NET Institute*

www.NETinst.org

Working Paper #06-12

September 2006

Network Effects, Switching Costs, and Underlying Preferences in Operating Systems for Servers: A Case of Linux vs. Windows

Seung-Hyun Hong
University of Illinois
and
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.

Network Effects, Switching Costs, and Underlying Preferences in Operating Systems for Servers: 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

September 30, 2006

Abstract

We seek to investigate to what extent network effects and switching costs affect 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. To account for unobserved preferences for either operating system, we employ recently developed dynamic discrete choice panel data methods (Arellano and Carrasco 2003). The results from our empirical analysis suggest that among network effects, switching costs, and unobserved preferences, the last two are important factors in the market for operating systems for servers. We find that switching costs are significant, but can be severely overestimated by methods that do not account for unobserved heterogeneity in establishment-specific tastes for operating systems. We also find that once taste heterogeneity is taken into account, network effects are not significant.

*We thank the NET Institute for financial support. 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

This paper examines firms' choices of operating systems for computer servers. In particular, we focus on Linux and Windows, and empirically measure the degree of network effects and switching costs of each operating system, accounting for both observed and unobserved heterogeneity across firms. We believe that measuring these effects is a crucial step to empirically assess the driving forces behind the persistence of Microsoft's Windows as the dominant player in the market for operating systems.

The continuing dominance of Windows can be explained by three potential factors: *network effects*, in that consumers prefer to use an operating system that is popular among other users; *switching costs*, in that consumers are locked into their previous investment in Windows; or simply *unobserved consumer preferences* – most consumers may simply consider Windows a better product than the existing alternatives.

Though network effects and switching costs of Windows have played a prominent role in the debate over the antitrust case against Microsoft (e.g., Bresnahan 2001), few empirical studies have measured either network effects or switching costs associated with Windows or operating systems in general. Moreover, little empirical work has investigated economic determinants underlying the adoption of Linux. Though Linux is often considered to be a viable competitor to challenge the Windows monopoly, the effectiveness of this competition depends on the extent of network effects and switching costs in this market, which is a matter of empirical investigation. We thus attempt to fill these gaps by quantifying the extent to which network effects, switching costs, and both observed and unobserved heterogeneity explain the adoption (or non-adoption) of Windows and Linux among business firms.

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 the choice of operating systems among business establishments.

In this paper, we focus on the adoption decision for operating systems in computer servers, since it is the server segment in which Linux is perceived to be the greatest threat to Windows. Nevertheless, our analysis also considers the adoption of operating systems in the personal computer (PC) segment. Specifically, we investigate whether the choice in the PC segment affects the adoption in the server segment.

The empirical analysis consists of several steps. We first examine descriptive statistics of our data. Particular attention is given to the frequency of switching patterns in operating system adoption, as this is informative of whether switching costs are significant in this market. We then analyze the results from linear regressions. Though these have the defect of imposing a linear probability framework on the data and not accounting for unobserved heterogeneity across firms, they are helpful as a first indication of the main factors affecting the adoption of operating systems.

We next consider a probit model to relax the linear probability assumption. The non-linear nature of discrete choice models, however, does not allow for the common differencing strategy to address fixed effects in linear regressions. We thus examine a random effect probit model, another common approach to integrate out unobserved heterogeneity. This approach, nonetheless, contains serious bias as well in our context, since the previous adoption choice also affects the current adoption decision, while firm-specific unobserved preferences determine the adoption decisions not only in the current period but also in the previous period. For this reason, unobserved heterogeneity captured by the random effect cannot be integrated out.

Consequently, we employ a dynamic discrete choice panel data method, recently developed by Arellano and Carrasco (2003), which addresses the preceding empirical challenges and allows us to consistently estimate a model that includes network effects, switching costs, and unobserved preferences. We find that both switching costs and unobserved preferences are important. In fact, the methods that neglect the role of unobserved preferences tend to severely overestimate the impact of switching costs. We also find that network effects are not significant; at least the type of network effect considered here, which is popularity in the PC segment leading to the adoption in the server segment.

The paper is organized as follows: the next section provides a background on the history of the open source movement and the Linux operating system. Section 3 describes our data. Section 4 provides descriptive statisics and presents the results of the linear regressions. Section 5 presents the main empirical method used in this paper. Section 6 discusses the results of our main method and contrasts them with those obtained using probit methods. Section 7 concludes with a summary of the findings and directions for future research.

2 Background

2.1 Linux

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.[1] Fur-

[1]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.

thermore, 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.[2] 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[3] 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.[4]

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. 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

[2]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.

[3]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.

[4]The whole 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 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.)

web browser. Consequently, it is not entirely impossible that Linux might eventually supercede Windows as the leading operating system. To assess the extent to which Linux may weaken the Windows dominance, this paper studies underlying forces behind the adoption of Linux and attempts to measure their magnitudes.

2.2 Linux Adoption

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,[5] and most estimates point to the increasing rate of Linux adoption, particularly in server computers.[6]

In this paper, we consider business firms' decision to adopt Linux in servers. In essence, firms will choose Linux if the net benefits from Linux are higher than those from other operating systems, particularly Windows. There are several factors that may determine benefits

[5]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.

[6]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).

and costs associated with using Linux over other operating systems. First, potentially 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 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, there might be direct network effects between different computers within a segment of use within a firm, as well as across segments. Especially in the adoption of operating systems for servers, firms may experience network effects between the operating systems in the 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. Third, indirect network effects may also influence the adoption of Linux. In particular, Linux was the key component of the "LAMP" server software combination.[7] For instance, a firm may adopt Linux for Web servers in order to run Apache. This suggests that the adoption of some application software (e.g. Apache, MySQL, or Perl) may be another factor behind the adoption of Linux.

Despite the potential significance of switching costs and network effects, however, other

[7] 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.

factors can be important as well. Put simply, firms may have different preferences over alternative operating systems. For example, firms 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. Moreover, it is not unlikely that most users in many firms may consider Windows the best among all existing operating systems. Because we may not observe all firm characteristics that might influence their preferences over operating systems, we adopt a methodology that allows for unobserved firm-specific effects in our empirical exercise.

3 Data

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 data on the use of a variety of information and communication technologies by each establishment. 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.[8] 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

[8] 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.

remains similar each year,[9] but many establishments were not included in every year. In other words, the CITDB is an unbalanced panel data set.

We study the adoption of operating systems at the segment level. The CITDB groups computers into four segments: personal computers, not used for servers; Internet servers; network servers; and non-PCs such as mainframes and workstations that are not used for either network server or Internet server. For notational convenience, we refer to the first segment as the PC segment[10] and the last three segments combined as the server segment. Note that we can only investigate operating system adoption 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[11] 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,

[9]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). These newly added observations, however, do not report their uses of detailed technologies such as operating systems in web servers and thus cannot be used in our analysis. In fact, the number of observations relevant for our analysis – those who report detailed technology use – remains about 120,000 in each year.

[10]Though PCs are normally used for standalone desktops or client computers connected to servers, some PCs are powerful enough to be used as Internet servers or network servers. The PC segment in this paper, however, denotes only the former group and does not include PCs used as servers.

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

etc.) but also Berkeley Software Distribution (BSD).[12] 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 at least two 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[13]; and similarly for *server.windows* and *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 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.[14] 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

[12]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.

[13]We also use *internet.sv.linux* for Linux on the Internet server segment, *lansv.linux* for Linux on the network server segment, and *non-pc.linux* for Linux on the non-PC such as mainframes not used for either Internet server or network server. The main analysis in this version of the paper, however, does not consider this more refined division of the server segment.

[14]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. Another reason for potential underestimation of Linux market share is that the establishments in the CITDB may not report the use of technologies which they consider unimportant. As a result, some establishments may use Linux in a few servers, but may not report the use of Linux. We do not account for this measurement error due to lack of further information.

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. The next section examines these possibility in more detail.

4 Descriptive Statistics

This section presents descriptive statistics of the main variables and provide a preliminary analysis of the extent of switching costs and network effects. To this end, we first examine the switching patterns in operating system adoption. We then consider linear regressions of the adoption of Linux (or Windows) on various explanatory variables. Regression results suggest potential network effects and possibly substantial switching costs, but we cannot rule out the possibility that unobserved heterogeneity accounts for these correlations, particularly between the current choice of operating systems and the previous choice. This motivates our main empirical method, presented in Sections 5 and 6.

4.1 Switching Pattern

Table 2 shows basic patterns in our data as to switching operating systems in servers. Note that in this table and henceforth, we restrict our samples by excluding observations without any server. Therefore, we implicitly assume that our analysis is conditional on establishments' ownership of computer servers. Though it would be interesting to know what operating systems 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. Note also that we exclude establishments that are observed only once in our data, obviously because we cannot observer their operating systems in the previous period.

Panel A in Table 2 reports the share of establishments in each year that follow a certain switching pattern specified in each column. The first three columns show switching patterns of Linux adoption. 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 60% 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, 18.3% of establishments in the information sector (the first two digits of NAICS equal to 51) continued to use Linux and 5.3% of them started to use Linux, whereas only 5% of those in the medical sector (the first two digits of NAICS equal to 62) continued to use Linux and only 2.8% 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 2 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 3 presents the shares of such establishments among total observations. Columns 6 and 8 in the table report that approximately 0.3% of establishments switched to Linux, while about 1% of observations switched to Windows. Hence, more establishments switched from other operating systems to Windows than to Linux.[15] Table 3 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.

4.2 Linear Regression Results

Switching patterns presented above suggest potential switching costs. To investigate the presence of other factors beyond switching costs, we consider linear regressions of the adoption of Linux (or Windows) on various variables. Table 4 presents the results from these regressions. Columns 1-2 show coefficient estimates and p-values for the regression where the dependent variable is a dummy equal to 1 if an establishment has Linux in any server – Internet server,

[15] 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.

non-PC, or network server. Columns 3-4 presents the results from a similar regression, except that the dependent variable is a dummy for Windows adoption in server segments. The results suggest that factors discussed in Section 2.2 are important: switching costs, network effects, observed heterogeneity, and unobserved heterogeneity.

First, the coefficient for lagged Windows adoption on servers is negative and significant in the regression of Linux adoption, and conversely the coefficient of lagged Linux adoption in the regression of Windows adoption is negative and statistically significant. This suggests that switching costs may be important. Other coefficients seem to confirm this finding as well. Coefficient estimates for Linux (resp. Windows) adoption on servers in the previous period is statistically significant positive in the regression of Linux (resp. Windows) adoption. These results, however, are also consistent with considerable heterogeneity due to unobserved preferences for either Linux or Windows. Even if switching costs are inconsequential, a firm may continue to use the same operating systems as before, since its preference for a certain operating system in the previous period might remain the same and would continue to influence its choice of operating systems over time.

Second, for the regression of Linux (resp. Windows) adoption, we have significantly positive coefficients for Linux (resp. Windows) adoption in the PC segment. This is consistent with direct network effects between PCs and servers. The coefficient of the lagged dummy for Linux (resp. Windows) adoption in PCs, however, is small but negative and statistically significant in the regression of Linux (resp. Windows) adoption. This finding may not be entirely consistent with positive network effects between PCs and servers. It may be however due to mispecification bias due to the linear probability model assumed here.

Third, coefficient estimates for various establishment characteristics are statistically significant, implying that observed heterogeneity may be an important factor in the operating system adoption decision. Fourth, we have positive and significant coefficient estimates for the uses of Apache and Perl, suggesting that indirect network effects between Linux and these open source software applications may exist.

The results in Table 4, however, do not seem to support the possibility that firms may have switched to Linux mostly from proprietary Unix. Observe that the coefficient for the lagged variable of other operating system adoption in servers is negative and statistically significant. Therefore, the previous adoption of other operating system such as Unix is negatively correlated with the adoption of Linux in the current period. These results, nevertheless, may instead suggest that switching costs are high not only for Windows or Linux, but also for other operating systems.

Though these correlations suggest the potential importance of network effects and switching costs, they are confounded by other factors, mainly unobserved preference heterogeneity at the establishment level. To address this problem, the next section presents the main empirical model used in this paper.

5 Empirical Model

In this section, we model establishments' decision to choose among different operating systems and account for unobserved heterogeneity in the adoption decision. We follow the methodology proposed by Arellano and Carrasco (2003). For notational convenience, we focus on the adoption of Linux in this section. The extension of our methodology to the adoption of Windows is straightforward.

The dependent variable is L_{it}, a binary variable that indicates if establishment i adopted

Linux as the operating system for one or more of its servers in year t. This decision is assumed to be affected by a vector x_{it} of discrete predetermined variables, as follows:

$$L_{it} = \mathbb{1}\{\beta x_{it} + \gamma_t + \eta_i + v_{it} \geq 0\}$$

where β is a vector of coefficients to be estimated, γ_t is a period-specific effect, η_t is an establishment-specific effect, and v_{it} is an idiosyncratic shock. Since the number of establishments is large, the vector of η_i's cannot be directly estimated.

The variables in x_{it} are predetermined (as opposed to strictly exogenous) because they may depend on past realizations of the shock v_{it-k}, for $k \geq 1$; in particular, this allows L_{it-1} to be one of the regressors. Also, this makes the method convenient for our application since one of the regressors can be lagged dependent variables of other similar processes, such as Linux adoption in desktops. In other words, the method allows for switching costs as well as network effects across uses within an establishment. (We still assume independence over different establishments).

Let H_{it} be the history observed in the data for establishment i up to period t (that is, the realizations of $x_{i1}, \ldots, x_{it}$ and $L_{i1}, \ldots, L_{it-1}$). We assume that

$$\eta_i + v_{it} | H_{it} \sim \mathcal{N}(E[\eta_i | H_{it}], 1).$$

This is a slight departure from Arellano and Carrasco (2003), as they allow this random variable to be heteroskedastic over time. Once homoskedasticity is imposed, setting the variance equal to 1 is a mere normalization choice.

From this assumption, we can write the probability of adoption conditional on the history as follows:

$$p_{it} \equiv \Pr(L_{it} = 1 | H_{it}) = \Phi\left(\beta x_{it} + \gamma_t + E(\eta_i | H_{it})\right),$$

where Φ is the CDF of the standard normal distribution.

The method is based on the insight that this equation can be inverted to obtain an expression that can be first-differenced:

$$\Phi^{-1}(p_{it}) - \Phi^{-1}(p_{it-1}) - \beta(x_{it} - x_{it-1}) - (\gamma_{it} - \gamma_{it-1}) = E(\eta_i|H_{it}) - E(\eta_i|H_{it-1}) \equiv \epsilon_{it}.$$

By the the definition of a conditional expectation, $E(\epsilon_{it}|H_{it-1}) = 0$. If p_{it} was available, these moments conditions could be used to estimate β and γ by GMM.

Arellano and Carrasco (2003) propose a two-step method, where in the second step GMM is performed with p_{it} replaced by estimates $\hat{p}$ obtained in the first step. In the case of discrete x_{it}, $\hat{p}_{it}$ can simply be the frequency of occurences of L in all establishments with the same history. We followed this suggestion, using as moment conditions the sample analog of $E(\epsilon_{it}|H_{it-1}) = 0$, for every history observed 3 times or more.

6 Results

6.1 Sample Selection

We estimate the empirical model discussed in the previous section using selected samples of the dataset. This is done for computational tractability, as well as allowing for parameters to be estimated in a flexible way. The Arellano and Carrasco method involves nonparametrically estimating p_{it}, the probability of adoption conditional on the history of the dependent variable and all regressors. Nonparametric estimation of this probability as a function of the history of, say, of all variables listed in Table 4 is unfeasible. To make the estimation tractable, we could impose more structure and use a semi-parametric approach such as single-index (or multiple-index) model for p_{it}.[16] Alternatively, we can select less heterogeneous samples and estimate the

[16]See Horowitz (1998) for potential semi-parametric approaches. In our future work, we plan to incorporate this semi-parametric approach into our estimation.

model for each sample. By doing so, we can maintain our estimation tractable, while allowing the estimates to reflect patterns in the data in a flexible way.

We use six sets of samples. We focus on two industries: the information sector (the first two digits of NAICS code equal to 51) and the medical sector (the first two digits of NAICS equal to 62). We then estimate the model for the samples in 2000-2002, for those in 2001-2003, and for those in 2002-2004 separately.

We select these two industries for the following reasons. First, based on the regression results in Table 4, the industry dummy is one of the most important observed factors affecting the operating system adoption. Focusing on individual industries allows us to reduce potential heterogeneity within each sample, while maintaining a reasonable sample size. Second, both selected industries rely heavily on computer use, compared to several other sectors. On the other hand, these two industries are contrasting in that the Linux adoption rate is relatively high in the information sector, while it is rather low in the medical sector. Third, the coefficients of the main variables in linear regressions and probit regressions are only slightly different between these selected samples and the full sample.

Note also that we consider three separate periods of three consecutive years in order to increase the number of complete panels. Recall that the CITDB is an unbalanced panel data. As a result, if we use observations that appear in all five years, the total number of observations becomes small. Moreover, considering three years reduces our estimation time, while it still allows for identification. Because the extent of either switching costs or network effects can change over time, however, we consider three periods: 2000-2002, 2001-2003, and 2002-2004.

6.2 Estimation Results

We estimate the model for both Linux adoption and Windows adoption, and include four predetermined variables in our model: server.linux$_{t-1}$, server.windows$_{t-1}$, pc.linux$_{t-1}$, and pc.windows$_{t-1}$.[17] The first two variables capture switching costs, whereas the last two variables reflect potential network effects between the PC segment and the server segment. Therefore, the coefficient estimates for these four variables indicate the extent of two major factors. The other major factor to determine the adoption of operating systems is unobserved heterogeneity, particularly unobserved preference for operating systems. We account for this unobserved factor using the method discussed in Section 5.

Table 7 reports the results from the main estimation for the six sets of samples. For the purpose of comparison, Tables 5 and 6 present coefficient estimates from probit regressions. Table 5 reports results from a standard probit procedure, whereas in Table 6 establishment-specific random effects are included.

The results from probit regressions of Linux adoption report significantly positive estimates for server.linux$_{t-1}$, which remains similar across different samples. Likewise, the probit results for Windows adoption show significantly positive estimates for server.windows$_{t-1}$. These estimates therefore indicate substantial switching costs in the adoption of operating systems. The results from simple probit regressions also suggest moderate network effects between the PC segment and the server segment for both Linux and Windows in both industries, with the exception of Windows adoption in the medical sector. Nevertheless, these estimates from a

[17]In this version of the paper, we decided to include only these four variables in the estimation for three reasons. First, these four variables capture the main effects – switching costs and network effects. Second, using only four predetermined variables reduces our estimation time for the main model in Section 5. Third, the coefficient estimates for these four variables in linear regressions and probit regressions remain robust across different specifications. In our future work, however, we plan to include more variables.

simple probit model are likely to be biased if unobserved heterogeneity is in fact an important factor to determine the adoption of operating systems.

Before we discuss the results from the main estimation, we consider two common approaches that might be used to account for unobserved heterogeneity under discrete choice models.[18] The first, the random effect approach, is to impose known distributional assumptions for the unobserved heterogeneity η_i and integrate η_i out in the likelihood function. Results for this method are presented in Table 6.

For a model with predetermined variables such as lagged variables y_{t-1}, however, this approach yields inconsistent estimates. Even if η_i actually follows i.i.d. normal distribution, it is correlated with the lagged variable y_{t-1}, since η_i also determined y_{t-1} in the previous period. For this reason, the random effect η_i cannot be integrated out. In addition to inconsistency, the results from Table 6 suggest that the method is not successful in capturing the underlying unobserved preferences, since the estimated coefficients are very similar to those reported in Table 5.

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 $(y_{i1}, y_{i2}, y_{i3}, y_{i4}) = (0, 0, 1, 1)$ and $(y_{i1}, y_{i2}, y_{i3}, y_{i4}) = (0, 1, 0, 1)$. The problem of applying this

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

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. For this reason, we cannot apply this approach to our data.

We now examine the estimation results using the method developed by Arellano and Carrasco (2003). To be specific, consider the estimates of Linux adoption for the information sector during the period between 2000 and 2002. The coefficient for server.linux$_{t-1}$ is 1.03 and is precisely estimated. In contrast, the same coefficient from the simple probit is 2.64. These two estimates suggest that the simple probit model overestimates switching costs because of the positive correlation between η_i and server.linux$_{t-1}$, and that unobserved preference for Linux is likely to be a significant factor to determine the adoption of Linux. The coefficient for server.window$_{t-1}$ is negative, but its magnitude is insignificant, which suggests that potential switching costs associated with pervious Windows usage might not be important in the adoption of Linux. Moreover, the estimated coefficient for pc.window$_{t-1}$ is negative but statistically indistinguishable from zero, implying that network effects between the PC segment and the server segment are unlikely to have discouraged establishments to adopt Linux in 2001 and 2002, despite the strong dominance of Windows in the PC segment.

Windows adoption for the information sector in 2000-2002 also shows similar results as Linux adoption. The coefficient estimate for server.window$_{t-1}$ is 1.41, which is estimated precisely. The corresponding estimate in the simple probit, however, is 2.16. This suggests that both switching costs and unobserved heterogeneity explain why establishments continued to use Windows on the server segment. The coefficient for pc.windows$_{t-1}$ is positive but is not estimated precisely, suggesting that network effects between the PC segment and the server segment might not be important in the adoption of operating systems in the server segment.

The estimates from 2001-2003, on the other hand, suggest somewhat different implications. For both server.linux$_{t-1}$ in Linux adoption and server.window$_{t-1}$ in Windows adoption, the coefficient estimates using the method in Section 5 are slightly different from those using the simple probit, which suggests that switching costs now play more significant role in the adoption of operating systems, whereas unobserved preference becomes less important. Note that more establishments in later years, say 2003, continued to use the same operating systems than those in earlier years, say in 2001. Therefore, the estimated results are consistent with the interpretation that those who used either Linux or Windows in earlier years might be using the same operating systems because of their high preference for either operating system, while those in later years might not be switching because of higher switching costs. This interpretation applies to the adoption of Linux for the sample in 2002-2004 as well, since the coefficients for server.linux$_{t-1}$ are slightly different between the simple probit and the method based on Arellano and Carrasco (2003). Windows adoption during 2002-2004, however, is not consistent with this interpretation. The estimates from 2002-2004 suggest that switching costs associated with Windows are still significant, but more establishments are likely to continue their use of Windows because of their preference.

The estimates for the medical sector show similar results. However, there are two differences to note. First, the coefficients for server.linux$_{t-1}$ in Linux adoption steadily increase over time, which is consistent with the interpretation that only those with high preference for Linux used Linux in earlier years, while those who continued to use Linux in later years did so because of switching costs associated with Linux. Therefore, it is less likely that those who started to use Linux would switch to Windows, despite the strong dominance of Windows and the moderate

rate of Linux adoption at present.

Second, the coefficients for server.window$_{t-1}$ in Windows adoption are slightly lower in the estimation results from the Arellano and Carrasco method, than those from the simple probit. Note that these simple probit estimates from the medical sector are relatively lower than those from the information sector. As a result, if one would use only the results from the simple probit, one could incorrectly infer that switching costs associated with Windows were lower in the medical sector than in the information sector. In contrast, the results from the Arellano and Carrasco method suggest that switching costs of Windows are higher in the medical sector than in the information sector, and that most establishments in the medical sector continued to use Windows because of switching costs, rather than unobserved preference for Windows.

To sum up the results, we find that switching costs account for not only the persistent dominance of Windows, but also the continuing use of Linux among those who started to use Linux. Unobserved preference for either operating system, nevertheless, has a substantial impact on the adoption pattern of either operating system as well, except for the adoption of Windows in the medical sector. However, network effects between the PC segment and the server segment do not seem to be a significant factor in the adoption of operating systems for servers. We also find that switching costs associated with Windows are unlikely to be an important factor in the adoption of Linux.[19] These findings therefore suggest that the dominance of Windows in the server segment is due to both switching costs and unobserved

[19]This finding appears to be inconsistent with high switching costs of Windows. Note, however, that the server segment in this paper is composed of separate types of servers – Internet servers, network server, and mainframes or workstations that are not used for either Internet server or network server. For example, an establishment may continue to use Windows on its network server because of high switching costs, but it can adopt Linux on its Internet server and continue to use Linux on its Internet server because of high switching costs associated with Linux. Therefore, our finding is not necessarily inconsistent with high switching costs of Windows. However, this possibility suggests that we may consider more refined server segment such as Internet server. We plan to do so in our future work.

preference for Windows, but that this dominance is not likely to induce the current users of Linux to switch to Windows in the future.

7 Concluding Remarks

In this paper, we investigate the extent to which network effects, switching costs, and unobserved preferences affect the choice of operating systems for computer servers in business establishments. We find that switching costs are significant but can be severely overestimated by methods that do not account for unobserved heterogeneity in establishment-specific tastes for operating systems. We also find that once taste heterogeneity is taken into account, network effects between the PC segment and the server segment are not significant. These results suggest that both switching costs and unobserved preferences are important factors to determine the adoption of either Linux or Windows.

In this version of the paper, we restrict the sample and the explanatory variables in order to deal with observed heterogeneity in a tractable way. In the future we plan to extend the model to make use of further information from the data. In particular, we plan to incorporate a single-index approach into the main method discussed in Section 5. Horowitz and Härdle (1996) develop a direct semi-parametric estimation method for a single-index model, which does not require a nonparametric maximum likelihood estimation. This approach is thus tractable and will enable us to flexibly estimate the first stage probability of the adoption of operating systems conditional on many predetermined variables. Using this approach and the full sample, we plan to examine robustness of the findings in this paper.

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

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

Horowitz, J. *Semiparametric Methods in Econometrics*, New York, NY: Springer-Verlag. 1998.

Horowitz, J. and Härdle, W. "Direct Semiparametric Estimation of a Single-Index Model with Discrete Covariates." *Journal of the American Statistical Association*, 1996, *91*, pp. 1632-1640.

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

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

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	0.45	0.52	0.53	0.52
server.linux	0.04	0.05	0.07	0.08	0.07
server.other	0.35	0.29	0.28	0.25	0.19
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
internet.sv.windows[c]	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
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
lansv.windows[d]	0.40	0.37	0.45	0.46	0.46
lansv.linux	0.02	0.03	0.04	0.05	0.05
lansv.other	0.11	0.11	0.11	0.11	0.10
perl[e]	0.01	0.01	0.01	0.01	0.01
apache[f]	0.02	0.03	0.03	0.03	0.03
#pc[g]	159.17	158.54	168.22	176.32	181.85
#non-pc[h]	2.71	2.27	2.29	1.97	1.34
#internet.server[i]	0.53	0.64	0.85	0.85	0.83
#lansv[j]	4.79	4.74	4.96	5.14	5.17
#lan[k]	0.74	0.69	0.71	0.68	0.68
#pc.server[l]	4.25	4.41	4.96	5.17	5.25
#employees[m]	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
#company.employees[n]	24,287.71	25,625.21	25,928.75	26,897.60	21,757.04
#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

[a]The table reports the mean of each variable.
[b]The dummy for Windows installed on any server – Internet server, non-PC, or network server.
[c]The dummy for Windows installed on Internet servers.
[d]The dummy for Windows installed on network servers.
[e]The dummy for perl installed on any computer.
[f]The dummy for apache installed on any computer.
[g]Total number of PCs that are not used for any server.
[h]Total number of non-PCs that are not used for either network server or Internet server.
[i]Number of internet servers.
[j]Number of lan servers.
[k]Number of LANs (local area networks) in establishment.
[l]Number of PCs used as servers.
[m]Number of employees in establishment.
[n]Number of employees in establishment's parent company.

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

Year	2000	2001	2002	2003	2004
	(1)	(2)	(3)	(4)	(5)
employee.100[b]	0.18	0.19	0.19	0.19	0.20
employee.250[c]	0.32	0.32	0.31	0.29	0.29
employee.500	0.16	0.15	0.14	0.14	0.14
employee.750	0.04	0.04	0.03	0.03	0.03
employee.1,000	0.03	0.02	0.02	0.03	0.03
employee.1,500	0.02	0.02	0.02	0.02	0.02
employee.2,000	0.01	0.01	0.01	0.01	0.01
employee.3,000	0.01	0.01	0.01	0.01	0.01
employee.5,000	0.01	0.01	0.01	0.01	0.01
employee.10,000	0.00	0.00	0.00	0.00	0.00
employee.100,000	0.00	0.00	0.00	0.00	0.00
revenue[d]	68.90	64.28	62.05	60.50	60.65
agriculture.1[e]	0.01	0.01	0.01	0.01	0.01
utility.2	0.05	0.05	0.05	0.05	0.05
manufacture.3	0.25	0.24	0.23	0.23	0.23
retail.4	0.14	0.14	0.14	0.13	0.10
information.51	0.06	0.06	0.06	0.07	0.07
banking.52	0.08	0.08	0.07	0.08	0.08
tech.54	0.07	0.07	0.07	0.08	0.09
other.info.sectors.5	0.03	0.04	0.04	0.04	0.04
education.61	0.08	0.08	0.08	0.09	0.09
medical.62	0.09	0.09	0.09	0.09	0.10
arts.7	0.04	0.04	0.04	0.04	0.03
other.service.8	0.02	0.03	0.03	0.03	0.03
federal.government	0.03	0.03	0.03	0.03	0.02
state.government	0.05	0.04	0.05	0.05	0.05
local.government	0.05	0.05	0.05	0.05	0.05
population<100k[f]	0.01	0.01	0.01	0.01	0.01
population.100k-250k	0.08	0.08	0.08	0.08	0.08
pop.250k-1million	0.23	0.23	0.23	0.23	0.23
pop>1million	0.52	0.52	0.51	0.51	0.52
#observations	120,880	124,324	120,984	121,324	120,269

[a]The table reports the mean of each variable.

[b]The dummy equal to 1 if number of employees in establishment ≤ 100.

[c]The dummy equal to 1 if number of employees in establishment > 100 and ≤ 250.

[d]The amount of revenue for each establishment estimated by Harte-Hanks (in $million).

[e]Industry sector of establishment. The number after the name of industry denotes the first digit (or first two digits) of NAICS.

[f]Establishment located in town with less than 100,000 people.

Table 2: Switching Patterns of the Adoption of Operating Systems in Servers[a]

Industry	Year	[Linux_t, Linux_{t-1}] [1,0] (1)	[1,1] (2)	[0,1] (3)	[Windows_t, Windows_{t-1}] [1,0] (4)	[1,1] (5)	[0,1] (6)	[Other_t, Other_{t-1}] [1,0] (7)	[1,1] (8)	[0,1] (9)	total obs.
		A. All Industries and By Year									
	2001	0.030	0.049	0.009	0.110	0.602	0.037	0.045	0.339	0.076	75,353
	2002	0.035	0.065	0.013	0.156	0.635	0.020	0.063	0.299	0.057	77,292
	2003	0.037	0.085	0.015	0.097	0.749	0.020	0.046	0.288	0.063	73,760
	2004	0.025	0.098	0.025	0.065	0.807	0.024	0.030	0.241	0.080	66,501
		B. All Industries and All Years									
		0.032	0.074	0.015	0.108	0.694	0.026	0.047	0.293	0.068	292,906
		C. By Industry and All Years									
agriculture.1		0.018	0.033	0.010	0.109	0.678	0.025	0.040	0.271	0.061	1,334
utility.2		0.023	0.045	0.012	0.120	0.702	0.021	0.041	0.252	0.065	11,862
manufacture.3		0.024	0.048	0.012	0.106	0.703	0.027	0.045	0.325	0.074	74,171
retail.4		0.023	0.052	0.010	0.102	0.599	0.028	0.048	0.370	0.064	29,970
information.51		0.053	0.183	0.023	0.090	0.658	0.032	0.047	0.304	0.065	21,007
banking.52		0.024	0.040	0.010	0.113	0.694	0.022	0.046	0.290	0.067	20,672
tech.54		0.038	0.101	0.018	0.097	0.752	0.021	0.035	0.203	0.057	24,222
other.info.5		0.024	0.049	0.009	0.125	0.660	0.023	0.044	0.241	0.056	7,420
education.61		0.052	0.138	0.027	0.118	0.701	0.026	0.059	0.296	0.070	31,552
medical.62		0.028	0.050	0.012	0.125	0.696	0.024	0.053	0.274	0.072	25,020
arts.7		0.022	0.043	0.012	0.127	0.661	0.024	0.047	0.214	0.058	8,080
other.service.8		0.033	0.063	0.018	0.122	0.710	0.024	0.038	0.230	0.058	6,418
government		0.036	0.065	0.018	0.100	0.747	0.026	0.047	0.284	0.077	31,178

[a]The table reports the share of the establishments that follow a certain switching pattern specified in each column. Samples without any server are excluded.

Table 3: Switching between Different Operating Systems in Servers[a]

Industry	Year	$Windows_t = 0$, $Windows_{t-1} = 1$		$Linux_t = 0$, $Linux_{t-1} = 1$		$Other_t = 0$, $Other_{t-1} = 1$			
		$L_t = 1$, $L_{t-1} = 1$	$L_t = 1$, $L_{t-1} = 0$	$W_t = 1$, $W_{t-1} = 1$	$W_t = 1$, $W_{t-1} = 0$	$L_t = 1$, $L_{t-1} = 1$	$L_t = 1$, $L_{t-1} = 0$	$W_t = 1$, $W_{t-1} = 1$	$W_t = 1$, $W_{t-1} = 0$
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. All Industries and By Year									
	2001	0.0020	0.0016	0.0059	0.0009	0.0040	0.0031	0.0541	0.0090
	2002	0.0023	0.0016	0.0080	0.0029	0.0039	0.0032	0.0407	0.0114
	2003	0.0028	0.0012	0.0105	0.0025	0.0058	0.0029	0.0483	0.0110
	2004	0.0040	0.0011	0.0200	0.0021	0.0087	0.0030	0.0663	0.0096
B. All Industries and All Years									
		0.0027	0.0014	0.0108	0.0021	0.0055	0.0031	0.0519	0.0103
C. By Industry and All Years									
agriculture.1		0.0007	0.0030	0.0052	0.0022	0.0030	0.0007	0.0472	0.0105
utility.2		0.0015	0.0008	0.0083	0.0024	0.0042	0.0026	0.0497	0.0094
manufacture.3		0.0018	0.0012	0.0082	0.0018	0.0037	0.0024	0.0577	0.0106
retail.4		0.0018	0.0011	0.0064	0.0013	0.0040	0.0018	0.0454	0.0107
information.51		0.0067	0.0019	0.0143	0.0034	0.0092	0.0044	0.0474	0.0089
banking.52		0.0012	0.0012	0.0067	0.0017	0.0034	0.0025	0.0500	0.0106
tech.54		0.0031	0.0016	0.0142	0.0018	0.0066	0.0030	0.0441	0.0072
other.info.5		0.0019	0.0019	0.0063	0.0019	0.0024	0.0019	0.0400	0.0115
education.61		0.0057	0.0021	0.0208	0.0032	0.0127	0.0055	0.0533	0.0101
medical.62		0.0015	0.0008	0.0087	0.0019	0.0036	0.0033	0.0538	0.0122
arts.7		0.0020	0.0015	0.0085	0.0017	0.0025	0.0028	0.0408	0.0125
other.service.8		0.0026	0.0009	0.0125	0.0031	0.0042	0.0026	0.0452	0.0090
government		0.0027	0.0014	0.0133	0.0022	0.0062	0.0036	0.0599	0.0103

[a]The table reports the share of the establishments that follow a certain switching pattern specified in each column. Samples without any server are excluded. Note that L_t (or W_t) denotes the use of Linux (or Windows) at period t.

Table 4: Regression Results for Samples with Servers[a]

Variable	Dependent Variable: Linux on Servers		Dependent Variable: Windows on Servers	
	Estimate	P-value	Estimate	P-value
	(1)	(2)	(3)	(4)
intercept	0.000 (0.006)	0.993	0.327 (0.009)	<.0001
pc.windows$_{t-1}$	-0.014 (0.002)	<.0001	-0.032 (0.002)	<.0001
pc.linux$_{t-1}$	-0.025 (0.002)	<.0001	-0.016 (0.003)	<.0001
pc.other$_{t-1}$	-0.002 (0.002)	0.389	-0.003 (0.003)	0.406
server.windows$_{t-1}$	-0.007 (0.001)	<.0001	0.530 (0.001)	<.0001
server.linux$_{t-1}$	0.701 (0.001)	<.0001	-0.043 (0.003)	<.0001
server.other$_{t-1}$	-0.003 (0.001)	0.026	0.007 (0.002)	<.0001
pc.windows	0.000 (0.002)	0.957	0.067 (0.002)	<.0001
pc.linux	0.129 (0.002)	<.0001	0.014 (0.002)	<.0001
pc.other	-0.002 (0.002)	0.484	-0.019 (0.004)	<.0001
server.windows	-0.008 (0.001)	<.0001		
server.linux			-0.017 (0.003)	<.0001
server.other	0.002 (0.001)	0.100	-0.164 (0.002)	<.0001
apache	0.157 (0.002)	<.0001	-0.003 (0.003)	0.250
perl	0.038 (0.004)	<.0001	-0.021 (0.005)	<.0001
utility.2	0.007 (0.003)	0.008	0.019 (0.004)	<.0001
manufacture.3	0.008 (0.002)	0.001	0.011 (0.003)	0.001
retail.4	0.009 (0.002)	<.0001	-0.025 (0.003)	<.0001
information.51	0.029 (0.003)	<.0001	-0.009 (0.004)	0.017
banking.52	0.009 (0.003)	0.001	0.012 (0.004)	0.001
tech.54	0.021 (0.003)	<.0001	0.017 (0.004)	<.0001
other.info.sectors.5	0.009 (0.003)	0.006	0.008 (0.005)	0.085
education.61	0.019 (0.002)	<.0001	0.011 (0.003)	0.000
medical.62	0.005 (0.002)	0.027	0.012 (0.003)	0.001
arts.7	0.006 (0.003)	0.065	-0.001 (0.004)	0.901
other.service.8	0.015 (0.003)	<.0001	0.021 (0.005)	<.0001
federal.government	0.013 (0.003)	<.0001	-0.007 (0.005)	0.118
state.government	0.016 (0.002)	<.0001	0.012 (0.003)	0.000
local.government	0.009 (0.002)	<.0001	0.017 (0.003)	<.0001
revenue	0.000 (0.000)	0.002	0.000 (0.000)	0.153
#white.collar.workers	0.000 (0.000)	0.007	0.000 (0.000)	0.742
#desk.workers	0.000 (0.000)	0.628	0.000 (0.000)	0.699
#company.employees	0.000 (0.000)	<.0001	0.000 (0.000)	0.265
#programmers	0.000 (0.000)	0.727	0.000 (0.000)	0.014
#it.workers	0.000 (0.000)	<.0001	0.000 (0.000)	<.0001

[a]Standard errors in parentheses. Samples without any server are excluded. The establishments that are observed only once during five years are also excluded. The table suppresses coefficient estimates for state dummies.

Table 4: Regression Results for Samples with Servers (continued)[a]

Variable	Dependent Variable: Linux on Servers		Dependent Variable: Windows on Servers	
	Estimate	P-value	Estimate	P-value
	(1)	(2)	(3)	(4)
employee.100	0.000 (0.001)	0.940	0.008 (0.002)	<.0001
employee.250	0.001 (0.001)	0.476	0.011 (0.002)	<.0001
employee.500	0.006 (0.001)	<.0001	0.015 (0.002)	<.0001
employee.750	0.005 (0.002)	0.012	0.022 (0.003)	<.0001
employee.1,000	0.014 (0.002)	<.0001	0.016 (0.004)	<.0001
employee.1,500	0.015 (0.003)	<.0001	0.013 (0.004)	0.001
employee.2,000	0.021 (0.004)	<.0001	0.013 (0.005)	0.015
employee.3,000	0.011 (0.004)	0.007	0.010 (0.006)	0.106
employee.5,000	0.018 (0.005)	0.000	0.015 (0.007)	0.044
employee.10,000	0.018 (0.007)	0.014	-0.019 (0.010)	0.066
employee.100,000	0.048 (0.013)	0.000	-0.027 (0.019)	0.163
non-pc.purchase[b]	0.003 (0.001)	0.001	0.026 (0.002)	<.0001
pc.purchase[c]	0.010 (0.001)	<.0001	0.001 (0.001)	0.359
population<100k	0.006 (0.004)	0.143	-0.002 (0.006)	0.708
population.100k-250k	0.005 (0.002)	0.002	0.001 (0.002)	0.553
pop.250k-1million	0.003 (0.001)	0.024	0.004 (0.002)	0.021
pop>1million	0.001 (0.001)	0.591	0.005 (0.002)	0.006
#pc	0.000 (0.000)	0.782	0.000 (0.000)	0.009
#non-pc	0.000 (0.000)	0.016	0.000 (0.000)	0.240
#internet.server	0.000 (0.000)	0.856	0.000 (0.000)	0.084
#lansv	0.000 (0.000)	0.030	0.000 (0.000)	0.005
#lan	0.022 (0.001)	<.0001	0.085 (0.001)	<.0001
#pc.server	0.000 (0.000)	0.023	0.000 (0.000)	<.0001
#internet.users	0.000 (0.000)	0.215	0.000 (0.000)	<.0001
#internet.developers	0.000 (0.000)	0.892	0.000 (0.000)	0.057
R^2		0.576		0.474
#observations		292,906		292,906

[a]Standard errors in parentheses. Samples without any server are excluded. The establishments that are observed only once during five years are also excluded. The table suppresses coefficient estimates for state dummies.

[b]The dummy = 1 if the establishment, not the parent company, makes a decision to purchase non-PCs.

[c]The dummy = 1 if the establishment, not the parent company, makes a decision to purchase PCs.

Table 5: Standard Probit Results

Variable	Sample Period: 2000-2002 Estimate		P-value	Sample Period: 2001-2003 Estimate		P-value	Sample Period: 2002-2004 Estimate		P-value
	Info51, Linux server adoption								
Intercept	-1.4002	(0.0533)	<.0001	-1.3492	(0.0551)	<.0001	-1.2387	(0.0603)	<.0001
$server.linux_{t-1}$	2.6418	(0.0456)	<.0001	2.4779	(0.0416)	<.0001	2.5243	(0.0417)	<.0001
$server.windows_{t-1}$	-0.0966	(0.0387)	0.0126	-0.1399	(0.0374)	0.0002	-0.3036	(0.0409)	<.0001
$pc.linux_{t-1}$	0.4496	(0.0458)	<.0001	0.5112	(0.0409)	<.0001	0.4673	(0.0414)	<.0001
$pc.windows_{t-1}$	-0.1604	(0.0554)	0.0038	-0.1027	(0.0559)	0.0659	-0.1215	(0.0601)	0.0434
	Info51, Windows server adoption								
Intercept	-0.5706	(0.0437)	<.0001	-0.4808	(0.0477)	<.0001	-0.5634	(0.0537)	<.0001
$server.linux_{t-1}$	-0.3101	(0.0458)	<.0001	-0.4114	(0.0444)	<.0001	-0.3902	(0.0461)	<.0001
$server.windows_{t-1}$	2.1625	(0.0335)	<.0001	2.2419	(0.0355)	<.0001	2.3379	(0.0379)	<.0001
$pc.linux_{t-1}$	-0.1581	(0.0461)	0.0006	-0.1898	(0.0448)	<.0001	-0.1645	(0.0460)	0.0003
$pc.windows_{t-1}$	0.1380	(0.0464)	0.0030	0.1631	(0.0502)	0.0012	0.1558	(0.0557)	0.0051
	Med62, Linux server adoption								
Intercept	-1.7394	(0.0673)	<.0001	-1.7610	(0.0697)	<.0001	-1.7714	(0.0814)	<.0001
$server.linux_{t-1}$	2.7172	(0.0682)	<.0001	2.5486	(0.0578)	<.0001	2.5150	(0.0523)	<.0001
$server.windows_{t-1}$	0.0025	(0.0459)	0.9569	0.0462	(0.0452)	0.3060	-0.0333	(0.0519)	0.5215
$pc.linux_{t-1}$	0.6153	(0.0755)	<.0001	0.7115	(0.0620)	<.0001	0.5247	(0.0588)	<.0001
$pc.windows_{t-1}$	-0.2235	(0.0680)	0.0010	-0.1521	(0.0676)	0.0244	-0.1189	(0.0769)	0.1220
	Med62, Windows server adoption								
Intercept	-0.1555	(0.0458)	0.0007	0.1492	(0.0521)	0.0042	0.3088	(0.0662)	<.0001
$server.linux_{t-1}$	-0.1865	(0.0750)	0.0129	-0.2398	(0.0698)	0.0006	-0.3694	(0.0672)	<.0001
$server.windows_{t-1}$	1.9453	(0.0310)	<.0001	1.8991	(0.0347)	<.0001	1.9718	(0.0380)	<.0001
$pc.linux_{t-1}$	-0.0340	(0.0751)	0.6504	-0.1257	(0.0694)	0.0702	-0.1030	(0.0679)	0.1292
$pc.windows_{t-1}$	-0.0734	(0.0481)	0.1269	-0.0867	(0.0545)	0.1115	-0.2693	(0.0689)	<.0001

Table 6: Random Effects Probit Results

Variable	Sample Period: 2000-2002 Estimate		P-value	Sample Period: 2001-2003 Estimate		P-value	Sample Period: 2002-2004 Estimate		P-value
				Info51, Linux server adoption					
Intercept	-1.400244	(.0533395)	0.000	-1.349243	(.0551476)	0.000	-1.238712	(.0602881)	0.000
server.linux$_{t-1}$	2.641811	(.0455773)	0.000	2.477917	(.0415612)	0.000	2.524325	(.0417189)	0.000
server.windows$_{t-1}$	-.0965654	(.0387081)	0.013	-.1398623	(.037433)	0.000	-.3035643	(.0408982)	0.000
pc.linux$_{t-1}$	.4495744	(.0457652)	0.000	.5111741	(.0409005)	0.000	.4673008	(.0413852)	0.000
pc.windows$_{t-1}$	-.1603683	(.0554179)	0.004	-.1027368	(.0558583)	0.066	-.1214683	(.0601437)	0.043
				Info51, Windows server adoption					
Intercept	-.5705755	(.043696)	0.000	-.4853501	(.0532402)	0.000	-.5634294	(.0537475)	0.000
server.linux$_{t-1}$	-.310072	(.0457597)	0.000	-.4503175	(.0505376)	0.000	-.3901895	(.0460826)	0.000
server.windows$_{t-1}$	2.1625	(.033462)	0.000	2.387153	(.0541162)	0.000	2.337943	(.0378976)	0.000
pc.linux$_{t-1}$	-.1581401	(.046057)	0.001	-.2072762	(.0501188)	0.000	-.1644641	(.045993)	0.000
pc.windows$_{t-1}$	.1379856	(.0464407)	0.003	.1836996	(.0561837)	0.001	.1558027	(.0556834)	0.005
				Med62, Linux server adoption					
Intercept	-1.73944	(.067324)	0.000	-1.761028	(.0696533)	0.000	-1.771358	(.0814)	0.000
server.linux$_{t-1}$	2.717215	(.0682205)	0.000	2.548592	(.0577892)	0.000	2.514998	(.0523366)	0.000
server.windows$_{t-1}$	.0024822	(.0459037)	0.957	.0462143	(.0451506)	0.306	-.0333001	(.0519417)	0.521
pc.linux$_{t-1}$	.6152779	(.0755467)	0.000	.7114679	(.0620186)	0.000	.5247	(.0587834)	0.000
pc.windows$_{t-1}$	-.223539	(.0680021)	0.001	-.1521095	(.0675794)	0.024	-.118874	(.0768634)	0.122
				Med62, Windows server adoption					
Intercept	-.1555358	(.0457627)	0.001	.2231923	(.0605591)	0.000	.4177865	(.0772992)	0.000
server.linux$_{t-1}$	-.1864971	(.0749639)	0.013	-.2535727	(.0818282)	0.002	-.4152562	(.0790975)	0.000
server.windows$_{t-1}$	1.945318	(.0310282)	0.000	2.056138	(.0428088)	0.000	2.148032	(.0468263)	0.000
pc.linux$_{t-1}$	-.0340454	(.075128)	0.650	-.1365703	(.0810999)	0.092	-.1024075	(.0802262)	0.202
pc.windows$_{t-1}$	-.0734237	(.0481018)	0.127	-.0779942	(.0627839)	0.214	-.2836442	(.0796567)	0.000

Table 7: Main Estimation Results

Variable	Sample Period: 2000-2002 Estimate	Std. Error	Sample Period: 2001-2003 Estimate	Std. Error	Sample Period: 2002-2004 Estimate	Std. Error
	Info51, Linux server adoption					
$server.linux_{t-1}$	1.0271	(0.1518)	2.2612	(0.0996)	2.3741	(0.2138)
$server.windows_{t-1}$	-0.0252	(0.0646)	-0.2068	(0.0758)	-0.0668	(0.1200)
$pc.linux_{t-1}$	-0.0693	(0.0953)	0.4805	(0.0834)	0.3601	(0.1401)
$pc.windows_{t-1}$	-0.1540	(0.1259)	-0.0802	(0.1781)	0.1116	(0.2356)
γ_{01}	-0.5154	(0.0536)				
γ_{02}	0.1280	(0.0409)	-0.0613	(0.0355)		
γ_{03}			-0.0610	(0.0428)	-0.1996	(0.0437)
γ_{04}					-0.2404	(0.0593)
	Info51, Windows server adoption					
$server.linux_{t-1}$	-0.0514	(0.1229)	-0.6463	(0.2209)	-0.3431	(0.2447)
$server.windows_{t-1}$	1.4118	(0.0561)	2.0071	(0.1800)	1.2020	(0.2206)
$pc.linux_{t-1}$	0.0102	(0.1083)	0.4885	(0.1388)	-0.8150	(0.1735)
$pc.windows_{t-1}$	0.1619	(0.1095)	0.2015	(0.1828)	-0.6782	(0.2780)
γ_{01}	0.8460	(0.0556)				
γ_{02}	0.3783	(0.0738)	0.6384	(0.0639)		
γ_{03}			-0.0638	(0.0491)	0.4475	(0.0726)
γ_{04}					-0.1997	(0.0476)
	Med62, Linux server adoption					
$server.linux_{t-1}$	0.4139	(0.1543)	1.6959	(0.1471)	2.4883	(0.0907)
$server.windows_{t-1}$	0.0407	(0.0573)	-0.0950	(0.0667)	-0.2166	(0.1038)
$pc.linux_{t-1}$	-0.1248	(0.1494)	-0.4513	(0.4169)	0.7617	(0.1512)
$pc.windows_{t-1}$	-0.1634	(0.0984)	-0.4762	(0.1266)	-0.2526	(0.1839)
γ_{01}	-0.3458	(0.0544)				
γ_{02}	0.2098	(0.0369)	0.1111	(0.0329)		
γ_{03}			-0.0108	(0.0390)	-0.0654	(0.0370)
γ_{04}					-0.1640	(0.0468)
	Med62, Windows server adoption					
$server.linux_{t-1}$	-1.7921	(0.4499)	0.0744	(0.3834)	-0.5694	(0.2092)
$server.windows_{t-1}$	1.6760	(0.0411)	2.0296	(0.1983)	1.7164	(0.2573)
$pc.linux_{t-1}$	0.0270	(0.1565)	-0.2383	(0.1524)	-0.1347	(0.1957)
$pc.windows_{t-1}$	-0.5323	(0.0885)	-0.0144	(0.2279)	-0.0437	(0.1925)
γ_{01}	0.9545	(0.0461)				
γ_{02}	0.2590	(0.0459)	0.6366	(0.0579)		
γ_{03}			0.0247	(0.0460)	0.3064	(0.0541)
γ_{04}					-0.0489	(0.0436)