologies over time. Organizations often adopt new technologies incrementally, experimenting with simple configurations on a subset of systems before implementing the technology fully. This will be particularly true when there exist high co-invention costs of adapting new technologies to user needs. The learning and experimentation hypothesis predicts that organizations adopt Internet applications sequentially, first gaining experience with low-cost applications such as access before adopting applications with higher organizational and technical costs such as e-commerce.

---

[12] For example, Malone, Yates, and Benjamin (1987) argue that technical improvements in IT will lead to an overall shift from hierarchical- to market-based transactions. Hitt and Brynjolfsson (1997) and Brynjolfsson and Hitt (1998) examine empirically the relationship between IT demand and organizational design.

[13] Athey and Stern (1999) show formally the problems of empirically identifying complementarities, and present a method for doing so.

To examine the role of learning and experimentation on multiple technology adoption, I estimate count data models of the number of Internet applications adopted. A lagged measure of adoption is included in the model to identify whether prior adopters of the Internet are likely to demand more or less Internet applications. If prior adopters are shown to adopt more applications, then I will argue there is evidence that learning and co-invention costs play some role in the adoption of Internet technologies.

4 Empirical models

I use discrete choice models to examine the adoption of access, intranet, and e-commerce. Proxies indicating the installed hardware and software within the organization are used to indicate whether the empirical evidence is consistent with each theory. These adoption models are discussed in section 4.1.

In section 4.2, I discuss the count data models that will estimate the determinants of *number* of Internet applications adopted. These models will be used to examine the timing of Internet adoption, in particular whether organizations front-load their adoption of Internet applications.

4.1 Discrete choice adoption model

Discrete choice techniques are used to model adoption behaviour. These discrete choice models must be able to control for two features of the data. First, adoption decisions will be driven by common factors both observable and unobservable to the econometrician, thus unobserved characteristics are likely to be correlated across decisions. Second, intranet and e-commerce can be adopted only if Internet access is as well.[14] Thus, there is a potential selection problem because adopters of access may differ systematically from other organizations in the population.

A decision rule is assumed in which organizations first decide whether to adopt basic access. In the second stage of the decision rule, organizations make simultaneous decisions to adopt intranet and e-commerce. Figure 1 presents a diagram of the decision rule.

Formally, I assume that the underlying adoption of a particular Internet technology is related to a latent variable. Let Y_{ik} be a dummy variable denoting organization i's adoption of Internet technology k, where $k=1$ denotes access, $k=2$ denotes intranet, and $k=3$ denotes e-commerce. Let Y_{ik}^{*} denote the related latent variable, such that $Y_{ik}=1$ if

14 Theoretically, a firm could operate an intranet without even obtaining access to the broader Internet, though in practice this is rarely done. Out of a total sample of 8,388 organizations in the analysis sample, 3,590 were observed to adopt intranet. Of those 3,590, 64 (about 1.8 per cent) reported adoption of intranet but not access. Many of these 64 probably represented reporting error.

and only if $Y_{ik}^* > 0$.[15] Y_{ik}^* is assumed to be a function of observable firm characteristics X_{ik} and a random error term e_{ik}. Thus, for a given organization *i*,

$$Y_{ik}^* = b_k' X_{ik} + e_{ik}$$
$$Y_{ik} = 1(Y_{ik}^* > 0)$$

where $1(\cdot)$ denotes the indicator function which is 1 when $Y_{ik}^* > 0$.

To complete the model, the assumption is made that the error terms are jointly distributed trivariate normal.

$$\begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{pmatrix} \sim N_3 \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & & \\ r_{21} & 1 & \\ r_{31} & r_{32} & 1 \end{pmatrix} \right]$$

If $q_{ik} = 2Y_{ik} - 1$, then the log-likelihood function for a sample of independent observations *i* will be

$$\log L = \sum_{i=1}^{N} \{ Y_{i1} \log \Phi_3 [q_{i1} b_1' X_{i1}, q_{i2} b_2' X_{i2}, q_{i3} b_3' X_{i3}, q_{i1} q_{i2} r_{12}, q_{i2} q_{i3} r_{23}, q_{i1} q_{i3} r_{13}] \\ + (1 - Y_{i1}) \log \Phi [q_{i1} b_1' X_{i1}] \} \qquad [1]$$

where $\Phi_3[\cdot]$ and $\Phi[\cdot]$ denote the cumulative distribution functions of the trivariate and univariate normal, respectively. The model is estimated using maximum likelihood.

4.2 Count data models

Count data models are used to identify the factors determining the number of applications that organizations adopt. The most common model used to analyse count data is the Poisson regression model. In this model, the distribution of a discrete random variable Y_i is assumed to be distributed Poisson

$$\Pr(Y_i = y_i) = \frac{e^{-l_i} l_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \ldots$$

15 More generally, one may say that organizations adopt Y_{ik} if and only if Y_{ik}^* crosses some threshold *T*, however *T* is not identified separately from a constant in the parameter vector.

In our model, Y_i is the number of Internet applications adopted by organization i over the relevant sample period. Generally, given a vector of explanatory variables X_i and coefficient vector b, a log-linear form is assumed, $\log l_i = b' X_i$.

Use of the Poisson regression model is sometimes criticized because the sample distribution of the dependent variable of interest violates some assumptions of the Poisson. The most common violation occurs because of the existence of overdispersion in the data. A common way of introducing overdispersion to the Poisson regression model is to add individual heterogeneity. Cameron and Trivedi (1997) list many possible ways to do this; the most common is the negative binomial model. In the negative binomial model, $\log l_i = b' X_i + e_i$ and Y_i assumes the following distribution

$$\Pr(Y_i = y_i) = \frac{\Gamma(q + y_i)}{\Gamma(q) y_i !} u_i^q (1 - u_i)^2, \qquad [2]$$

where $\Gamma(\cdot)$ denotes the gamma distribution, $q = 1/a$, $u_i = q/(q + l_i)$, and a is an additional parameter that induces overdispersion.

In my data, there may be a different form of non-Poissonness in the distribution of the dependent variable. In a given year, an organization can be observed to 'adopt' zero Internet applications in two different ways. First, it may be the case that the organization is a demander of Internet technologies, and zero is the number of new applications adopted in that year. However, an observation of zero can also arise because the organization is not a demander of Internet technologies at all.[16] In this case, the distribution of the dependent variable is a mixture of a Poisson and some other distribution indicating the organization is a demander of the Internet.

The most common implementation of this type of model is the zero inflated Poisson (ZIP) model (Lambert 1992). In this model, we define the indicator variable z_i and let $z_i = 1$ denote the regime in which an organization is not a demander of the Internet. We assume that the threshold model determining whether an organization is an Internet adopter is based on a probit, so that $\Pr(z_i = 1) = \Phi(g' W_i)$ where g is a coefficient vector and W_i is a vector of exogenous variables. If the organization is a demander of the Internet, it adopts a number Y_i applications distributed Poisson (l_i). Thus,

$$\Pr(Y_i = y_i) = \begin{cases} \Phi(g' W_i) + (1 - \Phi(g' W_i)) \Pr(Y_i = 0 \mid z_i = 0), & y_i = 0 \\ (1 - \Phi(g' W_i)) \Pr(Y_i = y_i \mid z_i = 0), & y_i > 0 \end{cases} \qquad [3]$$

[16] Greene (1997) provides an excellent discussion of this problem, and is the basis for much of the discussion in the rest of this section.

where $\Pr(Y_i = y_i \mid z_i = 0)$ is the Poisson probability that y_i applications are adopted given that the organization is a demander of the Internet. One can also reparameterize model [3] as a negative binomial, in which case $\Pr(Y_i = y_i \mid z_i = 0)$ is simply the negative binomial probability of y_i applications. This second model is known as the zero inflated negative binomial (ZINB).

The ZIP and ZINB models are not nested in the Poisson and negative binomial, creating problems for specification testing. Greene (1994) has shown that Vuong's (1989) test statistic for nonnested models may be able to discern between models. Asymptotically, the Vuong statistic is distributed standard normal, so one can use standard critical values from the normal to determine significance.

5 Data

As noted above, this study uses data from the Harte Hanks CI Technology Database. Earlier versions of the CI database have been used by other empirical researchers,[17] however this paper uses a newer version that includes information on establishment usage of TCP/IP-based technologies. The database is unique in that it contains detailed information on the hardware and software in use at individual firm establishments. Section 5.1 describes the sample that is used, and sections 5.2 and 5.3 detail the dependent and independent variables.

5.1 Sample

Data was obtained from the CI database over the period 1996-98. The CI database contains establishment-level data on (1) establishment characteristics such as number of employees, industry, and location; (2) usage of technology hardware and software such as computers, networking equipment, printers, and other office equipment; and (3) use of Internet applications and other networking services. Harte Hanks surveys establishments throughout the calendar year; my sample of annual data is assembled by obtaining the most current information as of December of each year.

To keep the analysis of manageable size, I obtained data from the CI database on SIC codes 60-67, 73, 87, and 27. These SIC codes correspond to the industrial groupings on finance, insurance, and real estate (60-67); business services (73); engineering, accounting, research, management, and related services (87); and printing and publishing (27). These industries were selected because they are generally regarded as heavy users of information technology. The sample contains data on all establishments of over 100 employees from the CI database in these industries over the sample period. All establishments are from the US.

---

[17] See, for example, Hitt and Brynjolfsson (1997), Ito (2000), and Bresnahan, Brynjolfsson, and Hitt (2002).

I use 1998 adoption data to estimate the static discrete choice and count data models described in section 4. To minimize potential endogeneity problems, I use prior year data on firm characteristics to analyse the factors driving Internet adoption. As a result, two consecutive years of data are required for each observation. Because there was some entry and exit of establishments and organizations in the CI database, this method of data construction required that a number of establishments and organizations be dropped from the analysis sample. Moreover, a small number of observations had to be dropped due to missing data.[18] In all, the establishment-level data originally obtained from the CI database contained 18725 *establishments* in 1998. The final analysis sample contains 8388 *organizations.*

5.2 Variables measuring Internet adoption

The primary focus of this paper will be with an organization's first adoption of technology. An organization will be recorded as adopting an Internet technology if it is adopted by at least one establishment within that organization. The CI database includes several measures of establishment Internet use. The data appendix describes how these measures are used to construct the adoption variables.

The count data models described in section 4.2 will be used to show the factors determining the total number of Internet applications that an organization adopts. To estimate count data models, a measure is needed of the number of Internet applications adopted. This is derived from additional surveys conducted by Hate Hanks in 1997 and 1998. A complete description of the methodology used to identify the number of applications adopted is included in the data appendix.

5.3 Exogenous variables

Table 4 contains the names of the variables, a short description, and their descriptive statistics. Descriptions of the variables are organized by hypothesis below.[19]

5.3.1 Lead user

The variable *PCPEREMP* represents the number of PCs in the organization per employee. Organizations with a high PC-to-employee ratio are likely to have more decentralized computing structures, potentially implying lower organizational costs of adoption. Moreover, because PCs represent the most common way of accessing the Internet, a high value for *PCPEREMP* likely represents lower technical costs of adoption.

---

[18] This occurred primarily when establishments did not report characteristics such as hardware and software installations or the number of employees.

[19] Note that organizational measures of the variables below are derived by aggregating the related establishment measures. The method of aggregation will sometimes depend on the construction of the variable. The description column in Table 4 will provide an explanation of construction when it is not clear from the context.

CLIENT indicates the percentage of establishments within the organization that have Internet-ready clients installed.[20] *CLIENT* indicates a minimal level of technical sophistication at the firm necessary for Internet adoption. Thus, low values of *CLIENT* should imply a lower probability of adoption among all three technologies.

NOAPP is a dummy variable that is one when all establishments are listed as having no software applications in the CI database.[21] Because measures of software use are frequently used as covariates, this variable is included primarily as a control.

The variable *PCTLAN* indicates the percentage of non-PC applications that are accessed through an organization's LAN. Organizations with no applications are coded as zero. Organizations with a high value for *PCTLAN* are ones that have adopted C/S heavily. These organizations should have lower technical and organizational costs of adoption.

SYSCOM is a dummy variable indicating the use of system software applications in the organization.[22] *NETWARE* is a variable that is one if the organization reports use of Novell's NetWare, an advanced LAN operating system best known for its uses as a file server. Since respondents to the Harte Hanks survey list only those software applications they feel are most important, a value of one for either for these variables likely represents a well-developed C/S network in use at the firm.

5.3.2 Competing effects of installed base

SYSCOM and *NETWARE* indicate prior investments in C/S that may lower the technical and organizational costs of adoption. However, they also indicate investments in platform-specific technologies that raise the costs of switching to a new platform. To identify how *SYSCOM* and *NETWARE* can raise adoption costs, I interact each with *PCTLAN*. A negative coefficient on these interaction terms indicates that the positive impact of *PCTLAN* on Internet adoption will be dampened if some C/S investments were directed toward technologies that are incompatible with Internet protocols. This negative impact will increase as *SYSCOM* and *NETWARE* are integrated into larger and more complicated networks (i.e., as *PCTLAN* increases).

The variable *PCTMAIN* indicates the percentage of non-PC applications that are accessed over a mainframe or minicomputer. *MAINOUT* indicates that an organization uses an outside consultant to handle the maintenance of its large-scale computing facilities (i.e.,

---

[20] For the purposes of this paper, Internet-ready clients include those with UNIX operating systems, Windows, Macintosh operating systems, and several smaller operating systems. Unfortunately, I was unable to identify between sites with Windows 95 or above versus those with Windows 3.1 or below, however I was able to identify and exclude sites with DOS installed.

[21] Establishments surveyed for the CI database do not supply a complete listing of software to Harte Hanks, rather they report the most important software in use. Thus, a value of 1 for *NOAPP* may indicate nonresponse or a potential lack of technical sophistication at the site.

[22] System software in this data set represents primarily software for the management, maintenance, and backup of LANs.

mainframes). These variables indicate investments in host-based hardware and services that should increase the costs of adoption.

5.3.3 Geographic concentration

The variable *MULTEST* is a dummy indicating there is more than one establishment in the organization. *MULTEST* is interacted with *EMPCON* to obtain a measure of the concentration of employees within the organization. *EMPCON* is calculated by summing the squared shares of employees in each establishment across all establishments in the organization.[23] A high value for *EMPCON* indicates that employees are concentrated within a small number of establishments, and should lower the probability of adoption.

MUTLTEST and *PCT100* are interacted to identify how greater geographic distance between establishments can influence the probability of adoption. To calculate *PCT100*, longitude and latitude are used to find the distance between establishments within an organization.[24] I then calculate the percentage of pairwise establishment combinations in the organization that are within 100 miles in distance.[25] A high value for *PCT100* indicates greater concentration of organization employees, and should lower the probability of adoption.

The variable *URBAN* indicates the percentage of establishments that are located in a major metropolitan statistical area (MSA). It is included to identify whether the benefits to Internet adoption are lower for organizations located in urban areas.

5.3.4 Controls

To capture the effects of organization size on the probability of adoption, I include the natural logarithm of the total number of employees in the organization. Because I expect ex-ante the marginal effects of increases in the number of employees on probability of adoption to vary with number of employees, I allow the coefficient on this variable to vary across ranges of firm size. Formally, the effect of organization size on probability of

---

23 Formally, let e_{ij} be the number of employees in establishment *i* and organization *j*, and $E_j = \sum_{i=1}^{N_j} e_{ij}$, where N_j is the number of establishments in organization *j*. Then, s_{ij}, the share of employees in organization *j* that are in establishment *i* will be $s_{ij} = e_{ij} / E_j$. Then $EMPHERF = \sum_{i=1}^{N} s_{ij}^2$. Ellison and Glaeser (1997) use a similar approach to measure geographic concentration in US manufacturing industries.

24 Wallsten (1999) used this approach to calculate distances between firms to measure the effects of agglomeration and spillovers on the probability of obtaining a grant from the federal government's Small Business Innovation Research (SBIR) programme.

25 Several other measures of distance were experimented with, including a linear measure as well as dummy variables for distances other than 100 miles. The results are robust to different measures of distance.

adoption is modelled as $b_{EMPLE}\log(EMPLE)$, where $EMPLE$ is number of organization employees and b_{EMPLE} is defined as

$$b_{EMPLE} = \begin{cases} b^1_{EMPLE}, & 100 \le EMPLE < 200 \\ b^2_{EMPLE}, & 200 \le EMPLE < 500 \\ b^3_{EMPLE}, & 500 < EMPLE \end{cases}.$$

To control for industry effects, the model includes variables that indicate the percentage of establishments in SIC codes 60-67, 73, and 87. SIC 27 is the omitted category. *PROGOUT* is a dummy variable included to capture the effects of IT outsourcing.

5.3.5 Complementarity

To examine whether organizational adoption decisions are consistent with the presence of complementarities, I examine the size and magnitude of the correlations in the error terms in the econometric model described in [1]. No exogenous covariates are included to test this theory.

5.3.6 Learning and experimentation

To examine whether organizations adopt multiple Internet applications simultaneously or sequentially, I include dummies indicating prior adoption of access (*DPACC*) in the count data model. It is expected that if learning and experimentation exert an important effect on Internet adoption, then the coefficient on this lagged dependent variable will be positive.

6 Results

This section presents the results of the empirical models. In section 6.1 the results of the adoption models are presented, with the results organized by hypothesis. Several extensions to the baseline model are presented, including the addition of a greater number of industry effects, controls for metropolitan areas, and interaction of industry and geographic effects. In section 6.2, I describe the results of the count data model, in particular examining the timing of Internet adoption.

6.1 Adoption model

Table 5 presents the baseline results. To ease interpretation of the coefficients, Table 6 shows the effect of a change in each variable from 0 to 1 on the probability of adoption. For continuous variables, these marginal effects show the impact on adoption of an increase in each variable from 0 to 100 per cent. The marginal effects are computed for a 'typical' organization in the sample.[26]

6.1.1 Lead user

Table 5 shows substantial evidence that technical sophistication and prior investment in C/S technologies play a role in Internet adoption, particularly in the adoption of access and intranet. The coefficients on *PCPEREMP*, *CLIENT*, and *NOAPP* are all significant and consistent with the hypothesis that the probability of access, intranet, and e-commerce adoption increases with technical sophistication. In contrast, organizations with low values of these basic measures of technical sophistication are unlikely to adopt even basic access. The expected probability of access adoption for an organization for which *PCPEREMP* and *CLIENT* are each one half standard deviation below the mean and for which *NOAPP*=0 (and for which all other variables take the values in footnote 26) is only 9.3 per cent.

Table 5 shows that prior investments in C/S technologies, as measured by *PCTLAN*, significantly increase the probability of access and intranet adoption but have less of an impact on the probability of e-commerce adoption. An increase in the percentage of LAN applications from 0 to 100 per cent increases the probability of adoption by 24.9 per cent and 21.1 per cent for access and intranet, but only 6.5 per cent for e-commerce.[27] Investments in *NETWARE* also have a stronger marginal impact on access (12.7 per cent) and intranet (11.8 per cent) than on e-commerce (4.9 per cent). Moreover, the coefficient estimates for *PCTLAN*, *SYSCOM*, and *NETWARE* in the e-commerce equation are all insignificant. One reason for this pattern is that access and intranet generally require complementary investments in an organization's internal network infrastructure to be effective. In contrast, many e-commerce applications may be hosted by third-party providers. These applications generally do not require the pervasive use of C/S in the organization that is captured by *PCTLAN*, *SYSCOM*, and *NETWARE*.[28]

6.1.2 Competing effects of installed base

The results show that investments in platform-specific technologies can raise the costs of Internet adoption. The coefficients on *NETWARE* × *PCTLAN* and *NETWARE* × *SYSCOM* have a negative impact on all three adoption decisions and a statistically and economically significant impact on the decisions to adopt access and intranet. Conditional on investment in *NETWARE*, an increase in *PCTLAN* from 0 to 100 per cent

---

[26] This typical organization has 200 employees and mean values for *SIC60-SIC87, URBAN, PCPEREMP, CLIENT, PCTLAN,* and *PCTMAIN*. The baseline organization is assumed to have one establishment, and to have values of zero for *SYSCOM, NETWARE, PROGOUT,* and *MAINOUT*. Marginal effects for *NETWARE* × *PCTLAN* are computed using values of one for *NETWARE*, and marginal effects for *SYSCOM* × *PCTLAN* and the geographic interaction terms are computed similarly.

[27] The marginal effect of *PCTLAN* on e-commerce is statistically significant. However, this is due primarily to the strong impact of *PCTLAN* on access adoption and the structure of our selection model.

[28] The fact that I am analysing the diffusion of early Internet applications is important. More recent e-business implementations, as described by, for example, Kalakota and Robinson (2001), emphasize the integration of e-commerce applications with the internal information systems infrastructure.

decreases the probability of access, intranet, and e-commerce adoption by 21.5 per cent, 24.5 per cent, and 6.6 per cent, respectively. These marginal effects show that for very complex networks—or ones with a very high value for *PCTLAN*—the negative lock-in effects will more than offset the benefits of investing in *NETWARE*. Similar results are obtained for *SYSCOM*.

On net, do *NETWARE* and *SYSCOM* have a positive or negative impact on the probability of access and intranet adoption? In other words, which is more important, technical sophistication or lock-in? The answer will depend on the complexity of the organization's C/S network. If the installed base of LAN applications is relatively small, then the positive impact of technical sophistication will outweigh the effects of additional switching costs. Investment in *NETWARE* and *SYSCOM* will then increase the probability of adoption. Conversely, if the installed base of LAN applications is very large, then the effects of large switching costs predominate, lowering the likelihood of access and intranet adoption. In most cases, the positive effects of technical sophistication will outweigh the negative impact of lock-in. Investment in *NETWARE* increases the probability of access adoption for 96.7 per cent of organizations and increases the likelihood of intranet adoption for 73.3 per cent organizations. For *SYSCOM*, the corresponding probabilities are 72.7 per cent for access and 90.6 per cent for intranet.

The coefficients on *PCTMAIN* have the expected negative impact on the probability of access, intranet, and e-commerce adoption, however, are economically and statistically insignificant.[29] The coefficients on *MAINOUT* are also statistically and economically insignificant.

6.1.3 Geographic concentration

The proxies for geographic concentration have among the strongest impact on the probability of adoption. Even controlling for organization size, multi-establishment organizations have a much higher probability of adopting all three Internet technologies. The marginal effects of *MULTEST* on access, intranet, and e-commerce adoption are 35.7 per cent, 50.8 per cent, and 38.6 per cent, respectively. These marginal effects are among the highest of any variable (including variables with positive and negative coefficients) on the probability of adopting these technologies.

The probability of adoption decreases significantly for organizations that are geographically concentrated. For multi-establishment organizations, an increase in *EMPCON* from 0 to 1 will decrease the probability of access (-26.7 per cent), intranet (-56.5 per cent), and e-commerce (-42.0 per cent) adoption. $MULTEST \times PCT100$ has an economically and statistically significant effect on access and intranet adoption: an increase in $MULTEST \times PCT100$ from 0 to 1 decreases the likelihood of adopting these technologies by 6.6 per cent and 13.6 per cent, respectively. However, it has relatively little effect on the probability of adopting e-commerce.

---

[29] This result is not due to collinearity between *PCTLAN* and *PCTMAIN*. The omitted categories of software in the adoption equation are applications on minicomputers, small servers, and workstations that are not accessed by users over the LAN. I also estimated the adoption equation without *PCTLAN* and any of its interaction terms but with *PCTMAIN*. *PCTMAIN* remained insignificant.

The coefficient for *URBAN* is negative and has a statistically significant marginal effect on access (-7.3 per cent) and e-commerce (-5.2 per cent) but has little effect on the adoption of intranet (-2.8 per cent). This is likely because access and e-commerce are technologies that allow communication between the organization and the outside world. Such technologies will be particularly valuable for organizations located predominantly in geographically isolated areas, and will be less valuable for organizations located in urban areas.

Overall, the results confirm the hypothesis that organizations that are geographically concentrated or located in urban areas will be less likely to adopt Internet applications.

6.1.4 Complementarity

How do organizations adopt multiple Internet technologies? There is strong evidence that access and intranet are adopted together, but less evidence that e-commerce is adopted with access or intranet. This is, of course, consistent with the aggregate data in Tables 1 and 2. The parameter r_{12} denotes the correlation between the access and intranet decisions. The estimate of r_{12} (0.88) is strong and significant. Conversely, the estimates of the correlations between access and e-commerce and intranet and e-commerce are small and statistically insignificant (the estimates are 0.12 and 0.09, respectively).

Positive correlation in the unobservables among adoption of practices is often interpreted as a sign of potential complementarities. However, it is well known that one should exercise care in interpreting positive correlations in this way, as they may simply reflect unobserved shocks that are common to both practices. [30] Unfortunately, there is no way of formally testing whether this observed clustering of activities represents true complementarity. [31]

Why would organizations adopt access and intranet together but not adopt e-commerce with other applications? Cost may be one reason. The technical and organizational costs of adopting access and intranet were lower than that of e-commerce throughout the sample. Most of the technical problems associated with the adoption of access and intranet had already been solved by scientists and engineers before the Internet was commercialized. Moreover, because early uses of access and intranet revolved around sending e-mail and viewing webpages, the co-invention costs of adapting these technologies to organization needs were quite low. In contrast, the costs of adopting e-commerce were much higher. Many e-commerce technologies were simply more complicated than other Internet technologies. Also, many of the complementary

---

[30] Arora and Gambardella (1990) show formally how correlation coefficients can be used as a test for complementarity. Athey and Stern (1999) describe formally the problems with use of this method.

[31] There does exist some anecdotal evidence that we should expect Internet technologies to be complementary. For example, a high-speed Internet access connection must be shared by several users to be cost-effective. This sharing can only be achieved if users can access the connection through an intranet. An establishment intranet becomes more valuable if it can be connected to other establishment intranets, although such a connection may only be feasible over the public Internet, for which access is required.

technologies necessary to the conducting of commerce over the Internet did not yet exist. The World Wide Web's core technologies: the TCP/IP communication protocol; the hypertext markup language (HTML) used to create webpages; and the uniform resource locator (URL) address system, were all ill-equipped to deal with security issues or to facilitate the dynamic communication between database servers and webpages necessary to conduct commerce transactions.[32]

6.1.5 Industry effects

Many of the SIC variables are significant, and tell an interesting story about differences in the costs and benefits of Internet adoption across industries. The coefficients on SIC 60—which includes commercial banks, credit unions, and savings institutions—were strong, negative, and significant for access, intranet, and e-commerce. Why were banks so slow to adopt, particularly given the promise of on-line banking? An important component of banking information systems is transaction processing, which requires sizeable technical investments to ensure security and reliability. Early Internet technologies were severely lacking in both these dimensions, implying high co-invention costs.

6.1.6 Robustness checks

Table 7 shows the results of re-estimating model 1 using three-digit rather than two-digit SIC codes.[33] The coefficient on $SYSCOM \times PCTLAN$ was no longer statistically significant, however otherwise the basic flavour of the results remain unchanged. The model was also estimated using variables indicating the percentage of establishments in the top 20 US metropolitan areas and interactions of these variables with employment. The results are in Table 8. In this model, some loss of significance occurred on the coefficient for *URBAN*. This is not surprising, because many *URBAN* establishments were located in the top 20 metropolitan areas. The impact of *PCT100* on the adoption of access was also no longer statistically significant. Otherwise, the results were unchanged.

One criticism of model 1 may be due to the aggregation of establishments to the organization level. Because organizations in the sample are counted as adopting as soon as one establishment adopts, the results from Table 5 may be driven by the behaviour of 'outlier' establishments who are the most likely to adopt Internet technologies. Table 9 shows the results of estimating the model 1 using only single-establishment organizations. The coefficient estimates are very similar to those in Table 5. Thus, it does not seem that the results are driven by outlier establishments.

---

[32] For a detailed description of some of the problems of adapting Internet technologies to commerce activities, see Orfali, Harkey, and Edwards (1999).

[33] In the baseline model I was unable to reject the hypothesis that the coefficients for *EMP200* and *EMP500* were equivalent. Therefore, for these robustness checks we redefine *EMP200* to include all employee quantities greater than 200. Observations in SIC 601 (central reserve depository institutions) and SIC 632 (security and commodity exchanges) were dropped because of the small number of observations in each class. As a result, 10 observations were dropped.

6.2 Count data results

Table 10 presents the results of estimating four count data models—the Poisson, negative binomial, ZIP, and ZINB—on counts of the number of Internet applications adopted in 1998. Columns 3 and 4 show the results of estimating the ZIP and ZINB models. The Vuong statistic comparing the results of the ZIP to the Poisson model is positive and very significant at 15.62. The Vuong statistic that compares the ZINB to negative binomial is much smaller at 8.47, suggesting that some of the improvement in moving from the Poisson to the ZIP may have been because the ZIP allows for overdispersion in the data in addition to inducing a regime splitting mechanism. However, it is still statistically significant. Thus, the discussion will focus attention primarily on the ZINB model.

The sign of the coefficient estimates for the Poisson and negative binomial models are generally the same as those in the adoption models in the previous section, and usually retain their statistical significance. This similarity may partially reflect the fact that these models are capturing adoption as well as quantity decisions. However, in the ZIP and ZINB models, some of the variables measuring prior networking investments show some differences with the adoption models. The variables *PCTLAN*, *NETWARE*, and $NETWARE \times PCTLAN$ lose their significance in the ZIP and ZINB models. The signs on *SYSCOM* and $SYSCOM \times PCTLAN$ are also reversed from the adoption equations. These differences in the networking variables may suggest these variables are important determinants of adoption but are less effective at predicting the quantity of applications demanded.

The effect of early Internet adoption on the number of new applications adopted in 1998 is negative and statistically and economically significant in all four models. For the typical organization that was described in section 6.1, the ZINB model implies that prior adoption will reduce the expected number of applications an organization adopts in 1998 from 0.42 to 0.35, a decline of 0.07 applications (or 16 per cent). The Poisson, negative binomial, and ZIP models all predict similar declines of 0.06, 0.06, and 0.07 applications. This suggests that organizations tended to adopt the largest number of applications when they adopted the Internet for the first time

Thus, the result of organizational front-loading of Internet adoption is robust to several reparameterizations of the model. However, one should interpret the results of this section with some care, since there are some potential alternative explanations for the results. One potential problem with the estimation strategy may be that prior access adoption may be endogenous with the error term in the model, leading to biased and inconsistent estimates of the effects of prior adoption on the number of applications adopted. Cameron and Trivedi (1997) show that nonlinear instrumental variables can be used to remove the endogeneity problem, however it is difficult to identify instruments with the proper characteristics.

A second potential problem is caused by the way the CI database is constructed. Recall that respondents only report those applications they feel are most important. If applications that are adopted later have lower marginal value to organizations, they are less likely to be reported. In this case, one should reinterpret the results of Table 10 as saying organizations adopt their most important Internet applications when they adopt access for the first time.

6.3 Assessment

What factors separated adopters of the Internet from non-adopters? Several forces were at work. Among the most important determinants were technical sophistication and prior investments in C/S technologies that lowered the technical and organizational costs of adoption. Although the average impact of C/S investments was unequivocally positive, the impact of marginal investments in platform-specific network software depended on the relative strengths of two competing effects, technical sophistication and lock-in. The net result of these two effects depended on the complexity of the installed base of LAN applications within the organization. In most cases, the positive effects of technical sophistication outweighed the negative effects of lock-in.

I find evidence that the organizations most likely to adopt the Internet were those that benefited most from the decreases in communications costs created by early Internet technologies. This result is consistent with papers that argue there exists a complementary relationship between IT investments and decentralization in organization authority. Increases in the geographic concentration of establishments had a strong negative impact on the likelihood of adoption. Urban organizations were less likely to adopt technologies such as access and e-commerce that permitted easier communication with parties outside the firm. Smaller firms that were geographically concentrated often benefited less from basic Internet communication technologies such as the World Wide Web and e-mail, and often chose not to adopt the Internet.

Organizations usually adopted multiple Internet technologies rather than adopting a particular technology in isolation, and in particular clustered their adoption of access and intranet. Some evidence of complementarity between Internet technologies was presented, and I argued that slower adoption of e-commerce and lack of clustering between e-commerce and other technologies were due to higher adoption costs.

Contrary to expectations, the analysis showed that learning and experimentation did not play a major role in the adoption of early Internet technologies. Organizations appeared to adopt multiple technologies simultaneously rather than spreading out their adoption over time. I believe this is because the co-invention costs of many early Internet technologies such as access and intranet were low.

These results suggest a common theme in the pattern of Internet adoption. Organizations first invested in access and intranet technologies that were relatively inexpensive to adopt. These early technologies lowered communication costs through applications such as e-mail and the World Wide Web, and were adopted first by organizations that had the most to gain from such reductions in communications costs. These technologies diffused rapidly. However, there remained many organizations for which the value of these early applications was relatively low, and for which non-adoption was the optimal response.

The diffusion of e-commerce applications took much longer than access or intranet. These technologies were technically more difficult to implement, and required investments in complementary technologies that had not yet been fully developed. The covariates in the adoption equation for e-commerce were less successful at explaining the variation in adoption behaviour because the organizational determinants of adoption are likely to be different. Moreover, non-organizational factors are likely to play a bigger role in e-commerce adoption than they did for access and intranet. In particular, competitive conditions, adoption decisions by incumbents, and market entry by new firms with Internet-based business models are all likely to play important roles.

7 Conclusion

In this study I examined reasons for the 'digital divide' between adopters and non-adopters of the Internet. Several hypotheses were suggested that identified organization features that potentially affected the costs and benefits of adoption. I then examined whether the empirical evidence was consistent with each hypothesis using a set of proxies.

Recent papers have empirically examined the relationship between IT demand and organizational structure. These papers have found a complementary relationship between IT and organizational decentralization. They further predict that organizations with decentralized, 'coordination-intensive' structures will benefit most from declines in communications costs created by technological change in IT. This paper has provided evidence that organizations with investments in decentralization and those that had a greater marginal benefit for low communications costs were among the first to adopt the Internet. Firms that invested previously in decentralized C/S computing systems and that were close to the technological frontier were usually among the first to adopt new Internet technologies. However, prior investments in C/S computing technologies that were platform-specific left some organizations ill-positioned to take advantage of the lower communications costs offered by the Internet.

I argued that the rapid diffusion of early Internet technologies was due to the low co-invention costs of adapting technologies like access and intranet to organizational needs. This hypothesis is supported by the manner with which organizations adopted the Internet. Organizations tended to cluster their adoption of access and intranet. Moreover, they tended to front-load their adoption of applications, rather than pursue an incremental strategy of adopting new technologies after periods of learning and adaptation.

This study was, in many ways, an exploratory analysis of the factors driving Internet adoption, and raises many new questions for future research. First, future work should more explicitly examine the roles of geographic and competitive factors on Internet adoption. In particular, what role did new web-based entrants play on organizational decisions to adopt access, intranet, and e-commerce? Further work should also examine intra-organization diffusion of Internet technologies and more carefully study the evolution of usage within organizations. Future studies should try to quantify the benefits of Internet adoption, examining the revenue-enhancing and cost-reducing effects of these technologies. Last, as noted above, this study has examined only the early stages of diffusion for a set of evolving technologies. More work should be done examining the diffusion of more complicated—and costly—Internet technologies. In all, this paper has taken a first look at understanding the forces driving investment in early Internet technologies. Much work remains to be done in understanding how these technologies have been implemented, and the impact of their usage.

References

Arora, A., and A. Gambardella (1990). 'Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology'. *The Journal of Industrial Economics*, 38: June.

Athey, S., and S. Stern (1999). 'An Empirical Framework for Testing Theories About Complementarity in Organization Design'. NBER Working Paper #6600. Cambridge, MA: National Bureau of Economic Research.

Bernard, R. (1998). *The Corporate Intranet*. New York: John Wiley & Sons, Inc.

Bresnahan, T., E. Brynjolfsson, and L. Hitt (2002). 'Information Technology, Work Organization, and the Demand for Skilled Labor: Firm-Level Evidence'. *Quarterly Journal of Economics*, 117 (1): 339-76.

Bresnahan, T., and S. Greenstein (1997). 'Technical Progress and Co-invention in Computing and in the Uses of Computers'. *Brookings Papers on Economic Activity: Microeconomics*,1-78.

Bresnahan, T., and M. Trajtenberg (1995). 'General Purpose Technologies: "Engines of Growth"?'. *Journal of Econometrics*, 65: 83-108.

Brown, J., and A. Goolsbe (2002). 'Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry'. *Journal of Political Economy*, 110 (3): 481-571.

Brynjolfsson, E., and L. Hitt (1998). 'Information Technology and Organizational Design: Evidence from Micro Data'. Working Paper. Cambridge: MIT Sloan School of Management.

Brynjolfsson, E., and M. Smith (2000). 'Frictionless Commerce? A Comparison of Internet and Conventional Retailers'. *Management Science*, 46: 563-85.

Cameron, C., and P. Trivedi (1997). *Models for Count Data.* Oxford: Oxford University Press.

Carlton, D., and J. Chevalier (2001). 'Free Riding and Sales Strategies for the Internet'. NBER Working Paper 8067. Cambridge, MA: National Bureau of Economic Research.

Ellison, G., and E. Glaeser (1997). 'Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach'. *Journal of Political Economy*, 105: 889-927.

Gertner, R., and R. Stillman (2000). 'Vertical Integration and Internet Strategies in the Apparel Industry', Working Paper. Chicago: Graduate School of Business, University of Chicago.

Greene, W. (1994). 'Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models'. Working Paper EC-94-10. New York: Dept. of Economics, Stern School of Business, New York University.

Greene, W. (1997). *Econometric Analysis*, 3rd Edition. New York: Prentice Hall.

Griliches, Zvi (1957). 'Hybrid Corn: An Exploration in the Economics of Technical Change'. *Econometrica*, 25 501-22.

Hannan, T., and J. McDowell (1984). 'The Determinants of Technology Adoption: The Case of the Banking Firm'. *RAND Journal of Economics*, 15: 328-35.

Hitt, L., and E. Brynjolfsson (1997). 'Information Technology and Internal Firm Organization: An Exploratory Analysis'. *Journal of Management Information Systems*, 14: 81-101.

Ito, H. (2000). 'The Structure of Adjustment Costs in Mainframe Computer Investment'. Working Paper. Providence: Dept. of Economics, Brown University.

Kalakota, R., and M. Robinson (2001). *e-Business 2.0: Roadmap for Success.* Reading, MA: Addison-Wesley.

Klemperer, Paul (1995). 'Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade'. *Review of Economic Studies*, 62: p. 515-39.

Lambert, D. (1992). 'Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing'. *Technometrics*, 34: 1-14.

Malone, T., J. Yates, and R. Benjamin (1987). 'Electronic Markets and Electronic Hierarchies'. *Communications of the ACM*, 30 (6): 484-97.

Milgrom, P., and J. Roberts (1990). 'The Economics of Modern Manufacturing: Technology, Strategy, and Organization'. *American Economic Review*, 80: 511-28.

Orfali, Robert, Dan Harkey, and Jeri Edwards (1999). *The Client/Server Survival Guide*. New York: John Wiley & Sons, Inc.

Rogers, E. (1995). *Diffusion of Innovations*, 4th Edition. New York: The Free Press.

Scott Morton, F., F. Zettelmeyer, and J. Risso (2000). 'Internet Car Retailing'. NBER Working Paper No. 7961. Cambridge, MA: National Bureau of Economic Research.

Vuong, Q. (1989). 'Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses'. *Econometrica*, 57, 307-34.

Wallsten, S. (1999) 'Geographic Clustering and Spillovers: An Empirical Approach Using Geographic Information Systems and Firm-Level Data'. Working Paper. Stanford: Stanford Institute for Economic Policy Research, Stanford University.

List of figures and tables

Figure 1 Decision process for Internet adoption 28

Table 1 Organization adoption of Internet technologies 29

Table 2 Internet applications adopted by organizations 29

Table 3 Total number of Internet applications adopted 29

Table 4 Description of variables 30

Table 5 Baseline adoption model results 31

Table 6 Marginal effects of change in dependent variables on probability of adoption in baseline model 33

Table 7 Adoption model with 3-digit SIC codes 34

Table 8 Adoption model with 3-digit SIC codes and top 20 metro areas 36

Table 9 Adoption model with single-establishment organizations only 38

Table 10 Parameter estimates for count data models 40

List of Appendix tables and figures

Table A.1 Establishments per organization 44

Table A.2 Most common intranet and Internet server applications 44

Figure A.1 Number of establishments adopting Internet applications in multi-establishment organizations 45

Figure 1
Decision process for Internet adoption

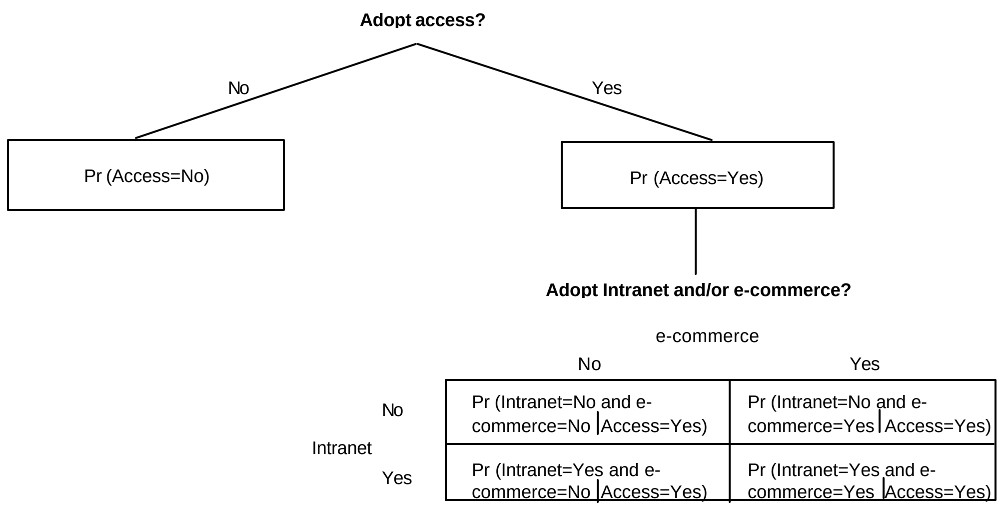


Table 1
Organization adoption of Internet technologies (%)

	Access	Intranet	E-commerce
1996	32.5	19.1	6.9
1997	63.4	38.6	19.3
1998	75.0	56.6	27.8

Note: Sample period is 1996-98, and includes only observations that were in the Harte Hanks database in 1996. Number of observations: 5389 in 1996; 4487 in 1997; and 3903 in 1998.

Source: Harte Hanks Market Intelligence and author's calculations.

Table 2
Internet applications adopted by organizations (%)

	Technologies adopted				
Year	None	Access only	Access + Intranet	Access + commerce	Access, Int., and comm.
1996	67.5	11.7	14.3	2.4	4.2
1997	36.6	19.0	25.4	6.4	12.6
1998	25.0	14.5	32.9	5.3	22.3

Note: Sample period is 1996-98, and includes only observations that were in the Harte Hanks database in 1996. Number of observations: 5389 in 1996; 4487 in 1997; and 3903 in 1998.

Source: Harte Hanks Market Intelligence and author's calculations from 1996-98 analysis sample.

Table 3
Total number of Internet applications adopted

	Time period	
Number of applications	1998	1996-98
0	72.7	51.6
1	13.1	19.8
2	7.0	11.1
3-5	6.3	13.5
6-10	0.7	3.2
over 10	0.1	0.8

Note: Number of observations=8388.

Source: Harte Hanks Market Intelligence and author's calculations.

Table 4
Description of variables

Variable	Description	Mean	Std deviation	Minimum	Maximum
Lead user					
PCPEREMP	Total organization PCs divided by total organization employees	0.5611	0.5646	0	10
CLIENT	% of establishments with Internet-ready clients	0.6779	0.4543	0	1
NOAPP	Dummy indicating no applications in organization	0.3399	0.4737	0	1
PCTLAN	% of non-PC applications accessed over LAN	0.2577	0.2687	0	1
SYSCOM	Dummy indicating presence of system LAN applications	0.2400	0.4271	0	1
NETWARE	Dummy indicating use of Netware LAN OS	0.4255	0.4944	0	1
Competing effects of installed base					
NETWARE × PCTLAN	NETWARE interacted with PCTLAN	0.1815	0.2518	0	1
SYSCOM × PCTLAN	SYSCOM interacted with PCTLAN	0.1107	0.2171	0	1
PCTMAIN	% of non-PC applications accessed over mainframe	0.1230	0.2320	0	1
MAINOUT	Dummy indicating maintenance vendor for large-scale computers	0.3043	0.4592	0	1
Geographic concentration					
MULTEST	Dummy indicating multiple establishments in organization	0.1046	0.3060	0	1
MULTEST × EMPCON	MULTEST interacted with employee Herfindahl	0.0434	0.1433	0	0.9357
MULTEST × PCT100	Percentage of pairwise establishment combinations within 100 miles in distance	0.0267	0.1411	0	1
URBAN	% of establishments that are in an MSA	0.9400	0.2316	0	1
Controls					
PROGOUT	Dummy indicating outsourcing of programming	0.1227	0.3281	0	1
EMP100-200	Log of no. of employees (between 100 and 200)	0.1726	0.2382	0	0.6931
EMP201-500	Log of no. of employees (between 201 and 500)	0.3197	0.5305	0	1.6094
EMP500+	Log of no. of employees (over 500)	0.4537	1.0815	0	6.9140
SIC60	% of establishments in SIC 60	0.0802	0.2678	0	1
SIC61	% of establishments in SIC 61	0.0234	0.1461	0	1
SIC62	% of establishments in SIC 62	0.0322	0.1734	0	1
SIC63	% of establishments in SIC 63	0.0732	0.2556	0	1
SIC64	% of establishments in SIC 64	0.0428	0.1976	0	1
SIC65	% of establishments in SIC 65	0.0468	0.2107	0	1
SIC67	% of establishments in SIC 67	0.0295	0.1604	0	1
SIC73	% of establishments in SIC 73	0.3184	0.4616	0	1
SIC87	% of establishments in SIC 87	0.1733	0.3752	0	1

Number of observations=8388

Table 5
Baseline adoption model results

	Access	Intranet	E-commerce
Lead user			
PCPEREMP	0.1892**	0.1383**	0.1684**
	(0.0324)	(0.0357)	(0.0423)
CLIENT	1.4289**	0.9091**	0.3100*
	(0.0654)	(0.2170)	(0.1642)
NOAPP	-0.5058**	-0.2124*	-0.2886**
	(0.0787)	(0.1158)	(0.1230)
PCTLAN	0.6949**	0.5277**	0.0596
	(0.1258)	(0.1278)	(0.1526)
SYSCOM	0.2000*	0.3246**	0.1602
	(0.1042)	(0.0994)	(0.1065)
NETWARE	0.3494**	0.2944**	0.0985
	(0.0720)	(0.0718)	(0.0830)
Competing effects of installed base			
NETWARE × PCTLAN	-0.5752**	-0.6789**	-0.0821
	(0.1569)	(0.1522)	(0.1817)
SYSCOM × PCTLAN	-0.4546**	-0.3824**	-0.2272
	(0.2018)	(0.1940)	(0.2177)
PCTMAIN	-0.0085	-0.0725	-0.0477
	(0.0927)	(0.0927)	(0.1096)
MAINOUT	-0.0430	0.0810	-0.0259
	(0.0503)	(0.0501)	(0.0527)
Geographic concentration			
MULTEST	1.4788**	1.4562**	0.8935**
	(0.2174)	(0.2166)	(0.1628)
MULTEST × EMPCON	-1.1805**	-1.6100**	-1.2320**
	(0.3946)	(0.3684)	(0.2714)
MULTEST × PCT100	-0.2996*	-0.3569**	-0.1355
	(0.1584)	(0.1440)	(0.1510)
URBAN	-0.1956**	-0.0604	-0.1714**
	(0.0760)	(0.0750)	(0.0828)
Controls			
PROGOUT	-0.0554	0.0296	0.1015*
	(0.0608)	(0.0586)	(0.0595)
EMP100	0.0906	0.1608*	0.1769*
	(0.0946)	(0.0922)	(0.1123)
EMP200	0.1823**	0.2108**	0.3213**
	(0.0418)	(0.0412)	(0.0489)
EMP500	0.1792**	0.2142**	0.2856**
	(0.0272)	(0.0271)	(0.0303)
SIC60	-0.4605**	-0.4265**	-0.2897**
	(0.0695)	(0.0735)	(0.0933)
SIC61	-0.1368	-0.1129	0.0336
	(0.1325)	(0.1195)	(0.1289)
SIC62	-0.3625**	-0.0577	-0.1705
	(0.1047)	(0.1181)	(0.1282)

Table 5 continues

Table 5: Description of variables
(con't)

	Access	Intranet	E-commerce
SIC63	-0.1884**	-0.1328*	-0.2159**
	(0.0774)	(0.0735)	(0.0833)
SIC64	-0.2690**	-0.1978*	-0.1160
	(0.1094)	(0.1011)	(0.1181)
SIC65	-0.4162**	-0.3287**	-0.4669**
	(0.0925)	(0.1002)	(0.1368)
SIC67	-0.4054**	-0.1150	-0.6018**
	(0.1076)	(0.1248)	(0.1375)
SIC73	-0.1919**	-0.0321	-0.0419
	(0.0547)	(0.0560)	(0.0584)
SIC87	-0.0952	0.0977	-0.1825**
	(0.0633)	(0.0635)	(0.0642)
CONSTANT	-0.7528**	-1.1729**	-1.0127**
	(0.1203)	(0.2525)	(0.2428)
r_{12}	0.8774**		
	(0.1385)		
r_{23}	0.1193		
	(0.0736)		
r_{13}	0.0900		
	(0.1631)		
Log-likelihood	-8945.21		
Number of observations	8388		

Note: ** significant at 5% level; * significant at 10% level.

Table 6
Marginal effects of change in dependent variables
on probability of adoption in baseline model

Variable	Decision		
	Access	Intranet	E-commerce
Lead user			
PCPEREMP	0.0732**	0.0542**	0.0463**
	(0.0125)	(0.0122)	(0.0083)
CLIENT	0.5251**	0.3360**	0.1544**
	(0.0203)	(0.0198)	(0.0127)
NOAPP	-0.1998**	-0.0966**	-0.0767**
	(0.0304)	(0.0263)	(0.0154)
PCTLAN	0.2491**	0.2109**	0.0652**
	(0.0413)	(0.0460)	(0.0338)
SYSCOM	0.0748**	0.1202**	0.0509**
	(0.0374)	(0.0366)	(0.0269)
NETWARE	0.1265**	0.1176**	0.0490**
	(0.0246)	(0.0263)	(0.0197)
Competing effects of installed base			
NETWARE × PCTLAN	-0.2153**	-0.2447**	-0.0656**
	(0.0588)	(0.0499)	(0.0372)
SYSCOM × PCTLAN	-0.1751**	-0.1542**	-0.0829*
	(0.0767)	(0.0683)	(0.0451)
PCTMAIN	-0.0033	-0.0244	-0.0095
	(0.0358)	(0.0322)	(0.0215)
MAINOUT	-0.0166	0.0236	-0.0082
	(0.0195)	(0.0169)	(0.0104)
Geographic concentration			
MULTEST	0.3568**	0.5078**	0.3864**
	(0.0241)	(0.0453)	(0.0576)
MULTEST × EMPCON	-0.2672**	-0.5648**	-0.4199**
	(0.0923)	(0.1040)	(0.0748)
MULTEST × PCT100	-0.0655*	-0.1363**	-0.0614
	(0.0378)	(0.0549)	(0.0473)
URBAN	-0.0731**	-0.0284	-0.0524**
	(0.0277)	(0.0271)	(0.0204)

Note: Number of observations=8388;
** significant at 5% level; * significant at 10% level;
Marginal effects are calculated by changing variables from 0 to 1.

Table 7
Adoption model with 3-digit SIC codes

	Access	Intranet	E-commerce
Lead user			
PCPEREMP	0.1741**	0.0986	0.1449**
	(0.0345)	(0.0452)	(0.0455)
CLIENT	1.4527**	0.7661**	0.2750
	(0.0681)	(0.3107)	(0.1935)
NOAPP	-0.4771**	-0.0639	-0.2780**
	(0.0813)	(0.1625)	(0.1376)
PCTLAN	0.7018**	0.4524**	0.0836
	(0.1300)	(0.1554)	(0.1595)
SYSCOM	0.1867*	0.3137**	0.1437
	(0.1067)	(0.1071)	(0.1083)
NETWARE	0.3476**	0.2761**	0.0957
	(0.0733)	(0.0838)	(0.0864)
Competing effects of installed base			
NETWARE × PCTLAN	-0.5923**	-0.6748**	-0.0973
	(0.1604)	(0.1686)	(0.1876)
SYSCOM × PCTLAN	-0.4446**	-0.3214	-0.2292
	(0.2074)	(0.2148)	(0.2221)
PCTMAIN	-0.0172	-0.1122	-0.0648
	(0.0942)	(0.1020)	(0.1123)
MAINOUT	-0.0434	0.1086*	-0.0362
	(0.0512)	(0.0563)	(0.0541)
Geographic concentration			
MULTEST	1.4882**	1.3956**	0.8458**
	(0.2224)	(0.2694)	(0.1730)
MULTEST × EMPCON	-1.1877**	-1.6439**	-1.2346**
	(0.4015)	(0.4046)	(0.2775)
MULTEST × PCT100	-0.2964*	-0.3572**	-0.1390
	(0.1628)	(0.1562)	(0.1542)
URBAN	-0.1991**	-0.0222	-0.1797**
	(0.0783)	(0.0855)	(0.0841)
Controls			
PROGOUT	-0.0707	0.0389	0.0880
	(0.0630)	(0.0658)	(0.0613)
EMP100	0.0667	0.1623*	0.1288
	(0.0916)	(0.0951)	(0.1071)
EMP200	0.1834**	0.2228**	0.3082**
	(0.0278)	(0.0305)	(0.0324)

Table 7 continues

Table 7: Adoption model with 3-digit SIC codes
(con't)

	Access	Intranet	E-commerce
CONSTANT	-0.5461**	-1.1302**	-0.9835**
	(0.1483)	(0.3785)	(0.2846)
r_{12}	0.6825**		
	(0.2765)		
r_{23}	0.1130*		
	(0.0693)		
r_{13}	0.0829		
	(0.2128)		
Number SIC dummies	48		
Log-likelihood	-8850.69		
Number of observations	8378		

Note: ** significant at 5% level; * significant at 10% level.

Table 8
Adoption model with 3-digit SIC codes
and top 20 metro areas

	Access	Intranet	E-commerce
Lead user			
PCPEREMP	0.1838**	0.0729	0.1554**
	(0.0350)	(0.0497)	(0.0472)
CLIENT	1.4542**	0.6541**	0.2594
	(0.0701)	(0.3010)	(0.2027)
NOAPP	-0.5027**	0.0228	-0.3006**
	(0.0834)	(0.1667)	(0.1491)
PCTLAN	0.7102**	0.4482**	0.0794
	(0.1324)	(0.1617)	(0.1656)
SYSCOM	0.1763	0.3170**	0.1450
	(0.1101)	(0.1141)	(0.1104)
NETWARE	0.3594**	0.2556**	0.1003
	(0.0762)	(0.0883)	(0.0888)
Competing effects of installed base			
NETWARE × PCTLAN	-0.6158**	-0.6363**	-0.1058
	(0.1658)	(0.1790)	(0.1935)
SYSCOM × PCTLAN	-0.4251**	-0.2892	-0.2276
	(0.2145)	(0.2273)	(0.2265)
PCTMAIN	-0.0363	-0.0812	-0.0587
	(0.0970)	(0.1073)	(0.1153)
MAINOUT	-0.0449	0.1034*	-0.0416
	(0.0531)	(0.0572)	(0.0556)
Geographic concentration			
MULTEST	1.4864**	1.3229**	0.8466**
	(0.2350)	(0.2892)	(0.1805)
MULTEST × EMPCON	-1.1864**	-1.6754**	-1.2347**
	(0.4231)	(0.4478)	(0.2843)
MULTEST × PCT100	-0.2717	-0.3233*	-0.1389
	(0.1778)	(0.1735)	(0.1594)
URBAN	-0.0858	-0.0271	-0.1469
	(0.0819)	(0.0857)	(0.0870)
Controls			
PROGOUT	-0.0934	0.0572	0.0847
	(0.0649)	(0.0697)	(0.0631)
EMP100	-0.0015	0.2554*	0.2400
	(0.1476)	(0.1535)	(0.1646)
EMP200	0.1709**	0.2022**	0.3132**
	(0.0433)	(0.0456)	(0.0438)

Table 8 continues

Table 8: Adoption model with 3-digit SIC codes and top 20 metro areas
(con't)

	Access	Intranet	E-commerce
CONSTANT	-0.5238** (0.1546)	-0.9622** (0.3882)	-0.9876 (0.2964)
r_{12}	0.5755** (0.2825)		
r_{23}	0.1182* (0.0612)		
r_{13}	0.0898 (0.2291)		
Number SIC dummies	48		
Number of metro area dummies	20		
Number of variables that interact metro area dummies with employment	40		
Log-likelihood	-8730.63		
Number of observations	8378		

Note: ** significant at 5% level; * significant at 10% level.

Table 9
Adoption model with single-establishment organizations only

	Access	Intranet	E-commerce
Lead user			
PCPEREMP	0.1912**	0.1611**	0.1697**
	(0.0333)	(0.0366)	(0.0457)
CLIENT	1.4327**	0.9021**	0.2930
	(0.0683)	(0.2752)	(0.2229)
NOAPP	-0.5085**	-0.2129*	-0.2644**
	(0.0820)	(0.1210)	(0.1293)
PCTLAN	0.6940**	0.5121**	0.0500
	(0.1281)	(0.1303)	(0.1621)
SYSCOM	0.1874*	0.3015**	0.1792
	(0.1072)	(0.1029)	(0.1182)
NETWARE	0.3534**	0.2911**	0.1055
	(0.0737)	(0.0752)	(0.0888)
Competing effects of installed base			
PCTMAIN	-0.0305	-0.0750	-0.0359
	(0.0949)	(0.0954)	(0.1172)
NETWARE × PCTLAN	-0.5843**	-0.6815**	-0.0951
	(0.1597)	(0.1556)	(0.1918)
SYSCOM × PCTLAN	-0.4604**	-0.3777*	-0.2295
	(0.2064)	(0.1990)	(0.2399)
MAINOUT	-0.0609	0.0740	-0.0409
	(0.0515)	(0.0524)	(0.0569)
Geographic concentration			
URBAN	-0.1833**	-0.0583	-0.1619**
	(0.0770)	(0.0758)	(0.0846)
Controls			
PROGOUT	-0.0579	-0.0012	0.0767
	(0.0642)	(0.0613)	(0.0681)
EMP100	0.0753	0.1373	0.1468
	(0.0906)	(0.0888)	(0.1095)
EMP200	0.1681**	0.1981**	0.2860**
	(0.0284)	(0.0286)	(0.0348)
SIC60	-0.4478**	-0.4414**	-0.3173**
	(0.0719)	(0.0764)	(0.1053)
SIC61	-0.1362	-0.1202	0.0273
	(0.1370)	(0.1262)	(0.1410)
SIC62	-0.3735**	-0.0551	-0.1798
	(0.1059)	(0.1279)	(0.1395)
SIC63	-0.2018**	-0.1244	-0.2424**
	(0.0792)	(0.0763)	(0.0890)

Table 9 continues

Table 9: Adoption model with single-establishment organizations only (con't)

	Access	Intranet	E-commerce
SIC64	-0.2858**	-0.2019*	-0.1068
	(0.1125)	(0.1058)	(0.1281)
SIC65	-0.4077**	-0.3274**	-0.4810**
	(0.0952)	(0.1026)	(0.1456)
SIC67	-0.3835**	-0.1367	-0.5942**
	(0.1126)	(0.1274)	(0.1474)
SIC73	-0.1856**	-0.0275	-0.0261
	(0.0561)	(0.0589)	(0.0622)
SIC87	-0.1160*	0.0906	-0.1754**
	(0.0647)	(0.0677)	(0.0684)
CONSTANT	-0.7392**	-1.1585**	-0.9867
	(0.1226)	(0.3044)	(0.3110)
r_{12}	0.8927**		
	(0.1516)		
r_{23}	0.1085		
	(0.0892)		
r_{13}	0.0875		
	(0.1922)		
Log-likelihood	-8047.33		
Number of observations	7479		

Note: ** significant at 5% level; * significant at 10% level.

Table 10
Parameter estimates for count data models

			ZIP models	
	Poisson	Negative binomial	ZIP Poisson	ZIP negative binomial
Lead user				
PCPEREMP	0.1730*	0.2222**	0.2199**	0.2511**
	(0.0277)	(0.0459)	(0.0356)	(0.0530)
CLIENT	0.8178**	1.0062**	-0.1023	-0.1201
	(0.0825)	(0.1028)	(0.0939)	(0.1187)
NOAPP	-0.5658**	-0.4702**	-0.2765**	-0.2469*
	(0.0922)	(0.1164)	(0.1093)	(0.1335)
PCTLAN	0.4499**	0.4043**	-0.1641	-0.1825
	(0.1174)	(0.1712)	(0.1617)	(0.2015)
SYSCOM	-0.1562*	-0.1760	-0.1217	-0.1169
	(0.0806)	(0.1273)	(0.1039)	(0.1370)
NETWARE	0.3278**	0.3301**	0.1240	0.1284
	(0.0654)	(0.0941)	(0.0848)	(0.1074)
Competing effects of installed base				
NETWARE × PCTLAN	-0.6922**	-0.6953**	-0.2283	-0.2672
	(0.1429)	(0.2072)	(0.1931)	(0.2426)
SYSCOM × PCTLAN	0.2803*	0.3188	0.5047**	0.6091**
	(0.1639)	(0.2486)	(0.2173)	(0.2848)
PCTMAIN	0.1838**	0.2394*	-0.0058	-0.0560
	(0.0847)	(0.1251)	(0.1148)	(0.1514)
MAINOUT	-0.0157	-0.0122	0.1160**	0.1442**
	(0.0409)	(0.0618)	(0.0534)	(0.0696)
Geographic concentration				
MULTEST	0.6406**	0.6678**	0.2199**	0.2111
	(0.0852)	(0.1518)	(0.1005)	(0.1403)
MULTEST × EMPCON	-0.5238**	-0.4301	-0.4771**	-0.3771*
	(0.1437)	(0.2689)	(0.1623)	(0.2236)
MULTEST × PCT100	-0.2848**	-0.2550	-0.2729**	-0.2789*
	(0.1029)	(0.1711)	(0.1339)	(0.1710)
URBAN	-0.1652**	-0.1543	-0.2209**	-0.2244**
	(0.0679)	(0.0958)	(0.0870)	(0.1117)
Learning and experimentation				
DPACC	-0.1472**	-0.1518**	-0.1446**	-0.1754**
	(0.0370)	(0.0536)	(0.0474)	(0.0610)

Table 10 continues

Table 10: Parameter estimates for count data models
(con't)

			ZIP models	
	Poisson	Negative binomial	ZIP Poisson	ZIP negative binomial
Controls				
PROGOUT	0.0723*	0.0361	0.0109	0.0408
	(0.0414)	(0.0686)	(0.0524)	(0.0728)
EMP100	0.3618**	0.3833**	0.3020**	0.3184**
	(0.0825)	(0.1163)	(0.1123)	(0.1398)
EMP200	0.3615**	0.3729**	0.2450**	0.2560**
	(0.0187)	(0.0314)	(0.0230)	(0.0322)
SIC60	-0.2239**	-0.2801**	-0.0464	-0.0876
	(0.0676)	(0.1003)	(0.0917)	(0.1229)
SIC61	-0.1231	-0.1744	-0.1840	-0.1955
	(0.1064)	(0.1582)	(0.1447)	(0.1778)
SIC62	0.1580*	0.1433	0.2072**	0.2276*
	(0.0827)	(0.1312)	(0.1016)	(0.1379)
SIC63	-0.0101	-0.0390	-0.0602	-0.1133
	(0.0612)	(0.0951)	(0.0809)	(0.1066)
SIC64	-0.1203	-0.2888**	-0.1100	-0.1729
	(0.0938)	(0.1359)	(0.1263)	(0.1600)
SIC65	-0.3059**	-0.3833**	0.0730	0.0526
	(0.1068)	(0.1394)	(0.1460)	(0.1827)
SIC67	-0.1991**	-0.2023	-0.1748	-0.1758
	(0.1013)	(0.1469)	(0.1387)	(0.1715)
SIC73	0.0247	-0.0383	0.1734**	0.1784**
	(0.0465)	(0.0671)	(0.0612)	(0.0783)
SIC87	0.0203	0.0061	0.0434	0.0234
	(0.0508)	(0.0750)	(0.0672)	(0.0861)
CONSTANT	-1.6922**	-1.8799**	0.2347	0.0142
	(0.1213)	(0.1649)	(0.1542)	(0.2035)
a		1.6245**		0.3985**
		(0.0785)		(0.0640)
Vuong			15.62	8.47
Log-likelihood	-8250.534	-7458.292	-7366.50	-7291.20
Number of observations	8388	8388	8388	8388

Note: ** significant at 5% level; * significant at 10% level.

Data appendix

A1 Unit of observation

The unit of observation in the CI database is an establishment/year. Roughly speaking, an establishment refers to a particular branch or location of a firm. It is similar to the concept of establishment used by government organizations such as the Bureau of Labor Statistics. Thus, the database will often have data on multiple establishments for a given firm.

The establishment is an inappropriate unit of observation because the technology adoption decision for an establishment is likely to depend on observable and unobservable attributes of other establishments within the same organization. Appendix Table A.1 and Figure A.1 illustrate the potential dangers associated with establishment-level analysis. Appendix Table A.1 presents the distribution of number of establishments per organization. The vast majority of organizations have only one or two establishments, although a significant percentage of establishments (over 10 per cent) are part of multi-establishment organizations. To show the manner in which multi-establishment organizations adopt, Appendix Figure A.1 shows the distribution of number of establishments adopting Internet technologies among multi-establishment organizations that have adopted some form of access. The distribution of total number of establishments in multi-establishment organizations among this subsample is included for comparison purposes. The Figure shows that even among multi-establishment organizations, many adopt Internet applications at only one location. The mode number of establishments adopting access, intranet, and e-commerce is one for all three technologies. In particular, for e-commerce, over 60 per cent of multi-establishment organizations adopt e-commerce at only one location. Although this graph is of course partially capturing within-firm diffusion, the high percentage of single-establishment adopters suggests that there is some danger that adoption decisions are being made at the firm rather than establishment level, and that the establishment may be an inappropriate unit of observation. To avoid these problems, I conduct all analyses of the paper at the organization level.

I define an organization in a year as the aggregation of all establishments within a firm that have been in the CI database for two consecutive years. The requirement that the establishment be in the database for two consecutive years is required to obtain independent variables for the analysis, as prior year variables on organization characteristics are used to determine adoption decisions.

The measure of organization will not, in general, correspond exactly to a firm. This is true for several reasons. First, the organizations in the dataset may consist of only a subset of the industries in a particular firm. This is because my sample is constructed from particular industries within the CI database. An analysis of the organizations in the dataset suggests that most are clustered in a small number of SIC codes, however. Ninety-four point nine per cent of the organizations in the dataset have business activities in only one two-digit SIC code, while another 3.4 per cent conduct business in only two. Second, the CI Technology database does not, in general, sample all the establishments from a firm. Entry and exit of establishments may also change the composition of establishments in an organization across two different years in the sample.

A2 Adoption variables

The CI database includes several measures of establishment use of Internet technologies. The first measure of Internet use contains data on an establishment's Internet Service Provider (ISP). Establishments that have responded to this survey by indicating use of an ISP are counted as adopting access. The second measure of Internet use is a direct survey on adoption of Internet technologies. Using these data, establishments are identified as adopting intranet if they responded positively in this survey to adopting either TCP/IP-based e-mail or intranet. An establishment is counted as adopting e-commerce if it responded positively to adopting any of the following: business-to-business electronic commerce, business-to-consumer electronic commerce, electronic commerce, customer service, education, extranet, publishing, purchasing, or technical support.

A3 Number of Internet applications adopted

The CI database contains three separate measures of the intensity of Internet usage within an organization. The first survey, described in the previous section on adoption variables, describes an establishment's use of the following technologies: business-to-business electronic commerce, business-to-consumer electronic commerce, electronic commerce, customer service, education, e-mail, extranet, homepage, intranet, publishing, purchasing, research, and technical support. I refer to this as the survey on basic Internet technologies. The CI database also contains information on an establishment's use of two types of Internet applications: Internet server applications and intranet applications. These last two categories differ from the survey on basic Internet technologies in that they contain the specific uses to which Internet technologies have been put, and are generally associated with more sophisticated uses of Internet technologies. Appendix Table A.2 lists the most common intranet and Internet server applications in use at establishments in 1998.

To derive a measure of number of Internet applications adopted, I sum the number of intranet and Internet server applications adopted with the number of basic Internet technologies adopted in the organization. Applications that are duplicated across surveys or across establishments within the same organization are deleted. Thus, the calculation of total number of applications adopted is the sum of unique applications across the three surveys and across all establishments within an organization.

There are two major reasons why I have combined applications from heterogeneous categories to arrive at a measure of the intensity of Internet adoption. First, the survey on basic Internet technologies is combined with the other two surveys to achieve a combined tally of 'basic' (basic Internet technologies) and 'advanced' (intranet and Internet server applications) Internet applications. Second, surveys on intranet and Internet server technologies are combined because of the difficulty of developing a useful way of classifying these applications separately. Many applications listed in the database as Internet server applications could also be used as intranet applications, making it difficult to separate applications into exclusive categories. For instance, e-mail is one of the most popular responses in all three surveys: basic Internet technologies, intranet applications, and Internet server applications. In the face of this uncertainty, it is best to take a conservative approach and combine classes of applications together.

Appendix Table A.1
Establishments per organization

Number of establishments	%
1	88.6
2	4.9
3-5	3.4
6-10	1.6
11-25	1.1
26-50	0.4
51-100	0.1
over 100	0.0

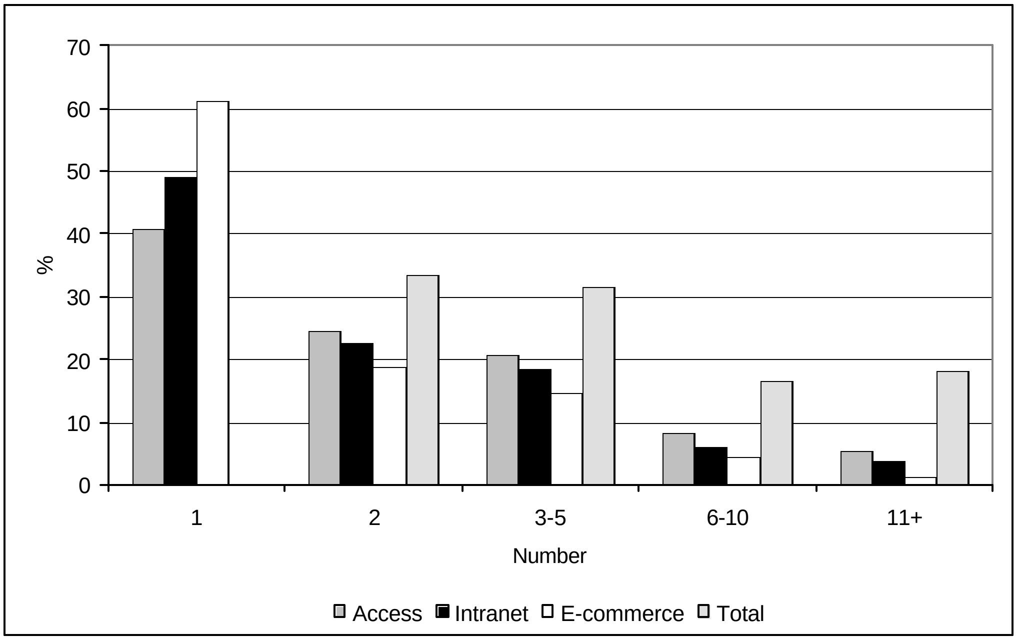
Note: Sample period is 1996-98.
Source: Harte Hanks Market Intelligence and author's calculations.

Appendix Table A.2
Most common intranet and Internet server applications

Intranet		Internet	
Application	Frequency	Application	Frequency
e-mail	921	e-mail	666
communications	417	web server	470
info-centre	175	homepage	223
data sharing	160	research	205
personnel	154	file transfer	201
documents	111	marketing	111
file-transfer	73	customer service	98
research	72	information centre	95
home page	60	advertising	78
database	51	firewall	78
bulletin board	48	database	58
business	44	EDI	43
document management	33	personnel	41
data transfer	27	sales	38
EDI	27	communications	34
technical support	26	development	34
documentation	23	file sharing	22
accounting	22	newspaper	20
file sharing	22	insurance	19
training	21	business	18
finance	18	accounting	17
sales	17	finance	17
customer service	16	publishing	16
publishing	16	technical support	16
banking	13	downloading	15

Source: Harte Hanks Market Intelligence and author's calculations.

Appendix Figure A.1
Number of establishments adopting Internet applications
in multi-establishment organizations



Notes: Sample period is 1996-98. Based on sample of organizations that have adopted basic access. Access, intranet, and e-commerce show distribution of number of establishments with these technologies among organizations that have adopted the technology. Total indicates the distribution of number of establishments among organizations that have adopted access.

Source: Harte Hanks Market Intelligence and author's calculations.