NET Institute*

www.NETinst.org

Working Paper #07-05

April 2007

Lock-In and Unobserved Preferences in Server Operating System Adoption: A Case of Linux vs. Windows

Seung-Hyun Hong, University of Illinois

Leonardo Rezende, PUC-Rio and University of Illinois

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

Lock-In and Unobserved Preferences in Server Operating System Adoption: A Case of Linux vs. Windows*

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

Leonardo Rezende
Department of Economics
PUC-Rio and University of Illinois
lrezende@econ.puc-rio.br

April 9, 2007

Abstract

This paper attempts to distinguish state dependence (or lock-in) from unobserved preferences in the decision to adopt Linux or Windows as the operating system for computer servers. To this end, we use detailed survey data of over 100,000 establishments in the United States. Without accounting for unobserved heterogeneity in establishment-specific preferences for operating systems, we find a strong positive correlation between the current choice and the previous choice, suggesting potentially high switching costs and lock-in. To account for unobserved preferences for either operating system, we impose weak identifying assumptions and employ recently developed dynamic discrete choice panel data methods (Arellano and Carrasco 2003). The results show little or no evidence of state dependence, implying that unobserved preferences, rather than switching costs and lock-in, are more important factors in the adoption decision. Once taste heterogeneity is taken into account, we additionally find little evidence of network effects between server operating systems and non-server operating systems.

*We appreciate the NET Institute for financial support. We thank George Deltas, Mark Jacobsen, Roger Koenker, and Zhongjun Qu for helpful discussion and comments. All remaining errors are our responsibility. Contact information: Seung-Hyun Hong: 470F Wohlers Hall, 1206 S. Sixth St., Champaign, IL 61820; Leonardo Rezende: Rua Marquês de São Vicente, 225 SL. 210F, Rio de Janeiro RJ Brazil 22453-900.

1 Introduction

The persistent dominance of Microsoft's Windows is often claimed to be due to high switching costs and lock-in, suggesting potential inefficiency. By contrast, it is also argued that consumers continue to use Windows because of its superior quality, indicating potential efficiency in the operating system market. Despite its importance in the debate over the antitrust case against Microsoft,[1] however, few empirical studies have attempted to separate out lock-in from superior quality, partly because of lack of data, but also because of the difficulty in identification – both lock-in and superior quality imply a strong positive correlation between the current choice and the previous choice in computer operating systems.[2]

In this paper, we examine establishments' decision to adopt operating systems for computer servers. In particular, we focus on Linux and Windows, and empirically distinguish state dependence (or lock-in) from establishment-specific unobserved preferences for either operating system. To this end, we use unbalanced panel data from 2000-2004 *Computer Intelligence Technology Database* (CITDB) collected by the Harte-Hanks Market Intelligence. The CITDB surveys over 100,000 establishments in the United States every year. It contains detailed information on establishment characteristics and ownership of information technologies such as operating systems for various computers. This detailed information allows us to examine establishment-level decisions to adopt server operating systems.

As expected, the descriptive statistics from the CITDB show significant positive correlations between the current choice of Windows and the previous choice of Windows, and similarly for Linux. We then estimate probit models for the adoption of server operating systems, allowing for various other factors to determine the adoption decision. Without accounting for unobserved heterogeneity, however, we still find strong state dependence in both Windows and Linux. This positive correlation is robust even after we control for various observed

[1] See, e.g., Bresnahan (2001) and Liebowitz and Margolis (1999).

[2] This problem reflects the well-known difficulty in identification between state dependence and unobserved heterogeneity. See, e.g., Heckman (1981a, 1981c), Hsiao (2003), and also Greenstein (1993).

heterogeneity. Nonetheless, we cannot interpret this result as evidence for lock-in in server operating systems, as it is also consistent with the importance of unobserved preferences for either operating system.

Therefore, we consider several existing econometric approaches to account for unobserved heterogeneity in discrete choice models. Most approaches including standard random effects models and conditional logit fixed effect models, however, impose very strong assumptions that are unlikely to hold in our application. Consequently, we employ a dynamic discrete choice panel data method developed by Arellano and Carrasco (2003), which addresses the preceding empirical challenge with weak assumptions in our context. The key identifying assumption is to assume that the quality of each operating system is not fully observed by establishments, so that at each period, establishments revise their assessment of the quality based on their history up to the current period. Given this assumption, we can invert the probit equation to obtain a moment condition,[3] which allows us to use a GMM method.

We apply this method to ten subsamples in the CITDB which we group based on the NAICS code. For most subsamples, we find little or no correlation between the current choice and the previous choice in the adoption of server operating systems. This result suggests that the adoption of Windows (or Linux) at the previous period does not necessarily lock establishments into Windows (or Linux) at the present period. Rather than switching costs and lock-in, establishment-specific unobserved preferences based on the assessment of unknown quality seem to be more important in the adoption decision. We additionally find little correlations between the current choice of server operating systems and the previous choice of non-server operating systems, providing little evidence for network effects (or benefits from the compatibility) between server operating systems and non-server operating systems.

These findings can be interpreted as evidence that superior quality, or at least perceived

[3]The moment condition is based on a law of iterated expectation, which implies that conditional on the same history, the expected value of the future assessment of product quality should be the same as the current assessment.

better quality, explains the persistent dominance of Windows in the operating system market. Equivalently, perceived better quality also explains the increasing popularity of Linux in server operating systems. These findings, nevertheless, may also indicate that true state dependence and lock-in might not be measured properly by simply estimating the coefficient for the adoption of Windows (or Linux) at the previous period, thus calling for further modeling.[4]

The paper is organized as follows. Section 2 summarizes various well-known factors explaining the adoption of operating systems. Section 2 also provides a short description of open source software and Linux adoption as another motivation for our study. Section 3 describes our data and provides descriptive statisics. Section 4 presents the main empirical method and discusses identifying assumptions. Section 5 reports the results from our main method and contrasts them with those obtained by using probit methods. Section 6 concludes the paper.

2 Background

2.1 Server Operating Systems Adoption

In this paper, we consider establishments' decision[5] to adopt operating systems for internet servers or network servers. In essence, a firm will choose a particular operating system for its server, if the net benefit from the operating system is higher than that from other operating systems. There are several factors that may determine benefits and costs associated with adopting a particular operating system. First, high switching costs might lock firms into the previous investment in Windows or Linux. Because an information system in business computing environments is composed of various interrelated components including computer hardware and networks, database, and application software, it would be difficult to change one component such as the operating system without changing other components in the information

[4]We pursue structural modeling in our future work.

[5]In this paper, we consider the operating system adoption by establishments, but not necessarily by firms, since a firm may own multiple establishments, that is, branches in different locations. For convenience, nevertheless, we use both terms – establishments and firms – interchangeably. At least for about a half of our samples from the CITDB, however, this is indeed correct, as they are single facility firms.

system. As a result, firms may continue to use the same operating system as before, unless there are strong reasons to reorganize the existing architecture. For this reason, the choice of the operating system in the previous period may be important, reflecting high switching costs.

Second, firms may have heterogeneous preferences over different operating systems, depending on firm characteristics or their assessment of the quality of operating systems. For example, a firm with in-house software programmers may have a strong preference for the open access nature of the Linux architecture because they need to customize the operating system for their special needs. Because of considerable complexity in operating system software, on the other hand, even users of operating systems may not know the true quality of operating systems, and thus need to assess the quality based on their experiences. Accordingly, observed heterogeneity or unobserved quality assessment may also determine benefits and possibly costs associated with adopting operating systems.

The main focus of this paper is to distinguish between lock-in due to high switching costs and unobserved preferences particularly based on the quality of operating systems. Other factors, nevertheless, might be important as well. First, there might be direct network effects between different computers within a sever segment as well as across different segments. Especially in the adoption of operating systems for servers, firms may experience network effects between the operating systems in a personal computer (PC) segment and the operating systems in the server segment. For example, if a firm uses only Windows for all the PCs, then the value of using Windows on servers might be higher because of the compatibility between PCs and servers. Hence, the operating systems choice in PCs may affect the adoption of Linux in servers. Second, indirect network effects may also influence the adoption of operating systems. In particular, Linux was the key component of the "LAMP" server software combination.[6] For instance, a firm may adopt Linux for Web servers in order to run Apache. This suggests that

[6]The LAMP refers to a set of open source software programs commonly used together to run Web sites or servers. The LAMP stands for Linux, Apache, MySQL, and Perl/PHP/Python. See http://en.wikipedia.org/wiki/Linux, and http://en.wikipedia.org/wiki/LAMP_%28software_bundle%29, both accessed September 7, 2006.

the adoption of some application software (e.g. Apache, MySQL, or Perl) may be another factor behind the adoption of Linux. Though we cannot fully account for all other factors in this paper, we attempt to consider some of them in our empirical analysis.

2.2 Linux Adoption

In addition to the continuing dominance of Windows, the increasing popularity of Linux also motivates our study. Linux is open source software. In contrast to proprietary software such as Windows, open source software provides users with free access to both source code and object code.[7] Furthermore, open source code is usually released under special licenses such as the GNU Public License (GPL) which allow users to freely use, copy, or modify the source code, provided that the modifications are also made publicly available. This has allowed numerous open source projects to flourish in the recent decade, of which Linux is the most widely known.

Linux has become an effective alternative in the operating systems market in the 1990s, but its development dates back to 1983, when Richard Stallman started the GNU project to develop a complete Unix-like operating system with open source code.[8] By the early 1990s, open source versions of most components of a Unix system were ready, except for the kernel, the lowest level software at the heart of the system. Meanwhile, Linus Torvalds started to develop a kernel of a Unix system for Intel-based personal computers, releasing the first version of the Linux kernel in 1991. Due probably to its modular structure[9] and the diffusion of the Internet, Linux became a popular open source project. It was eventually adapted for the GNU system, thereby creating a complete operating system.[10]

[7]Source code is the code written in human-readable computer programming language such as C, whereas object code can be executed by a computer, but is not readable. Most proprietary software provides users with only the object code, thus preventing users from modifying the program.

[8]The name GNU stands for "GNU's Not Unix", a recursive pun intended to stress the fact that this new clone of the Unix system had something different.

[9]A module is a self-contained component of a computer system, and an operating system is modular if it uses modules that can be interchanged as units without affecting the rest of the operating system. This feature allows developers to debug problems easily and improve a part of the program without understanding the entire structure. See Fink (2002) for more technical details.

[10]The system as a whole is sometimes referred to as GNU/Linux. For simplicity this paper uses Linux to indicate an operating system that consists of the Linux kernel, the GNU components, and possibly other open

Linux has drawn attention from many observers of the computer software market not only because it is successful open source architecture, but also because it could be a viable alternative to replace Microsoft's Windows, especially in the server market. The adoption of Linux seems to be increasing over time, but the open source nature of Linux makes it difficult to obtain official statistics of the exact magnitude of Linux adoption. Unlike proprietary software, there is no single producer who controls the production of Linux and knows the exact number of copies sold. Moreover, some companies may purchase only one copy of Linux and customize it in order to reuse it throughout the companies, in which case one shipped copy of Linux may be equivalent to thousands or more of Linux in use. Many consumers also obtain Linux by downloading it from various web sites for free. For this reason, it is practically impossible to know the exact number of copies of Linux in use. Nevertheless, there have been attempts to estimate the adoption rate,[11] and most estimates point to the increasing rate of Linux adoption, particularly in server computers.[12] This growing popularity of Linux has further invigorated or coincided with the development of other open source projects, some of which can even substitute other Microsoft products including database, web server, or web browser. Consequently, it is not entirely impossible that Linux might eventually supplant Windows as the leading operating system. To assess the extent to which Linux may weaken the Windows dominance, we study underlying forces behind the adoption of Linux as well as Windows, focusing on lock-in and unobserved preferences in particular.

---

source, non-GNU code. (Because there are other components such as X11 graphic interface system, desktop interfaces, and file sharing interoperability tools, which are often delivered with Linux as well but are not GNU components, the name GNU/Linux may not be entirely precise, either.)

[11] A nonprofit organization named the Linux Counter Project has counted the number of Linux users based on voluntary registrations of Linux users on its web site at http://counter.li.org/. A for-profit research firm named IDC has published reports called "Worldwide Linux Operating Environments Forecast and Analysis" for 2002-2006 (IDC report #27521) and 2005-2009 (IDC report #34390) that estimate the rate of Linux server unit shipments.

[12] Some articles (e.g., Varian and Shapiro 2003), books (e.g., Fink 2002), and web sites (e.g., Wikipedia) cite the IDC report #27521 for high growth rate of Linux adoption in computer servers. In contrast, Linux adoption in non-servers (mostly, personal computers) appears to be still limited. According to the IDC report #202388, titled "Worldwide Operating Systems and Subsystems 2005 Vendor Shares", Linux accounts for 1.4% of the overall operating systems and subsystems in 2005, compared to Microsoft's 71.4% share, despite robust growth in Linux adoption (refer to abstract of this report).

3 Data and Descriptive Statistics

3.1 Data Description

We use data from the *Computer Intelligence Technology Database* (CITDB) collected by Harte-Hanks Market Intelligence. The CITDB is an yearly survey of over 100,000 establishments in the United States. It contains detailed establishment-level data on the use of a variety of information and communication technologies. This dataset has been used in several papers (e.g. Bresnahan and Greenstein 1996; Bresnahan, Brynjolfsson, and Hitt 2002). For our study, we were able to obtain 2000-2004 data.[13] The CITDB is useful for our purpose because it contains detailed information on establishment characteristics and ownership of computer hardware and software including operating systems. The unit of observation is an establishment in a year. The CITDB has attempted to survey the same establishment each year, so that the data set contains panel information of many establishments. Because the survey is voluntary, however, some establishments did not respond to survey requests, and the CITDB has added new establishments each year. As a result, the total number of observations remains similar each year,[14] but many establishments were not surveyed in every year.

We study the adoption of operating systems at the segment level. The CITDB groups computers into four segments: internet servers; network servers; personal computers, not used for either internet servers or network servers; and non-PCs not used for servers. In this paper, we consider three mutually exclusive segments: server, including both internet servers and network servers[15]; PC, including personal computers that are used for standalone desktops or

[13] Harte-Hanks releases a new dataset every January, containing information collected the previous year. Our reference year is the collection year, not the year of release; e.g. the 2000 dataset was released January 2001.

[14] From 2003 on, the CITDB actually increased the number of observations in order to include more small establishments in its survey (e.g. 264,595 in 2002 to 482,933 in 2003). However, note that the CITDB collects two kinds of information on the use of technology: *standard*, meaning rather limited information such as the number of workstations in the establishment; *enhanced*, meaning detailed product information such as brands of workstations or operating systems on internet server. The newly added observations report only standard information and thus cannot be used in our analysis. Only the observations with enhanced information are relevant for our study, and the number of such observations remains about 120,000 in each year.

[15] We combine internet servers and network servers for two reasons: to simplify our analysis and to increase the size of samples with any kind of server.

client computers connected to servers[16]; and non-PC, including mainframes, midrange, and workstations that are not used for servers. Note that we can only investigate the adoption of operating systems up to the segment level, since the information on operating system choices at the individual computer level is not available in the CITDB. In other words, we observe which kinds of operating systems are used for computers in each segment, but we do not know exactly which operating system is running on each individual computer. The segment-level information, nonetheless, is valuable because most establishments in the CITDB tend to use only one kind of operating system for each segment[17] and many of them use only a small number of computers for each segment, except the PC segment.

Table 1 presents summary statistics of variables used in our analysis. We use Windows to denote Windows-family operating systems such as Windows 95, 98, ME, NT, 2000, 2003, and XP. Linux indicates not only various versions of Linux (e.g. Debian, Red-Hat, Mandrake, SuSE, etc.) but also Berkeley Software Distribution (BSD).[18] We use other to denote other operating systems including Mac OS X as well as a variety of proprietary Unix (e.g. Solaris, HP-UX, AIX). Because we consider three segments, we use the following notations to denote the choice of operating systems on each segment: server.linux for Linux on the server segment; pc.linux for Linux on the PC segment; non-pc.linux for Linux on the non-PC segment.[19]; and similarly for server.windows, pc.windows, and non-pc.windows.

At least three observations emerge from Table 1. First, Windows is dominant in the PC segment as well as in most server segments, except the non-PC segment where other operating system is the most popular, probably because most non-PCs are IBM computers running IBM operating systems. Note also that the adoption of Windows has increased in most server

[16]Some PCs can be used as servers, but such PCs are included in the server segment in this paper.

[17]This suggests potentially strong network effects within the same segment.

[18]BSD is the Unix derivative developed by the University of California, Berkeley. BSD is not Linux and follows its own licensing agreement different from the GPL. Nevertheless, we include BSD in the Linux category, because BSD is similar to Linux in that it is a Unix-like operating system and is available for free. The percentage of establishments using BSD, however, is negligible in our data.

[19]We also use internet.sv.linux for Linux on the Internet server segment, network.sv.linux for Linux on the network server segment.

segments. This may suggest that potential network effects between the PC segment and server segments could have led Windows to gain popularity even in server segments. Second, the adoption of Linux has increased in both server segments and the PC segment, though its share seems to be still moderate. However, note that many establishments do not own server computers, implying that 2-3% of Linux adoption in Internet servers, for example, is translated to over 10% of the share for Internet server operating systems.[20] Third, the adoption of other operating systems has declined over time. One possibility for these trends is that firms may have switched to Linux, not from Windows, but from proprietary Unix. However, this does not imply that the competition occurs only between Linux and Unix, and Windows is irrelevant to Linux adoption. Note that quite a few establishments have switched from Windows to Linux, and many establishments have switched from Unix to Windows, suggesting that the competition between Linux and Windows might be indeed intense. Section 3.3 examines these possibility in more detail.

3.2 Sample Restriction

For our empirical analysis in the following sections, we restrict our sample in order to meet three considerations. First, we restrict our sample to the establishments that report which server operating systems they are using.[21] Establishments may not report information on server operating systems for two reasons: either because they do not have any server computer, or because they consider server operating systems unimportant. By excluding the former case, we implicitly assume that our analysis is conditional on establishments' ownership of computer servers. Though it would be interesting to know which operating system an establishment without any server would choose if it started to use a server, our analysis does not allow us to construct such a counterfactual. The latter case is a common problem in many survey

[20]If some establishments own a large number of servers running only Linux, the actual market share of Linux in servers would be much higher. Hence, we may underestimate the market share for Linux.

[21]Total number of observations with enhanced information is 607,781 in the 2000-2004 CITDB. Among them, 328,109 observations report information on operating systems for either internet server or network server.

data – respondents do not answer every question in the survey, either because they do not remember, or because they do not consider it important. The CITDB is not an exception in this regard. This problem can result in a potential underestimation of the number of establishments using each operating system. If this measurement error occurs randomly, however, it does not affect the estimated market share of each operating system.[22] For lack of further information, nevertheless, we cannot account for this potential measurement error.

Second, we do not use the observations whose information on computing technology was outdated. The CITDB does not survey all the establishments every year. For some observations, the CITDB reuses information collected in the previous year.[23] If an establishment continues to use the same operating system as before, information on operating systems can be current even though it was collected in the previous year. On the other hand, if the establishment actually switched to different operating systems, using outdated information would result in a spurious positive correlation between the current choice and the previous choice. To avoid this problem, we use only those with up-to-date information. For the initial observation of each establishment in our sample, however, there is no issue regarding reusing the same information. For this reason, we include the initial observation of each establishment as long as information on computing technology was collected within the last one year.[24]

Third, we use only complete panels for our main analysis in Sections 4 and 5. Obviously, we cannot use information from establishments that are observed only once in our data. We further restrict our sample to complete panels in order to use the econometric methodology discussed

[22]It is also possible that establishments with only a few servers may tend to consider server operating systems less important and may not report their operating systems, which implies a systematic error in measurement. This possibility, on the contrary, may help our measure of market share reflect the true population, because we can examine the adoption of operating systems only at the segment level. Note that we assign a dummy for server.linux, regardless of the number of servers running Linux, which implies a potential overestimation of the Linux market share if the establishments with a small number of servers tend to install Linux, while the establishments with a large number of servers tend to install Windows. Therefore, if establishments do not report operating systems running on a small number of servers, the estimated market share for each operating system in our sample is likely to be close to the true population.

[23]Because the CITDB records when the survey on each establishment was conducted, we can find whether its information is outdated.

[24]Among the 328,109 observations with any kind of server operating system, approximately 68% of them report up-to-date information on computing technology.

in Section 4. Because this restriction reduces the size of sample considerably, however, we use a shorter panel, rather than using panels observed in all five years. Specifically, we consider four separate periods of consecutive years: 2000-2002, 2001-2003, 2002-2004, and 2000-2003. This increases the size of our sample. It also allows us to examine the sensitivity of our findings against different initial conditions.[25]

To examine potential implications of our sample restriction, Table 2 presents the market share of each operating system from the unrestricted sample and the restricted samples as discussed above. To compute the market share, a dummy for each operating system is assigned to each observation, and the table reports its mean. In panel B, for example, the mean of server.windows is 0.914 in 2000, indicating that 91.4% of establishments reported to use Windows in server computers. Because an establishment can have more than one kind of operating system, the sum of columns (1)-(3) can be larger than 1. According to the table, our restrictions do not seem to create systematically different samples from the others.[26] Though we do not report in this paper, summary statistics of various observed characteristics do not differ much across different samples. As a result, we believe that our sample restriction does not create systematic bias in our results. Nevertheless, we do not attempt to generalize our findings beyond the samples examined in our analysis.

3.3 Switching Pattern

Before we present our empirical analysis in the following sections, we examine basic patterns in our data, focusing on switching operating systems in servers. Table 3 shows the results. Panel A in the table reports the share of establishments in each year that follow the switching pattern specified in each column. The first three columns show switching patterns of Linux adoption.

[25]For different periods, we may find different estimates for switching costs or network effects. This could suggest potential sensitivity to initial conditions, but it can also imply that the extent of switching costs or network effects have changed over time.

[26]One exception is that the market shares in panels B through G are about twice larger than those in panel A, but this is expected because approximately half of observations in the unrestricted sample does not report to own any kind of server computer.

The next three columns present those of Windows adoption, and the last three columns show switching patterns for the adoption of other operating systems. Column 1 reports the share of establishments in each year that used Linux in that year but had not used Linux in the previous year. More establishments started to use Linux until 2003, but the percentage of those switching to Linux declined in 2004. This might be because more establishments continued to use Linux (see column 2). Column 3, however, reports that those who discontinued using Linux have increased over time, suggesting that Linux could also lose its current market share. Columns 4-6 show the strong dominance of Windows in server operating system markets. Over 90% of establishments in 2001 used Windows in both 2000 and 2001, and more establishments have continued to use Windows in the following years.

Similar shares for all years are reported in panel B. These are the average of the shares presented in panel A. In addition to the dominance of Windows, panel B also shows potentially significant switching costs in the adoption of operating systems. Note that the percentage of establishments who used the same operating systems in two consecutive years is higher than the percentage of those who either started or discontinued the use of any of the operating systems. Panel C reports similar shares by different industries. It shows substantial heterogeneity in Linux adoption across industries. For example, 19.4% of establishments in the information sector (the first two digits of NAICS equal to 51) continued to use Linux and 6.8% of them started to use Linux, whereas only 5.9% of those in the medical sector (the first two digits of NAICS equal to 62) continued to use Linux and only 3.5% of them started to use Linux. Considerable heterogeneity across industries is also observed in the adoption of Windows and other operating systems.

One additional observation from Table 3 is that less establishments continued to use other operating systems over time, while more observations discontinued to use other operating systems than those who started to use other operating systems (see columns 7-9 in panel A). Though it is possible that most of establishments who stopped using other operating systems could switch to Linux, the table does not provide any evidence. For this reason, we decompose these establishments into those who started to use Linux and those who switched to Windows. Table 4 presents the shares of such establishments among total observations. Columns 6 and 8 in the table report that approximately 0.39% of establishments switched to Linux, while about 1.32% of observations switched to Windows. Hence, more establishments switched from other operating systems to Windows than to Linux.[27] Table 4 also shows that those who switched to Linux are not necessarily those who stopped using other operating systems. Columns 1-4 report that those switching from Windows to Linux are not negligible, although quite a few establishments have also switched from Linux to Windows.

4 Empirical Framework

4.1 Econometric Model

Switching patterns presented in Section 3.3 suggest strong state dependence (or lock-in) and potentially high switching costs. Other factors, nevertheless, can confound the relationship between the current choice and the previous choice of server operating systems. For this reason, we consider a simple model and examine various factors that determine establishments' decision to adopt Linux (or Windows). Specifically, we employ the following reduced-form function for the net benefit of adopting server operating system j, where $j \in \{\textsf{Linux, Windows, other}\}$. For establishment i at period t, the net benefit is given by

$$\pi_{ijt} = \gamma_{jt} + \sum_{k} \beta_{jk} y_{ik(t-1)} + \alpha_j x_{it} + \delta_j Z_{it} + \eta_{ij} + \epsilon_{ijt}, \tag{1}$$

where $y_{ik(t-1)}$ is a binary variable that indicates whether establishment i adopted server operating system k at the previous period, x_{it} is a vector of predetermined variables related to non-server segments, Z_{it} is a vector of observed characteristics of the establishment, γ_{jt} captures a time effect, η_{ij} reflects unobserved preferences for the operating system j, and ϵ_{ijt} is

[27] Some establishments might experiment with new operating systems while using old operating systems, and then decided to switch to new ones in the next period. This kind of switching is reported in columns 5 and 7 which show similar patterns as those reported in columns 6 and 8.

an idiosyncratic error term.

We assume that establishment i decides to adopt operating system j at period t if the net benefit is non-negative. In other words, $y_{ijt} = \mathbb{1}\{\pi_{ijt} \geq 0\}$. Note that we do not use multinomial models commonly used in empirical literature, because an establishment in our data can own multiple servers and thus adopt more than one operating system at the segment level. In this paper, however, we do not attempt to model the joint decision of adopting multiple operating systems.[28] Instead, we consider the adoption of each operating system separately, since this approach still allows us to examine the extent of state dependence or lock-in in the adoption of operating system j. In the equation for Linux adoption, for example, the coefficient for $\mathsf{server.linux}_{t-1}$ captures the degree of lock-in. One caveat is that the coefficients for $\mathsf{server.windows}_{t-1}$ and $\mathsf{server.other}_{t-1}$ in Linux adoption do not entirely reflect costs associated with switching from Windows or other operating systems to Linux.[29]

The net benefit function in (1) also includes x_{it}, a vector of predetermined variables related to operating system choices in non-server segments. We consider x_{it} to examine potential network effects (or benefits from the compatibility) between the server segment and non-server segments. Note that our method below allows for strictly exogenous variables as well as predetermined variables, but it does not permit correlation between explanatory variables and the time varying error ϵ_{ijt}. In this regard, we use $\mathsf{pc.linux}_{t-1}$, $\mathsf{pc.windows}_{t-1}$, and $\mathsf{pc.other}_{t-1}$ (and similarly for non-pc), instead of the current choices in non-server segments.

The final elements in (1) is establishment-specific heterogeneity, where $\delta_j Z_{it}$ reflects observed heterogeneity while $\eta_{ij} + \epsilon_{ijt}$ captures unobserved heterogeneity. We assume that the idiosyncratic error term follows an i.i.d. normal distribution such that $\epsilon_{ijt} | H_i^t \sim \mathcal{N}(0, \sigma_t)$, where

[28]We could use multivariate models in the sense that we extend multinomial models by treating a segment-level joint adoption of Linux, Windows, and other operating systems as one choice. The key factor in this joint decision, however, is network effects within the same segment, which cannot be fully captured by the multivariate models. Investigating such network effects requires further modeling, which we will pursue in our future work.

[29]For example, some observation i may have $y_{ijt} = y_{ij(t-1)} = 1$, while $y_{ik(t-1)} = 1$ for $k \neq j$. In this case, establishment i did not exactly switch from k to j, and therefore, the coefficient for $y_{ik(t-1)}$ does not necessarily reflect switching costs from k to j.

$H_i^t = (H_{i0}, H_{i1}, \ldots, H_{it})$, and $H_{it} = (\mathsf{server.linux}_{t-1}, \mathsf{server.windows}_{t-1}, \mathsf{server.other}_{t-1}, x_{it}, Z_{it})$. Accordingly, we can write the probability of the adoption conditional on the history as follows:

$$\Pr(y_{ijt} = 1 | H_i^t) = \Phi\left(\frac{\gamma_{jt} + \sum_k \beta_{jk} y_{ik(t-1)} + \alpha_j x_{it} + \delta_j Z_{it} + \eta_{ij}}{\sigma_t}\right), \tag{2}$$

where Φ is the standard normal CDF. The main issue in estimating the equation (2) is the presence of η_{ij} which generates spurious state dependence between y_{ijt} and $y_{ij(t-1)}$. Ignoring the presence of η_{ij} therefore biases the estimates for the coefficients in (2).

Before we present our main method to address this issue, we briefly discuss two common approaches that might be used to account for unobserved heterogeneity in discrete choice panel data models.[30] The first, so called the random effect approach, is to impose known distributional assumptions for the unobserved heterogeneity η_{ij} and integrate η_{ij} out in the likelihood function. For a model with predetermined variables such as lagged variables $y_{ij(t-1)}$, however, this approach yields inconsistent estimates. Even if η_{ij} actually follows an i.i.d. normal distribution, it is correlated with $y_{ij(t-1)}$, since η_{ij} also determined $y_{ij(t-1)}$ at $t-1$. Consequently, the random effect η_{ij} cannot be simply integrated out. Alternatively, one could consider the following likelihood for an establishment with $T+1$ observations as

$$L_i = \int \prod_{t=1}^{T} \Pr(y_{ijt} | H_i^{t-1}, \eta_{ij}) f(y_{ij0} | \eta_{ij}) dG(\eta_{ij}), \tag{3}$$

where η_{ij} follows distribution $G(\eta_{ij})$, $f(y_{ij0}|\eta_{ij})$ denotes the marginal probability of the initial choice in the data given η_{ij}, and we assume $H_i^t = (y_{ij0}, y_{ij1}, \ldots, y_{ijt})$ for simplicity. The key difficulty of using (3) is how to specify the distribution of the initial condition given η_{ij}. One could assume that y_{ij0} is independent of η_{ij}, but this assumption is very strong because we do not observed the very beginning of the process and y_{ij0} should be determined by η_{ij} and the history before y_{ij0}. One could also assume that $f(y_{ij0}|\eta_{ij})$ represents a steady-state distribution, but such a stationary distribution has been found only in limited special cases.[31]

[30] See Hsiao (2003) for review of the literature.

[31] See Heckman (1981b) and Wooldridge (2005) for more discussion on the initial conditions problem in dynamic discrete choice panel data models.

A second approach to account for unobserved heterogeneity under discrete choice models is to treat η_i as fixed effects (i.e. not imposing distributional assumptions for η_i), while assuming a logit model for the idiosyncratic error terms. This method relies on conditional maximum likelihood methods and exploits the functional form of a conditional logit in order to difference out the fixed effects. Honoré and Kyriazidou (2000) extend this method to the case with predetermined variables. The identification of their method, however, requires conditioning the analysis on observations where the dependent variable follows specific patterns, namely $(y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4}) = (0, 0, 1, 1)$ and $(y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4}) = (0, 1, 0, 1)$. The problem of applying this method to our data is that we rarely observe the latter case. Few establishments experiment with the same operating system by not using it at the first period, using it at the second period, and not using it again at the third period, and finally using it again at the fourth period. As a result, we cannot apply this approach to our data.

These standard approaches require very strong assumptions that are unlikely to be plausible in our context. This motivates our main method below.

4.2 AC Method and Identification

Our approach follows the methodology proposed by Arellano and Carrasco (2003) – henceforth, the AC method. We begin with the model in (1) and (2). The main idea of the AC method is to modify (1) as follows:

$$\text{(A-1)} \qquad \pi_{ijt} = \gamma_{jt} + \sum_k \beta_{jk} y_{ik(t-1)} + \alpha_j x_{it} + \delta_j Z_{it} + E(\eta_{ij} | H_i^t) + \epsilon_{ijt}.$$

That is, we assume that $E(\eta_{ij}|H_i^t)$, rather than η_{ij}, determines the net benefit of adopting server operating system j. Our interpretation of the assumption (A-1) is that the quality of operating system j is not fully observed by establishments, so that establishment i revises its expectation (or forecast) of the quality of product j based on its previous experiences (i.e. H_i^t). Establishment-specific unobserved preferences for product j, however, might include some additional component not captured by $E(\eta_{ij}|H_i^t)$. The assumption (A-1) implies that such component is included in the idiosyncratic error term ϵ_{ijt}. One more implication of (A-1) is that we implicitly specify the conditional distribution of η_{ij} given the initial condition y_{ij0}, as opposed to $f(y_{ij0}|\eta_{ij})$ in (3). The corresponding likelihood under our assumption is then given by

$$L_i = \Pr(y_{ij0}) \int \prod_{t=1}^{T} \Pr(y_{ijt}|H_i^{t-1}, \eta_{ij}) dG(\eta_{ij}|y_{ij0}).$$

Therefore, the AC method allows for dependence between η_{ij} and y_{ij0}, while leaving the initial conditions of the process unrestricted.[32]

Under (A-1), the probability of the adoption conditional on the history is written as

$$p_{ijt} \equiv \Pr(y_{ijt} = 1|H_i^t) = \Phi\left(\frac{\gamma_{jt} + \sum_k \beta_{jk} y_{ik(t-1)} + \alpha_j x_{it} + \delta_j Z_{it} + E(\eta_{ij}|H_i^t)}{\sigma_t}\right). \tag{4}$$

The advantage of this equation over (2) is that the equation (4) can be inverted as

$$E(\eta_{ij}|H_i^t) = \Phi^{-1}(p_{ijt})\sigma_t - \gamma_{jt} - \sum_k \beta_{jk} y_{ik(t-1)} - \alpha_j x_{it} - \delta_j Z_{it}.$$

Given this equation, the AC method identifies the parameters in the model by using the following moment condition:

$$\text{(A-2)} \qquad E(\nu_{ijt}|H_i^t) \equiv E[E(\eta_{ij}|H_i^{t+1})|H_i^t] - E(\eta_{ij}|H_i^t) = 0,$$

where $\nu_{ijt} \equiv E(\eta_{ij}|H_i^{t+1}) - E(\eta_{ij}|H_i^t)$. The moment condition in (A-2) should hold because of the law of iterated expectation. Our interpretation of (A-2) is that if we condition only on the history up to t (i.e. H_i^t), the assessment of the quality of product j revised at period $t+1$ (i.e. $E(\eta_{ij}|H_i^{t+1})$) should be the same as the assessment of the quality of j at period t (i.e. $E(\eta_{ij}|H_i^t)$). An establishment may change its opinion about the unobserved quality of server operating system j after its experience at the present period, but conditional on its previous experience, the expected future opinion at the next period should be the same as the

[32]See Wooldridge (2005) for related approaches and further discussion on the advantages of modeling the distribution of the unobserved heterogeneity conditional on the initial conditions.

current opinion. Other than (A-2), however, we do not restrict the sequence of conditional expectations of the quality of product j, $\{E(\eta_{ij}|H_i^t), t = 1, \ldots, T\}$.

To estimate the parameters in the model, we consider the case where H_i^t is a discrete random vector with a finite support.[33] Let ϕ_h^t denote a value that H_i^t can take, and define $d_{ih}^t = \mathbb{I}\{H_i^t = \phi_h^t\}$. Under (A-1) and (A-2), the estimation is based on the moment condition:

$$E\left\{d_{ih}^{t-1}\left[\sigma_t\Phi^{-1}(p_{ijt}) - \sigma_{t-1}\Phi^{-1}(p_{ij(t-1)}) - \Delta\gamma_t - \sum_k \beta_{jk}\Delta y_{ik(t-1)} - \alpha_j\Delta x_{it}\right]\right\} = 0, \qquad (5)$$

where $\Delta\gamma_t = \gamma_t - \gamma_{t-1}$, $\Delta y_{ij(t-1)} = y_{ij(t-1)} - y_{ij(t-2)}$,[34] and $\Delta x_{it} = x_{it} - x_{i(t-1)}$. Note that the moment condition (5) contains $p_{ijt}(H_i^t) \equiv \Pr(y_{ijt} = 1|H_i^t)$. Arellano and Carrasco (2003) propose a two-step method – in the first step, $p_{ijt}(H_i^t)$ is estimated nonparametrically, and in the second step, a GMM method based on (5) is used with $p_{ijt}(H_i^t)$ replaced by $\hat{p}_{ijt}(H_i^t)$. In the case of a discrete vector H_i^t, $\hat{p}_{ijt}$ can simply be the frequency of occurences of y_{ijt} in all establishments with the same history. Arellano and Carrasco (2003) also present a minimum distance estimation as well as a maximum likelihood estimation. For our application, however, we employ their two-step GMM method because of its natural connection with (A-2). We then estimate standard errors using the formula derived in Arellano and Carrasco (2003).

Before we present the results below, we need to discuss a practical issue in applying the AC method. To obtain consistent estimates, the AC method requires nonparametric estimates for $p_{ijt}(H_i^t)$. In practice, however, it is not feasible to nonparametrically estimate this probability as a function of the history of many regressors in the net benefit function. Our approach to this issue is to estimate the model using relatively homogeneous samples, so that we can reduce the dimensionality in Z_{it}, a vector of variables capturing observed heterogeneity. Specifically, we consider 10 subsamples based on the NAICS code and estimate the model separately for each subsample. Alternatively, one could also consider a semi-parametric approach to estimate

[33] Note that $y_{ij(t-1)}$ and x_{it} are discrete variables, and we drop Z_{it} in our estimation using the AC method. Therefore, H_i^t takes a discrete value in our application.

[34] The presence of $\Delta y_{ij(t-1)}$ in the moment condition implies that identification of the parameters would require at least three observations available on each establishment.

$p_{ijt}(H_i^t)$. For example, one might use a single-index model such that $p_{ijt}(H_i^t) = p_{ijt}(H_i^{t\prime}\theta)$, where θ is a vector of coefficients corresponding to each element in H_i^t. Note, however, that the right hand-side of (4) is a function of H_i^t. As a result, using the single-index $H_i^{t\prime}\theta$ implicitly imposes a structure which is unlikely to be internally consistent with the original model in (4). For this reason, we do not consider semi-parametric approach to estimate $p_{ijt}(H_i^t)$.

5 Results

5.1 Probit Results

Tables 5 and 6 report the results obtained by using standard probit (Columns 1 and 2) as well as the AC method (Column 3). These tables present the estimation results for the 2000-2003 complete panel as discussed in Section 3.2. Note that we do not report the results for the 2000-2002, 2001-2003, and 2002-2004 complete panels, merely because of a space concern. The results from these samples in different periods are similar to those in the tables, suggesting that our findings based on Tables 5 and 6 are unlikely to be artifacts from restricting our samples.

Table 5 reports parameter estimates for Windows adoption, wheras Table 6 presents parameter estimates for Linux adoption. We group relatively homogenous samples based on kinds of business, and estimate the model separately for different subsamples.[35] The main parameters of interest are the coefficients for $\mathsf{server.windows}_{t-1}$ in Windows adoption and $\mathsf{server.linux}_{t-1}$ in Linux adoption, potentially capturing state dependence or lock-in. We also include predetermined variables for operating systems adoption in non-server segments, in a way to reflect potential network effects between servers and non-servers. Note, however, that we exclude some variables such as $\mathsf{pc.windows}_{t-1}$ (or $\mathsf{non\text{-}pc.linux}_{t-1}$) for some samples, since they are almost all one (or zero), resulting in multicollinearity.

According to the results from standard probit without accounting for unobserved prefer-

[35]We group ten subsamples based on the NAICS codes, but exclude those whose first digit NAICS code is equal to 7 or 8, because total number of such observations in our complete panels is not enough to enable us to apply the AC method.

ences (Columns 1 and 2), $\mathsf{server.windows}_{t-1}$ in Windows adoption and $\mathsf{server.linux}_{t-1}$ in Linux adoption appear to be the most important factors to determine the choices of the current server operating systems. The coefficients for these variables are all positive and statistically significant. For all the samples, the previous choice of server operating systems seem to dominate any other factors. Interestingly, the magnitudes of these coefficients are also very similar across different samples. These results do not change much even after we account for observed heterogeneity by including various establishment-specific characteristics (Column 2). We also find that most of coefficient estimates for these characteristics are small and statistically indistinguishable from zero,[36] suggesting that observed heterogeneity is less likely to be important, or at least that our subsamples are reasonably homogenous.

5.2 AC Method Results

The preceding section shows the seemingly robust positive correlation between the current choice and the previous choice of server operating systems, potentially implying high switching costs and lock-in. Because the positive correlation is consistent with the importance of unobserved preferences as well, however, we attempt to distinguish them by imposing the assumptions (A-1) and (A-2) and use the AC method. The results are reported in Column 3 in Tables 5 and 6.

In Windows adoption, the magnitudes of the coefficient estimates for $\mathsf{server.windows}_{t-1}$ vary across different subsamples. Nevertheless, the estimates are considerably smaller than those reported in Columns 1 and 2. For many subsamples, the coefficient estimates are statistically insignificant. All the coefficient estimates for $\mathsf{server.windows}_{t-1}$ therefore point to one conclusion: the previous use of Windows does not seem to be the most important factor that explains the current use of Windows across various industries. These results suggest that the adoption of Windows at the previous period does not necessarily lock establishments into Windows.

The estimates for Linux adoption show similar results. The magnitudes of the coefficient

[36] To save space, we do not report coefficient estimates for these characteristics.

estimates for $\mathsf{server.linux}_{t-1}$ differ across various industries. The estimates, nonetheless, are substantially smaller than those in Columns 1 and 2. They are statistically insignificant for most samples, although standard errors seem to be quite high for some samples, possibly reflecting small sample sizes. The estimates for $\mathsf{server.linux}_{t-1}$, nevertheless, indicate that the adoption of Linux at the previous period may not be the important factor to determine the current choice of Linux.

For other variables in tables, most coefficients for non-server operating systems are small and statistically insignificant in both Windows adoption and Linux adoption. This suggests that network effects or benefits from compatibility between server operating systems and non-server operating systems might not be important. Some of year dummies, however, are statistically significant, though their magnitudes are not substantial.

These results contrast with those discussed in the previous section. They imply that the strong positive correlations between the current choice and the previous choice from standard probits seem likely to reflect unobserved preferences for either operating systems, rather than high switching costs and lock-in. In the adoption decision, more important factors might be establishment-specific unobserved preferences based on the assessment of unknown quality of server operating systems.

6 Concluding Remarks

In this paper, we focus on two leading factors – lock-in and superior quality – in the decision to adopt Windows or Linux as operating systems for server computers. Using detailed establishment-level data, we attempt to distinguish state dependence or lock-in from unobserved preferences for either operating system. Without accounting for unobserved preferences, we find a seemingly robust positive correlation between the current choice and the previous choice. To account for unobserved preferences, we impose weak identifying assumptions and employ recently developed dynamic discrete choice panel data methods. The results show little or no evidence of state dependence, implying that unobserved preferences, rather than switching costs and lock-in, are more important factors in the adoption decision.

These findings may be interpreted as evidence that perceived better quality accounts for the persistent dominance of Windows as well as the growing popularity of Linux in server operating systems. Nevertheless, we do not rule out the possibility that true state dependence and lock-in might not be measured properly by simply estimating the coefficient for the choice of Windows or Linux at the previous period. Moreover, unobserved preferences and quality may rather reflect potential network effects particularly within the same segment, as opposed to between different segments. This motivates future work on further modeling.

References

Arellano, M. and Carrasco, R. "Binary Choice Panel Data Model with Predetermined Variables." *Journal of Econometrics*, 2003, *115*, pp. 125-57.

Bresnahan, T. "Network Effects and Microsoft." Manuscript. Stanford University, 2001.

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

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

Fink, M. *The Business and Economics of Linux and Open Source*, New Jersey: Prentice Hall PTR, 2002.

Greenstein, S. "Did Installed Base Give an Incumbent Any (Measurable) Advantages in Federal Computer Procurement?." *RAND Journal of Economics*, 1993, *24*, pp. 19-39.

Heckman, J. "Statistical Models for Discrete Panel Data." *Structural Analysis of Discrete Panel Data with Econometric Applications*, edited by C. Manski and D. McFadden. Cambridge, MA: MIT Press. 1981a, pp. 114-78.

Heckman, J. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process." *Structural Analysis of Discrete Panel Data with Econometric Applications*, edited by C. Manski and D. McFadden. Cambridge, MA: MIT Press. 1981b, pp. 179-95.

Heckman, J. "Heterogeneity and State Dependence." *Studies in Labor Markets*, edited by S. Rosen. University of Chicago Press. 1981c, pp. 91-139.

Honoré, B. and Kyriazidou, E. "Panel Data Discrete Choice Models with Lagged Dependent Variables." *Econometrica*, 2000, *70*, pp. 839-74.

Hsiao, C. *Analysis of Panel Data.* Cambridge, United Kingdom: Cambridge University Press. 2003.

Liebowitz, S. and Margolis, S. *Winners, Losers, & Microsoft: Competition and Antitrust in HIgh Technology.* The Independent Institute. 1999.

Varian, H. and Shapiro, C. "Linux Adoption in the Public Sector: An Economic Analysis." Manuscript. University of California, Berkeley, 2003.

Wooldridge, J. "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics*, 2005, *20*, pp. 39-54.

Table 1: Summary Statistics of Variables for Each Year[a]

Year	2000	2001	2002	2003	2004
	(1)	(2)	(3)	(4)	(5)
server.windows[b]	0.46	0.44	0.51	0.52	0.51
server.linux	0.04	0.05	0.06	0.07	0.07
server.other	0.13	0.13	0.13	0.12	0.11
internet.sv.windows	0.13	0.15	0.17	0.15	0.11
internet.sv.linux	0.02	0.02	0.03	0.03	0.02
internet.sv.other	0.03	0.03	0.03	0.03	0.02
network.sv.windows	0.40	0.37	0.45	0.46	0.46
network.sv.linux	0.02	0.03	0.04	0.05	0.05
network.sv.other	0.11	0.11	0.11	0.11	0.10
pc.windows	0.84	0.85	0.88	0.90	0.90
pc.linux	0.04	0.07	0.08	0.10	0.11
pc.other	0.09	0.06	0.06	0.05	0.03
non-pc.windows	0.02	0.02	0.04	0.03	0.02
non-pc.linux	0.00	0.00	0.00	0.00	0.00
non-pc.other	0.28	0.21	0.20	0.16	0.10
perl[c]	0.01	0.01	0.01	0.01	0.01
apache[d]	0.02	0.03	0.03	0.03	0.03
#pc[e]	159.17	158.54	168.22	176.32	181.85
#non-pc	2.71	2.27	2.29	1.97	1.34
#internet.server	0.53	0.64	0.85	0.85	0.83
#network.server	4.79	4.74	4.96	5.14	5.17
#pc.server	4.25	4.41	4.96	5.17	5.25
#employees[f]	316.25	299.01	298.42	297.53	288.86
#white.collar.workers	175.51	167.19	172.24	174.04	170.53
#desk.workers	137.75	129.85	134.85	133.73	131.78
#programmers	3.10	3.31	3.23	3.10	3.14
#it.workers	n/a	4.61	6.18	6.81	8.44
#internet.users	61.08	68.63	72.14	77.27	82.11
#internet.developers	0.60	0.65	0.70	0.73	0.78
revenue[g]	68.90	64.28	62.05	60.50	60.65
#observations	120,880	124,324	120,984	121,324	120,269

[a]The table reports the mean of each variable.
[b]Dummy equal to 1 if Windows is installed on either internet server or network server in the establishment.
[c]Dummy equal to 1 if Perl is installed on any computer in the establishment.
[d]Dummy equal to 1 if Apache is installed on any computer in the establishment.
[e]Total number of PCs that are not used for any server.
[f]Total number of employees in the establishment.
[g]The amount of revenue for each establishment estimated by Harte-Hanks (in $million).

Table 2: Yearly Market Share of Each Operating System for Different Samples[a]

Year	server. windows	server. linux	server. other	pc. windows	pc. linux	pc. other	non-pc. windows	non-pc. linux	non-pc. other	#obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	A. Unrestricted sample									
2000	0.456	0.036	0.133	0.839	0.044	0.087	0.021	0.003	0.279	120,880
2001	0.444	0.050	0.133	0.854	0.066	0.062	0.025	0.004	0.211	124,324
2002	0.510	0.064	0.133	0.878	0.085	0.057	0.036	0.005	0.196	120,984
2003	0.517	0.074	0.124	0.896	0.099	0.050	0.034	0.005	0.162	121,324
2004	0.509	0.069	0.111	0.900	0.106	0.033	0.023	0.003	0.099	120,269
	B. Sample with any server operating system									
2000	0.914	0.072	0.267	0.920	0.071	0.116	0.031	0.005	0.380	60,282
2001	0.881	0.099	0.265	0.897	0.103	0.080	0.036	0.006	0.288	62,641
2002	0.897	0.112	0.234	0.909	0.121	0.073	0.051	0.007	0.259	68,786
2003	0.908	0.130	0.218	0.918	0.136	0.063	0.047	0.007	0.216	69,087
2004	0.910	0.124	0.199	0.898	0.137	0.040	0.031	0.005	0.134	67,313
	C. Sample with any server o/s and with up-to-date information									
2000	0.914	0.072	0.267	0.920	0.071	0.116	0.031	0.005	0.380	60,282
2001	0.878	0.103	0.258	0.978	0.116	0.081	0.040	0.006	0.274	37,065
2002	0.903	0.117	0.229	0.984	0.135	0.076	0.060	0.007	0.272	46,001
2003	0.911	0.140	0.219	0.982	0.155	0.062	0.047	0.008	0.205	40,119
2004	0.913	0.128	0.200	0.978	0.151	0.035	0.026	0.005	0.111	39,642
	D. Complete panel: 2000-2003[b]									
2000	0.923	0.081	0.283	0.953	0.072	0.122	0.032	0.003	0.413	11,010
2001	0.926	0.111	0.262	0.983	0.125	0.089	0.043	0.004	0.302	11,010
2002	0.936	0.140	0.246	0.989	0.151	0.076	0.064	0.007	0.293	11,010
2003	0.938	0.159	0.234	0.984	0.175	0.059	0.049	0.008	0.203	11,010
	E. Complete panel: 2000-2002									
2000	0.920	0.078	0.280	0.949	0.071	0.117	0.031	0.003	0.406	18,061
2001	0.925	0.106	0.260	0.982	0.122	0.086	0.043	0.004	0.299	18,061
2002	0.933	0.133	0.245	0.988	0.148	0.073	0.064	0.007	0.293	18,061
	F. Complete panel: 2001-2003									
2001	0.900	0.108	0.262	0.980	0.116	0.093	0.040	0.005	0.301	18,435
2002	0.923	0.135	0.246	0.987	0.148	0.079	0.064	0.008	0.283	18,435
2003	0.929	0.155	0.234	0.984	0.172	0.062	0.050	0.008	0.204	18,435
	G. Complete panel: 2002-2004									
2002	0.914	0.124	0.242	0.984	0.130	0.087	0.060	0.007	0.287	19,674
2003	0.931	0.148	0.228	0.984	0.156	0.068	0.051	0.008	0.219	19,674
2004	0.926	0.137	0.205	0.980	0.155	0.036	0.028	0.005	0.116	19,674

[a] To compute the market share, a dummy for each operating system is assigned to each observation, and the table reports its mean. Because an establishment can have more than one kind of operating system, the sum of columns (1)-(3) can be larger than 1.

[b] Complete panels include only those with any server operating system and also with up-to-date information.

Table 3: Switching Patterns of the Adoption of Operating Systems in Servers[a]

Industry	Year	[$Linux_{t-1}$, $Linux_t$]			[$Windows_{t-1}$, $Windows_t$]			[$Other_{t-1}$, $Other_t$]			total obs.
		[0,1]	[1,1]	[1,0]	[0,1]	[1,1]	[1,0]	[0,1]	[1,1]	[1,0]	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
		A. All Industries and By Year									
	2001	0.041	0.064	0.014	0.023	0.899	0.019	0.030	0.226	0.049	27,578
	2002	0.046	0.080	0.019	0.041	0.873	0.020	0.038	0.207	0.057	37,244
	2003	0.046	0.095	0.022	0.034	0.887	0.019	0.034	0.190	0.050	35,968
	2004	0.030	0.101	0.036	0.022	0.896	0.024	0.024	0.180	0.048	36,055
		B. All Industries and All Years									
		0.041	0.086	0.023	0.030	0.888	0.020	0.032	0.199	0.051	136,845
		C. By Industry and All Years									
agri_utility.1-2[b]		0.029	0.051	0.019	0.026	0.914	0.017	0.027	0.148	0.045	6,026
manufacture.31-32		0.025	0.043	0.015	0.026	0.906	0.018	0.027	0.167	0.043	13,720
manufacture.33		0.034	0.065	0.019	0.024	0.908	0.019	0.028	0.188	0.047	19,622
retail.4		0.033	0.064	0.020	0.032	0.868	0.020	0.030	0.200	0.046	10,529
information.51		0.068	0.194	0.034	0.045	0.794	0.033	0.044	0.294	0.060	9,429
financial.52-53		0.030	0.049	0.015	0.028	0.907	0.014	0.025	0.147	0.046	9,578
professional.54-56		0.045	0.100	0.025	0.027	0.900	0.018	0.029	0.159	0.049	13,308
education.61		0.063	0.164	0.039	0.044	0.851	0.029	0.046	0.292	0.068	17,743
medical.62		0.035	0.059	0.020	0.027	0.914	0.015	0.027	0.172	0.048	12,043
arts_service.7-8		0.028	0.055	0.020	0.028	0.894	0.022	0.026	0.145	0.046	3,550
government		0.041	0.072	0.022	0.028	0.900	0.019	0.031	0.207	0.053	17,976

[a]The table reports the share of the establishments that follow the switching pattern specified in each column. The sample includes only those that reported information on server operating systems. The sample also excludes the observations whose information on computing technology was outdated.

[b]The number after the name of industry denotes the first digit (or first two digits) of NAICS.

Table 4: Switching between Different Operating Systems in Servers[a]

Industry	Year	$\text{Windows}_{t-1}=1$, $\text{Windows}_t=0$		$\text{Linux}_{t-1}=1$, $\text{Linux}_t=0$		$\text{Other}_{t-1}=1$, $\text{Other}_t=0$			
		$L_{t-1}=1$, $L_t=1$	$L_{t-1}=0$, $L_t=1$	$W_{t-1}=1$, $W_t=1$	$W_{t-1}=0$, $W_t=1$	$L_{t-1}=1$, $L_t=1$	$L_{t-1}=0$, $L_t=1$	$W_{t-1}=1$, $W_t=1$	$W_{t-1}=0$, $W_t=1$
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		A. All Industries and By Year							
	2001	0.0030	0.0037	0.0111	0.0019	0.0041	0.0048	0.0387	0.0073
	2002	0.0036	0.0032	0.0123	0.0050	0.0045	0.0049	0.0375	0.0167
	2003	0.0040	0.0023	0.0158	0.0041	0.0054	0.0035	0.0323	0.0159
	2004	0.0055	0.0024	0.0288	0.0037	0.0067	0.0024	0.0348	0.0115
		B. All Industries and All Years							
		0.0041	0.0028	0.0173	0.0038	0.0052	0.0039	0.0357	0.0132
		C. By Industry and All Years							
agri_utility.1-2		0.0028	0.0018	0.0136	0.0045	0.0043	0.0030	0.0299	0.0131
manufacture.31-32		0.0018	0.0025	0.0097	0.0037	0.0026	0.0018	0.0289	0.0125
manufacture.33		0.0029	0.0029	0.0146	0.0030	0.0031	0.0028	0.0348	0.0110
retail.4		0.0031	0.0027	0.0141	0.0039	0.0032	0.0037	0.0302	0.0144
information.51		0.0097	0.0045	0.0225	0.0056	0.0117	0.0070	0.0396	0.0138
financial.52-53		0.0018	0.0022	0.0105	0.0030	0.0024	0.0030	0.0306	0.0144
professional.54-56		0.0047	0.0033	0.0203	0.0031	0.0052	0.0041	0.0349	0.0120
education.61		0.0079	0.0035	0.0311	0.0046	0.0125	0.0068	0.0492	0.0151
medical.62		0.0021	0.0023	0.0140	0.0041	0.0027	0.0037	0.0331	0.0140
arts_service.6-7		0.0045	0.0037	0.0138	0.0042	0.0045	0.0031	0.0276	0.0163
government		0.0038	0.0022	0.0171	0.0032	0.0041	0.0034	0.0387	0.0129

[a]The table reports the share of the establishments that follow the switching pattern specified in each column. The sample includes only those that reported information on server operating systems. The sample also excludes the observations whose information on computing technology was outdated. Note that L_t (or W_t) denotes the use of Linux (or Windows) at period t.

Table 5: Results for Windows Adoption from 2000-2003 Complete Panel

Variable	Probit				AC Method	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
	(1)		(2)		(3)	
	A. NAICS 1-2: Agriculture, Utility, and Construction					
$server.linux_{t-1}$	-0.2400	0.2696	-0.5097	0.3078	-0.6100	0.2339
$server.windows_{t-1}$	**2.1137**	**0.2095**	**1.9766**	**0.2308**	**-0.0833**	**0.6379**
$server.other_{t-1}$	-0.7869	0.1743	-0.9168	0.2041	0.1239	0.5652
$pc.linux_{t-1}$	-0.1149	0.2725	-0.3150	0.3110	0.8880	0.2224
$non\text{-}pc.other_{t-1}$	0.0558	0.1697	-0.0624	0.1879	0.1180	0.5157
γ_{2001}	0.2487	0.2593	0.1869	0.3234	0.4868	0.2713
γ_{2002}	0.7109	0.2642	0.7263	0.3265	0.8897	1.1733
γ_{2003}	0.1411	0.2625	0.0414	0.3244	-0.6852	0.1546
	B. NAICS 31-32: Manufacturing					
$server.linux_{t-1}$	-0.5008	0.1726	-0.5263	0.1980	-0.2709	1.0126
$server.windows_{t-1}$	**2.1847**	**0.1239**	**2.0919**	**0.1332**	**0.6396**	**2.3218**
$server.other_{t-1}$	-0.4742	0.1110	-0.5019	0.1207	0.0531	1.2414
$pc.linux_{t-1}$	0.2031	0.1972	0.2373	0.2173	-0.7982	2.3477
$pc.windows_{t-1}$	0.5063	0.2594	0.6117	0.2822	-0.3763	0.6367
$pc.other_{t-1}$	0.1175	0.1691	0.0831	0.1775	0.2351	0.5087
$non\text{-}pc.windows_{t-1}$	0.0394	0.2412	-0.2629	0.2577	0.3860	1.7883
$non\text{-}pc.other_{t-1}$	-0.0244	0.0996	-0.0958	0.1078	-0.3122	0.4762
γ_{2001}	-0.4670	0.2908	-0.5999	0.3330	0.9521	0.3261
γ_{2002}	-0.5218	0.2960	-0.6655	0.3366	0.0308	0.2160
γ_{2003}	-0.2289	0.2970	-0.3337	0.3352	-0.0905	0.2391
	C. NAICS 33: Manufacturing					
$server.linux_{t-1}$	-0.5034	0.1150	-0.5635	0.1263	-0.3455	0.8980
$server.windows_{t-1}$	**2.1623**	**0.0990**	**2.1013**	**0.1027**	**-0.6462**	**0.8558**
$server.other_{t-1}$	-0.5233	0.0870	-0.5675	0.0907	-0.5201	0.5028
$pc.linux_{t-1}$	0.1642	0.1330	0.2031	0.1423	0.0123	0.3987
$pc.windows_{t-1}$	0.1292	0.2641	0.2091	0.2628	-0.9718	0.6142
$pc.other_{t-1}$	-0.1372	0.1537	-0.1786	0.1602	1.2666	0.5523
$non\text{-}pc.windows_{t-1}$	-0.0104	0.1456	-0.0792	0.1509	-0.0673	2.6958
$non\text{-}pc.other_{t-1}$	-0.1034	0.0816	-0.1268	0.0856	0.0498	0.4198
γ_{2001}	0.0368	0.2787	-0.0588	0.2942	0.7928	0.2696
γ_{2002}	0.1156	0.2885	0.0279	0.3030	-0.0635	0.1416
γ_{2003}	0.0542	0.2924	-0.0330	0.3068	0.6299	0.6177
	D. NAICS 4: Wholesale, Retail, and Transportation					
$server.linux_{t-1}$	-0.2247	0.1923	-0.2996	0.2055	1.1918	3.2555
$server.windows_{t-1}$	**2.3338**	**0.1458**	**2.2925**	**0.1515**	**0.0752**	**3.9759**
$server.other_{t-1}$	-0.6339	0.1298	-0.6647	0.1340	-0.4619	0.9639
$pc.linux_{t-1}$	0.1018	0.2070	0.0152	0.2140	-1.1091	2.6280
$pc.other_{t-1}$	-0.1677	0.2448	-0.2530	0.2550	-0.9709	1.6089
$non\text{-}pc.other_{t-1}$	-0.0778	0.1222	-0.1020	0.1268	0.6219	0.8612
γ_{2001}	0.0778	0.1901	-0.0351	0.2569	0.7353	0.4763
γ_{2002}	0.0184	0.1920	-0.1014	0.2569	0.6099	0.9014
γ_{2003}	0.0044	0.1965	-0.1113	0.2613	0.0049	0.0090
additional control[a]	No		Yes		No	

[a]Additional control includes revenue, #it.workers, #programmers, #desk.workers, apache, #pc, #non-pc, #internet.server, #network.server, #pc.server, and dummies for population where establishments are located.

Table 5: Results for Windows Adoption (Continued)

Variable	Probit				AC Method	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
	(1)		(2)		(3)	
	E. NAICS 51: Information					
server.linux$_{t-1}$	-0.4975	0.1195	-0.4025	0.1325	0.8256	6.9077
server.windows$_{t-1}$	**2.2486**	**0.1018**	**2.1982**	**0.1044**	**0.4355**	**2.4374**
server.other$_{t-1}$	-0.4048	0.1056	-0.4216	0.1089	-1.2925	1.4588
pc.linux$_{t-1}$	-0.0230	0.1228	-0.0352	0.1296	-0.8058	4.6321
pc.other$_{t-1}$	-0.2789	0.1088	-0.3080	0.1136	-0.9122	1.0541
non-pc.other$_{t-1}$	0.1057	0.1048	-0.0039	0.1096	-0.0004	0.7098
γ_{2001}	-0.2052	0.1442	-0.3748	0.1804	1.9090	2.5531
γ_{2002}	-0.1170	0.1439	-0.3213	0.1805	-0.0334	0.1403
γ_{2003}	-0.0914	0.1468	-0.3025	0.1818	0.0092	0.0350
	F. NAICS 52-53: Finance and Real Estate					
server.linux$_{t-1}$	-0.4957	0.2618	-0.6325	0.2872	-1.0016	7.2711
server.windows$_{t-1}$	**2.1608**	**0.1913**	**2.1177**	**0.2054**	**1.3606**	**5.1082**
server.other$_{t-1}$	-0.5334	0.1696	-0.5960	0.1918	-0.3885	1.2477
pc.linux$_{t-1}$	-0.1447	0.2494	-0.1239	0.2925	-0.1071	0.2630
pc.windows$_{t-1}$	0.3811	0.3964	0.6651	0.4296	-0.6780	1.0850
pc.other$_{t-1}$	-0.3540	0.3363	-0.2686	0.3802	0.3199	1.0869
non-pc.other$_{t-1}$	0.2010	0.1533	0.1317	0.1659	-0.0732	0.5104
γ_{2001}	-0.1662	0.4276	-0.4036	0.5041	0.7067	0.3482
γ_{2002}	-0.1930	0.4391	-0.4370	0.5136	-0.1901	0.1221
γ_{2003}	-0.2039	0.4425	-0.3744	0.5169	0.4203	0.3260
	G. NAICS 54-56: Professional and Technical Services					
server.linux$_{t-1}$	-0.5688	0.1326	-0.7029	0.1508	0.1558	1.4408
server.windows$_{t-1}$	**2.1411**	**0.1283**	**2.0749**	**0.1351**	**-0.1237**	**1.6132**
server.other$_{t-1}$	-0.6060	0.1212	-0.6916	0.1321	-1.0279	1.3229
pc.linux$_{t-1}$	-0.2390	0.1413	-0.2367	0.1495	0.0630	1.1458
pc.windows$_{t-1}$	0.4084	0.2896	0.4214	0.3009	0.1236	1.9699
pc.other$_{t-1}$	-0.4057	0.1619	-0.4051	0.1682	0.1885	0.1971
non-pc.windows$_{t-1}$	0.2611	0.2895	0.1900	0.3252	0.5298	0.8448
non-pc.other$_{t-1}$	0.1031	0.1232	0.0872	0.1333	-0.5819	0.9349
γ_{2001}	-0.1600	0.3107	-0.0448	0.3877	1.5581	1.2706
γ_{2002}	-0.0178	0.3232	0.1351	0.3955	-0.3242	0.1738
γ_{2003}	-0.0648	0.3252	0.0915	0.3988	-0.0893	0.2111
additional control[a]	No		Yes		No	

[a]Additional control includes revenue, #it.workers, #programmers, #desk.workers, apache, #pc, #non-pc, #internet.server, #network.server, #pc.server, and dummies for population where establishments are located.

Table 5: Results for Windows Adoption (Continued)

Variable	Probit				AC Method	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
	(1)		(2)		(3)	
	H. NAICS 61: Educational Services					
server.linux$_{t-1}$	-0.3368	0.0792	-0.4037	0.0854	0.5116	1.9153
server.windows$_{t-1}$	**2.1196**	**0.0774**	**2.0707**	**0.0799**	**0.8832**	**1.9319**
server.other$_{t-1}$	-0.3487	0.0759	-0.4059	0.0794	-0.6752	1.0749
pc.linux$_{t-1}$	0.1069	0.0908	0.0267	0.0939	-0.2430	2.2783
pc.windows$_{t-1}$	0.4299	0.1169	0.4319	0.1186	0.1788	1.9397
pc.other$_{t-1}$	0.0388	0.0787	0.0393	0.0810	-0.9587	0.8733
non-pc.other$_{t-1}$	0.1734	0.0801	0.0888	0.0850	-0.1366	0.4145
γ_{2001}	-0.5306	0.1436	-0.5162	0.1536	1.6607	1.4075
γ_{2002}	-0.3427	0.1496	-0.3128	0.1591	0.0156	0.1285
γ_{2003}	-0.4502	0.1537	-0.4310	0.1632	-0.6401	0.2095
	I. NAICS 62: Health Care					
server.linux$_{t-1}$	-0.2452	0.1756	-0.2213	0.1878	0.2400	3.0700
server.windows$_{t-1}$	**2.3643**	**0.1547**	**2.3603**	**0.1638**	**0.5515**	**0.4469**
server.other$_{t-1}$	-0.4665	0.1308	-0.5420	0.1384	-0.5111	0.3282
pc.linux$_{t-1}$	-0.3100	0.1789	-0.3499	0.1966	0.0874	1.4727
pc.windows$_{t-1}$	0.1951	0.3657	0.1727	0.4199	0.0324	2.3790
pc.other$_{t-1}$	-0.1320	0.2459	-0.2082	0.2637	-0.2396	0.3413
non-pc.windows$_{t-1}$	0.3188	0.4237	0.3380	0.4553	-0.4225	0.4917
non-pc.other$_{t-1}$	-0.0990	0.1280	-0.1947	0.1427	0.2187	0.5515
γ_{2001}	-0.1606	0.3777	-0.1906	0.4406	0.9710	0.4012
γ_{2002}	0.0076	0.4017	-0.0564	0.4637	-0.0206	0.2097
γ_{2003}	-0.0106	0.4084	-0.0822	0.4697	0.2945	0.5217
	J. NAICS 9: Public Administration					
server.linux$_{t-1}$	-0.3623	0.1085	-0.3667	0.1148	0.0751	2.1366
server.windows$_{t-1}$	**2.2756**	**0.0896**	**2.2323**	**0.0925**	**0.9074**	**1.3132**
server.other$_{t-1}$	-0.6647	0.0844	-0.6886	0.0867	-0.6583	0.5504
pc.linux$_{t-1}$	-0.0570	0.1142	-0.0799	0.1194	-0.1388	2.2270
pc.windows$_{t-1}$	0.2253	0.2259	0.3430	0.2296	-0.5168	0.5483
pc.other$_{t-1}$	0.0835	0.1565	0.0314	0.1631	-1.1044	1.0443
non-pc.windows$_{t-1}$	0.3114	0.1822	0.1924	0.1910	-0.2976	1.5553
non-pc.other$_{t-1}$	-0.0148	0.0786	-0.0489	0.0818	0.1332	0.3063
γ_{2001}	-0.1203	0.2398	-0.2557	0.2568	0.7842	0.1745
γ_{2002}	0.0335	0.2455	-0.1033	0.2621	0.6225	0.5270
γ_{2003}	-0.1724	0.2471	-0.3187	0.2646	-0.2299	0.1544
additional control[a]	No		Yes		No	

[a] Additional control includes revenue, #it.workers, #programmers, #desk.workers, apache, #pc, #non-pc, #internet.server, #network.server, #pc.server, and dummies for population where establishments are located.

Table 6: Results for Linux Adoption from 2000-2003 Complete Panel

Variable	Probit				AC Method	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
	(1)		(2)		(3)	
	A. NAICS 1-2: Agriculture, Utility, and Construction					
server.linux$_{t-1}$	**2.4462**	**0.1699**	**2.3363**	**0.1790**	**1.0791**	**5.8312**
server.windows$_{t-1}$	0.7208	0.3325	0.5816	0.3289	-0.2788	4.1046
server.other$_{t-1}$	0.4711	0.1399	0.4369	0.1567	-0.0259	1.0404
pc.linux$_{t-1}$	0.8031	0.1724	0.8092	0.1792	0.1641	2.8536
non-pc.other$_{t-1}$	0.0978	0.1303	0.0441	0.1405	0.1506	0.4113
γ_{2001}	-2.9677	0.3589	-2.8164	0.3684	-0.6111	0.3112
γ_{2002}	-2.6777	0.3475	-2.5431	0.3551	-0.0225	0.2305
γ_{2003}	-2.8198	0.3642	-2.6649	0.3718	-0.6757	0.9315
	B. NAICS 31-32: Manufacturing					
server.linux$_{t-1}$	**2.3700**	**0.1306**	**2.2591**	**0.1380**	**0.2332**	**5.5547**
server.windows$_{t-1}$	-0.1815	0.1518	-0.2321	0.1638	0.5191	0.6253
server.other$_{t-1}$	0.3081	0.1015	0.2122	0.1127	0.0270	0.6849
pc.linux$_{t-1}$	0.8513	0.1285	0.8108	0.1373	-0.6578	0.5925
pc.windows$_{t-1}$	-0.1106	0.3103	0.1258	0.3607	0.9593	0.5632
pc.other$_{t-1}$	0.1360	0.1540	0.1225	0.1681	-0.6139	0.6819
non-pc.windows$_{t-1}$	0.4346	0.1679	0.4231	0.1836	0.1070	1.7725
non-pc.other$_{t-1}$	0.0942	0.0933	0.0888	0.1010	-0.0817	0.1626
γ_{2001}	-1.8947	0.3321	-2.2070	0.3872	-0.3864	0.1543
γ_{2002}	-1.8543	0.3369	-2.1662	0.3936	-0.5878	0.6378
γ_{2003}	-1.9292	0.3366	-2.1920	0.3902	0.5239	0.1865
	C. NAICS 33: Manufacturing					
server.linux$_{t-1}$	**2.3072**	**0.0819**	**2.1901**	**0.0863**	**0.7922**	**2.6546**
server.windows$_{t-1}$	0.1082	0.1246	0.0334	0.1330	0.0310	1.0613
server.other$_{t-1}$	0.2470	0.0696	0.2048	0.0747	-0.0158	0.1525
pc.linux$_{t-1}$	0.7528	0.0819	0.7035	0.0860	-0.2638	0.7866
pc.windows$_{t-1}$	-0.2233	0.2453	-0.1543	0.2654	0.0533	0.4515
pc.other$_{t-1}$	0.2053	0.1275	0.1467	0.1363	-0.2386	0.8431
non-pc.windows$_{t-1}$	0.1324	0.1094	0.0934	0.1157	-0.2596	1.8652
non-pc.other$_{t-1}$	0.0882	0.0639	0.0406	0.0692	0.1408	0.3620
γ_{2001}	-1.7877	0.2647	-1.9231	0.2946	-0.4419	0.2205
γ_{2002}	-1.7485	0.2714	-1.9012	0.3008	-0.0987	0.1382
γ_{2003}	-1.8741	0.2738	-1.9485	0.3030	-0.0142	0.1668
	D. NAICS 4: Wholesale, Retail, and Transportation					
server.linux$_{t-1}$	**2.1845**	**0.1263**	**2.0777**	**0.1318**	**0.7717**	**3.7202**
server.windows$_{t-1}$	0.0534	0.1645	-0.0212	0.1721	0.2324	1.0611
server.other$_{t-1}$	0.5155	0.1048	0.4638	0.1112	0.1861	0.2117
pc.linux$_{t-1}$	0.7675	0.1276	0.7075	0.1340	-1.0716	3.7800
pc.other$_{t-1}$	-0.2636	0.2621	-0.3489	0.2856	-0.0432	1.5115
non-pc.other$_{t-1}$	0.1361	0.0975	0.1641	0.1033	0.2186	0.6815
γ_{2001}	-2.1205	0.1922	-2.2946	0.2524	-0.7731	0.5927
γ_{2002}	-2.1630	0.1939	-2.3271	0.2533	-0.1651	0.3980
γ_{2003}	-1.9746	0.1928	-2.1259	0.2532	-0.2192	0.8206
additional control[a]	No		Yes		No	

[a]Additional control includes revenue, #it.workers, #programmers, #desk.workers, apache, #pc, #non-pc, #internet.server, #network.server, #pc.server, and dummies for population where establishments are located.

Table 6: Results for Linux Adoption (Continued)

Variable	Probit (1) Estimate	Probit (1) S.E.	Probit (2) Estimate	Probit (2) S.E.	AC Method (3) Estimate	AC Method (3) S.E.
	E. NAICS 51: Information					
server.linux$_{t-1}$	**2.4257**	**0.1029**	**2.2381**	**0.1088**	**1.5468**	**6.7492**
server.windows$_{t-1}$	-0.0945	0.1148	-0.0293	0.1214	-0.0349	3.5709
server.other$_{t-1}$	0.1891	0.0903	0.1456	0.0945	0.8704	1.0939
pc.linux$_{t-1}$	0.5573	0.0954	0.4402	0.0997	-0.4815	2.1890
pc.other$_{t-1}$	0.0187	0.1053	0.0318	0.1111	0.2139	1.0916
non-pc.other$_{t-1}$	0.0997	0.0884	0.1238	0.0937	-0.4183	1.5637
γ_{2001}	-1.6606	0.1456	-1.8918	0.1821	-1.0970	0.6969
γ_{2002}	-1.6480	0.1456	-1.8566	0.1825	0.3772	0.3674
γ_{2003}	-1.4767	0.1438	-1.6348	0.1771	-0.3131	0.6167
	F. NAICS 52-53: Finance and Real Estate					
server.linux$_{t-1}$	**2.5984**	**0.1780**	**2.5967**	**0.1889**	**-1.4322**	**8.8333**
server.windows$_{t-1}$	0.0211	0.2386	-0.0077	0.2644	-0.5526	0.8692
server.other$_{t-1}$	0.4080	0.1381	0.1754	0.1616	-0.1705	1.0137
pc.linux$_{t-1}$	0.9052	0.1604	0.7169	0.1745	1.2972	5.0129
pc.windows$_{t-1}$	-0.4127	0.3759	-0.3540	0.4123	-0.6040	0.8511
pc.other$_{t-1}$	0.0830	0.3013	-0.0958	0.3377	0.6070	0.3700
non-pc.other$_{t-1}$	0.0472	0.1173	-0.0050	0.1305	-0.0002	0.2845
γ_{2001}	-1.7912	0.4053	-2.1278	0.4836	-0.1002	0.1864
γ_{2002}	-1.6484	0.4209	-1.9882	0.4973	0.1101	0.1117
γ_{2003}	-1.6554	0.4211	-2.0104	0.4988	-0.8400	1.1267
	G. NAICS 54-56: Professional and Technical Services					
server.linux$_{t-1}$	**2.3657**	**0.0968**	**2.2186**	**0.1013**	**-0.1799**	**2.6967**
server.windows$_{t-1}$	-0.1857	0.1449	-0.2730	0.1547	-0.4210	1.4712
server.other$_{t-1}$	0.2133	0.0950	0.0323	0.1051	0.4673	1.4864
pc.linux$_{t-1}$	0.5903	0.0987	0.5547	0.1036	0.1630	3.4858
pc.windows$_{t-1}$	-0.0299	0.2827	0.1335	0.3058	0.1530	1.6833
pc.other$_{t-1}$	0.0334	0.1412	-0.0287	0.1513	-0.0281	0.2968
non-pc.windows$_{t-1}$	-0.1224	0.1655	-0.1588	0.1774	0.2234	3.9757
non-pc.other$_{t-1}$	0.1808	0.0876	0.0671	0.0952	0.4267	0.4551
γ_{2001}	-1.5799	0.3060	-1.3959	0.3524	-0.5611	0.1328
γ_{2002}	-1.5707	0.3129	-1.4414	0.3594	0.0187	0.1265
γ_{2003}	-1.5869	0.3141	-1.4074	0.3604	0.1639	0.1167
additional control[a]	No		Yes		No	

[a] Additional control includes revenue, #it.workers, #programmers, #desk.workers, apache, #pc, #non-pc, #internet.server, #network.server, #pc.server, and dummies for population where establishments are located.

Table 6: Results for Linux Adoption (Continued)

Variable	Probit (1) Estimate	(1) S.E.	Probit (2) Estimate	(2) S.E.	AC Method (3) Estimate	(3) S.E.
	H. NAICS 61: Educational Services					
server.linux$_{t-1}$	**2.2183**	**0.0580**	**2.1225**	**0.0600**	**-0.0576**	**1.3704**
server.windows$_{t-1}$	-0.0925	0.0828	-0.1003	0.0876	0.0860	1.5217
server.other$_{t-1}$	0.2144	0.0547	0.1091	0.0581	-0.1269	0.4114
pc.linux$_{t-1}$	0.4491	0.0600	0.3696	0.0626	0.7678	1.9329
pc.windows$_{t-1}$	0.2487	0.1216	0.2301	0.1283	0.0938	1.9257
pc.other$_{t-1}$	0.0662	0.0584	0.0667	0.0606	-0.2937	0.2270
non-pc.other$_{t-1}$	0.2694	0.0545	0.1865	0.0579	0.4367	0.4700
γ_{2001}	-1.6688	0.1396	-1.6606	0.1517	-0.5648	0.1549
γ_{2002}	-1.6570	0.1422	-1.6591	0.1542	-0.3021	0.2611
γ_{2003}	-1.7626	0.1454	-1.7251	0.1568	0.0562	0.2510
	I. NAICS 62: Health Care					
server.linux$_{t-1}$	**2.3016**	**0.1156**	**2.2475**	**0.1192**	**0.3865**	**0.6098**
server.windows$_{t-1}$	-0.3311	0.1633	-0.3419	0.1691	-0.6120	0.7247
server.other$_{t-1}$	0.1648	0.0968	0.1418	0.1003	0.3524	0.7846
pc.linux$_{t-1}$	0.8028	0.1184	0.7189	0.1222	-0.0374	1.2735
pc.windows$_{t-1}$	-0.5905	0.2899	-0.6020	0.2933	-0.1556	2.3653
pc.other$_{t-1}$	0.3116	0.1637	0.3502	0.1693	0.0411	0.5765
non-pc.windows$_{t-1}$	0.1769	0.2336	0.2486	0.2327	0.0169	0.1853
non-pc.other$_{t-1}$	0.2265	0.0876	0.2058	0.0923	0.3165	0.1384
γ_{2001}	-1.1277	0.3191	-1.1803	0.3278	-0.4321	0.1610
γ_{2002}	-0.9138	0.3340	-0.9628	0.3436	0.1681	0.1190
γ_{2003}	-1.0088	0.3367	-1.0261	0.3457	-0.3117	0.2162
	J. NAICS 9: Public Administration					
server.linux$_{t-1}$	**2.3387**	**0.0746**	**2.2644**	**0.0778**	**1.2505**	**2.2006**
server.windows$_{t-1}$	0.0656	0.1099	0.0153	0.1150	-0.8214	0.9837
server.other$_{t-1}$	0.2020	0.0641	0.1452	0.0672	-0.2358	0.2263
pc.linux$_{t-1}$	0.6681	0.0744	0.6039	0.0778	0.1334	0.9280
pc.windows$_{t-1}$	0.1550	0.2462	0.1080	0.2536	0.2577	0.5039
pc.other$_{t-1}$	0.0583	0.1221	-0.1734	0.1368	-0.4116	1.3493
non-pc.windows$_{t-1}$	0.1287	0.1093	0.0218	0.1166	0.1122	0.9879
non-pc.other$_{t-1}$	0.1579	0.0587	0.1152	0.0618	-0.1000	0.1701
γ_{2001}	-2.0987	0.2614	-2.0302	0.2749	-0.2141	0.0793
γ_{2002}	-2.0717	0.2656	-2.0218	0.2799	-0.4161	0.1542
γ_{2003}	-2.1240	0.2669	-2.0421	0.2811	0.0434	0.1320
additional control[a]	No		Yes		No	

[a]Additional control includes revenue, #it.workers, #programmers, #desk.workers, apache, #pc, #non-pc, #internet.server, #network.server, #pc.server, and dummies for population where establishments are located.