NET Institute*

www.NETinst.org

Working Paper #05-17

October 2005

Competing Technologies in the Database Management Systems Market

Tobias Kretschmer
London School of Economics

* 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.

Competing technologies in the database management systems market[*]

Tobias Kretschmer[**]

London School of Economics and *CEP*

Abstract

In this paper, we study the dynamics of the market for Database Management Systems (DBMS), which is commonly assumed to possess network effects and where there is still some viable competition in our study period, 2000 – 2004. Specifically, we make use of a unique and detailed dataset on several thousand UK firms to study individual organizations' incentives to adopt a particular technology. We find that there are significant internal complement effects – in other words, using an operating system and a DBMS from the same vendor seems to confer some complementarities. We also find evidence for complementarities between enterprise resource planning systems (ERP) and DBMS and find that as ERP are frequently specific and customized, DBMS are unlikely to be changed once they have been customized to an ERP. We also find that organizations have an increasing tendency to use multiple DBMS on one site, which contradicts the notion that different DBMS are near-perfect substitutes.

Keywords: Database software, indirect network effects, technology adoption, microdata.

JEL Codes: L86, O33.

---

[*] Preliminary, comments welcome. This research has been supported by a grant from the NET Institute (http://www.netinst.org), and purchase of the data has been financed by the ESRC, the Anglo-German Foundation, and the Interdisciplinary Institute of Management at the LSE. I thank Benedikt Gamharter for excellent research assistance.

[**] Address for correspondence: Tobias Kretschmer, Interdisciplinary Institute of Management, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Tel: +44 207 955 6041, Fax: +44 207 955 6887, email: t.kretschmer@lse.ac.uk.

INTRODUCTION

Standards battles are common in many industries. Whenever there exist significant network effects or demand-side economies of scale, it is likely that one version of the technology emerges as the *industry standard* (Arthur, 1989). Since an industry standard often guarantees monopoly profits over a period of time, firms will expend significant resources on winning the race for it. Further, since de-facto standards are likely to persist for some time, settling on the "wrong" standard can have important welfare implications (Cabral and Kretschmer, forthcoming).

One area in which standards battles are especially prevalent is computer software. The existence of network effects in specific software markets has been documented by several studies (Brynjolfsson and Kemerer, 1996, Gandal, 1994, Gandal et al., 1999), and recent history has shown that software markets tend to settle on a single technology that often remains dominant over several product generations (Kretschmer, 2004, Liebowitz and Margolis, 2001) – Microsoft is the best-known example of vendor dominance, but there is ample evidence of similar processes occurring in other software industries – for example, SAP R/3 in ERP Software, Apache for web-server Software[1] and Google in search engines.

In this paper, we study the dynamics of the market for Database Management Systems (DBMS), which is commonly assumed to possess network effects and where there is still some viable competition in our study period, 2000 – 2004. Specifically, we make use of a unique and detailed dataset on several thousand UK firms (*LSE micro-data set on Information and Communication Technologies*, ICT-LSE, described in the data section) and complement it with in-depth information about the DBMS market to study individual firms' incentives to adopt a particular technology. The data allows us to look at the use of complementary technologies *within the firm* over time and to assess their effect on organizations' DBMS choice. Specifically, we look at two technologies that are complementary to DBMS – Enterprise Resource Planning systems (ERP) and operating systems (OS). This is one of the first studies to explicitly consider *internal complement effects* (ICE) and their impact.

We find that there are significant internal complement effects, even on the vendor level. In other words, using an operating system and a DBMS from the same vendor

[1] See, e.g., http://news.netcraft.com/archives/web_server_survey.html.

seems to confer some complementarities. We also find evidence for complementarities between ERP and DBMS and find that as ERP are frequently specific and customized, DBMS are unlikely to be changed once they have been customized to an ERP. We also find that organizations have an increasing tendency to use multiple DBMS on one site, which contradicts the notion that different DBMS are near-perfect substitutes.

EXISTING LITERATURE

In this section, we highlight some of the issues that previous studies have had to deal with and highlight how this paper will address these issues.

EXISTING LITERATURE AND DATA ISSUES

The study of software markets and software standards battles in particular has often been limited by the availability and quality of data, and the proposed study intends to address some of these shortcomings. Existing studies have two features of their data to identify and test for network effects:

Tracking aggregate usage. Many papers look at aggregate usage figures to proxy for the sum of individual decisions (e.g. Gandal et al., 1999, Ohashi, 2003, Brynjolfsson and Kemerer, 1996, Bayus and Shankar, 2003, Koski and Kretschmer, 2005). This is useful for gaining a general idea of the strength of network effects. Put crudely, the residual in a demand or a willingness-to-pay function is interpreted as *network effects*, since other potential demand shifters are controlled for. However, this does not consider the effect of individual users' characteristics since the distribution of unobserved adopter characteristics are assumed to be constant over time, which is at odds with standard practice in marketing studies (Rogers, 2003).[2]

Single technology history. Most studies study a single technology to analyze standards battles and network industries.[3] One of the most prominent features of software markets however is that significant complementarities exist across related, but

[2] There are some exceptions, however. Greenstein (1993) uses microdata on governmental purchase decisions to track computer diffusion in the US, and Breuhan (1997) uses individual switching decisions of firms from one word processing software to another to estimate the extent of switching cost across different vendors, and Astebro (2004) analyses firms' decision to adopt CAD and CNC.
[3] See, for example, Liebowitz and Margolis (2001), Rohlfs (2001) or Sarnikar (2002) for studies on individual software industries.

different products, such as spreadsheets and word processing software, or database applications and Enterprise Resource Planning (ERP) systems. Complementarities in network industries have only been identified so far for pure complements such as CDs and CD players (Gandal et al., 2000). Failing to control for the effect of complementary products will tend to overstate network effects originating from the studied technology.

Previous studies on network industries have been useful in establishing a general set of conclusions on the existence and strength of network effects in network industries. However, the lack of data on individual users' decisions and characteristics has made it difficult to go beyond this. This paper will use a new *LSE micro-data set on Information and Communication Technologies* (ICT-LSE, described in the data section) to deal with some of the problems outlined above.

THE DATABASE MANAGEMENT SYSTEMS MARKET – HISTORY AND DYNAMICS

Ever since groups of people worked or lived together the requirement existed to store, handle and access great amounts of data for administration or information purposes. In the 1960s when the very first computers emerged initial efforts to manage data by utilizing information technology (IT) began. This section first describes the history and key stages of the relational database market. It then introduces the key competitors and the nature of competition in the DBMS market.

Stand-alone relational databases

The 1970s brought a big leap forward for the database market. In 1970 Edgar Codd invented the relational model for data storage. Codd was a scientist working for IBM at its San Jose Research Laboratory. His paper on database management[4] was a landmark publication and is understood to be the theoretical foundation for the relational database market as we know it today. First derivatives of the Structured Query Language (SQL) were also developed mainly by IBM in this initial phase.[5]

The fundamental building blocks of relational databases are tables of data and queries to link the data. All data is stored "centrally" and referred to as required. If the data is

[4] A Relational Model of Data for Large Shared Data Banks Communications of the ACM, Vol. 13, No. 6, June 1970, pp. 377-387, see http://www.acm.org/classics/nov95/toc.html.
[5] SQL is a computer language to create, modify and retrieve data from databases.

updated once, all records using this data reflect the change. Queries for instance create, change, link, manipulate or structure data and thereby create records. The logic behind the relational database structure is to simplify data maintenance and to reduce the chance of having duplicate data or inconsistencies which occur if data is not centrally maintained.

IBM research launched a research project, codenamed System R, to develop a database system based on Codd's idea and SQL. System R led to IBM's first commercial database which was launched in 1979 and its flagship relational database, called DB2, in the early 1980s. Although IBM was leading R&D efforts, other companies were faster when it came to introducing the first products. Honeywell introduced the first relational database as early as 1976 and Oracle its first version in 1979.

Behind the DBMS system of Honeywell is Charles Bachman; another key innovator in the DBMS field, who initially developed a database system for General Electric when the market was still in its infancy and GE still maintained an in-house Computer Division that was subsequently spun out and merged with Honeywell in the 1970s. During this short stint the Honeywell DBMS system (Integrated Data Storage) was brought to market.

Oracle was founded by Larry Ellison and four partners in 1977 in Redwood, CA. Inspired by Codd's paper they wanted to commercially explore the database market. Their initial idea was to partner with IBM in developing a database system. But IBM kept its product development efforts secret and had no interest to get the young team on board. Hence, Ellison and his team decided to launch their own database product. The highly motivated team unexpectedly managed to outrace "Big-Blue" and successfully launched a database product. Ever since the 1970s Oracle has been competing in the database market with great success.[6]

Personal Computers and Client/Server relational databases

In the 1970s, all aspects of the computer industry were dominated by IBM. The first database technologies were developed for the server market, e.g. various IBM operating systems, UNIX, etc. The next big effect on the database market was the

[6] In the recent past, Oracle has been buying up rival DBMS vendors in a series of hostile and friendly takeovers and is regarded as one of the most powerful players in the database market (see: Oracle and Siebel, *The Economist*, Sept. 15, 2005) and the second largest software house after Microsoft.

vertical disintegration of the computer industry and the emergence of the minicomputer and later personal computer segments in the 1980s. During this wave new players entered the database market. The most prominent companies to enter the field were Sybase and Informix. Of both companies the founders and intellectual capital came from an earlier database research project called Ingres at Berkeley University. Those companies entered the market to race for market share in the non IBM-dominated mainframe market (for instance Microsoft, HP-Unix, etc.). Sybase as well as Oracle introduced a client/server relational database in the late 1980s utilizing new technologies.

Microsoft was selected by IBM to develop its operating system for the desktop market.[7] Initially IBM and Microsoft's efforts were closely linked. The development of the desktop operating system DOS and early versions of Windows were joint efforts. As Microsoft grew in size it also increasingly gained independence. IBM and Microsoft's operating software alliance ended when cooperation on their jointly developed OS/2 broke down. At that time Microsoft also decided to compete against IBM, Sun, Digital and other firms in the server market. In 1988 IBM launched its OS/2 and Microsoft its first server software, Windows NT. Microsoft had just entered the server market and had not yet developed its own database capabilities. It therefore explored the marked and looked for know-how to fill the gap. The company Sybase was working in the Unix database segment and Ashton-Tate operated in the desktop market. Ashton-Tate's core product was called dBASE (which was bought by Borland in Sept. 1991). Microsoft signed a license agreement for Sybase's technology. The two firms partnered operationally and ported its Unix Sybase SQL Server technology into the Windows NT environment. The partnership continued until the early 1990s. In 1993 Microsoft decided to use Sybase's technology as a basis but to develop its own version, Microsoft SQL server technology. Sybase was forced to rename its core database product to Adaptive Server Enterprise to differentiate the products in subsequent years.

Internet and enterprise application integration

In the 1990s the two major advances in the computer industry were the commoditization of PC hardware and the emergence of the internet. They created a

[7] For a brief history of the PC Operating Systems market, see Kretschmer (2004).

new focus for providers of enterprise system solutions: to build integrated or interconnected systems. This changed the solution space for classical database manufacturers away from providing data management to providing information management solutions for networked organizations.

With the proliferation of computer networks throughout organization system architects created a new data-layer in the IT topology. Emerging middleware technologies supported this layer. One reason was that more robust and scalable enterprise software design was possible by moving to this multi-tier architecture. Another reason was the deployment of applications. Database management firms identified this segment and enabled their applications to support various upstream applications. For instance, Sybase, launched its own middleware technology in 1994.

In order to ensure application integration across platforms database functionality was changed to utilize internet technologies wherever possible. In 1997 Oracle for instance moved its client/server application to the web. It also launched its first web-based database. Further, database manufacturers supported the development of new open protocols and standards. Oracle launched the first database with XML support. Oracle9i (the “i” standing for “Internet”) was a complete information management suite from Database, Application Server to Developer Suite. Subsequently, modern databases were developed to support Microsoft's Open Database Connectivity (ODBC) interface, the Java Database Connectivity (JDBC) interface, or a CORBA interface broker and thereby allow various front-end systems to interact with databases.

Multidimensional databases and Business Intelligence

Another database solution space formed around the Internet-Web-Data integration push of the late 1990s and the early phase of the new millennium. The market place was looking for open-standard database technologies which could be customized and deployed cost effective hand-in-hand with other open-source technologies. An example of a successful player in this segment is MySQL. MySQL for instance can be used with Apache Web-Server technology running on Linux. Such a solution provides an end-to-end open-source solution. However, established firms also tried to utilize opportunities from the open source market. Sybase, for example, was the first established vendor to provide software for the Linux operating system a market which was entered thereafter also by Oracle and IBM.[8]

One key emerging theme of the 2000s was the convergence of computing and communication technologies. Research and development efforts in all segments were launched; from new communication protocols to improvements to numerous aspects of mobile hardware technology (e.g. displays for mobile devices, batteries). As in all other segments from semiconductor to handheld manufacturers, the database industry was making big strides forward in this direction. The mobile database market was soon populated with products such as Sybase SQL Anywhere, IBM DB2 Everyplace, Microsoft SQL Server 2000 Windows CE Edition or Oracle 9i Lite competing for market share.

The last 30 years have revolutionized the data management industry. The progress in database technology played a pivotal role for the exponential growth of the IT industry. The current focus of database management systems is to provide data warehousing capabilities and to enable grid computing. Data warehouses are repositories for all sorts of enterprise data: they store everything from human resource data (such as payroll-information) over financial data (such as sales forecasts across product groups or actual unit sales data) to supply-chain details (such as stock-inventory information). In 2002 Oracle launched its first fully integrated relational and multidimensional database. Multidimensional databases are used for data warehouses when two-dimensional tabular structures are not sufficient anymore to represent the relationships between data. Since SQL is not sufficient to query multidimensional databases, a standard called Online Analytical Processing (OLAP) was developed. OLAP data marts or cubes are generated out of data warehouses.
In order to streamline IT hardware costs, the current trend is to move away from using standalone high-specification servers but to use server farms (grids) of low-specification. Oracle developed its first database which can be run on such a grid environment; Oracle 10g ("g" – standing for "Grid").

Competition and Product Pricing

Since 2000 three key players dominate the market: IBM, Microsoft and Oracle. They compete very vigorously and constantly fight for market leadership. Depending which

[8] For an overview of the economics of open-source, see Lerner and Tirole (2002).

industry league table or consultant you trust, any of the three is the current market leader.

Microsoft has two primary DBMS systems: MS Access and MS SQL Server. MS Access is sold as part of the Microsoft Office Professional Suite. MS Access is a one-size-fits-most application and broadly serves two consumer types. Firstly, it is used by end users creating their own databases in a Windows environment. In this segment of the DBMS market MS Access has virtually no competition and is the de facto standard. The key benefit of MS Access is its graphical user interface which enables users to quickly design stand-alone applications without the need to learn how to code software. Users can not only define tables and queries but they can also create front-end screens and reports within MS Access. Secondly, Access is used by application developers for small development jobs or when prototyping.

MS SQL Server is Microsoft's professional database application. System developers can use MS SQL Server as the back-end technology when building software applications. The MS SQL Server development environment is very user-friendly and professionals familiar with other MS development tools can quickly start programming in SQL. MS SQL Server is available in various editions. The basic, or Free, edition can be downloaded from the internet. This edition has limited functionality and the licence restricts the database size as well as the hardware specification on which the DBMS can be installed (e.g. CPU type, memory size). This edition is suitable for small DBMS jobs. However, it predominantly eliminates the hold-up problem associated with the investment decision for potential buyers. Users can freely test and learn the basics of the tool prior to investing into the technology. Alternatively, MS SQL Server is available as the Basic-, Standard- and Enterprise edition. The editions differ marginally in functionally but mainly around the size of database supported. Scalability is the key driver for product price. Users can buy three types of MS SQL Server licence depending on their requirement: a processor licence, a server-plus-device licence or a server-plus-user-client-account licence (CALs). For scalable databases or applications where the number of users is hard to quantify, for instance a database behind an internet information site, firms are advised to purchase a processor licence.

The other types of licences are for applications with a defined set of devices or users. Hence, such licenses can be priced more specifically. Historically, MS has priced its

DBMS systems very aggressively and transparent making it a very attractive option for small and medium size businesses.[9] For example, MS hardly charges users for additional functional add-ons (e.g. specific data-mining tools) and therefore the total cost of ownership can be easier quantified.

IBM is a key player in the database market. As described, IBM has a long history in the market and sponsored many of the product innovations. Historically, IBM developed applications for its own hardware and operating system. Today IBM's databases run virtually on all servers. Again, IBM's R&D is at the forefront and IBM is an expert in providing tailormade solutions. Its core DBMS system is DB2 which comes in various editions such as a dedicated for mobile applications (DB2 Everywhere), a free edition for developers (DB2 Personal Developer) or various Workgroup editions to mention a few. A particular edition of DB2 is its Warehouse Enterprise Edition. The focus of this series is on business intelligence. It comes with tools to create cubes, data-mining tools, scoring, modelling and other analytical libraries. IBM uses very sophisticated licensing. Although this creates custom-made solutions for clients, it also comes with a tradeoff, as users will need to understand their requirements in detail and need the expertise to select the most suitable bundle.

Historically Oracle's databases have been used predominantly for large-scale applications. Oracle technology is renowned for its reliability and focus on performance and scalability. With the proliferation of computing Oracle has been pushing hard to enter the small and medium-size business market over the last years.[10] Oracle launched Oracle9i and Oracle10g between 2000-2004. Both products came in a variety of editions tailored for various markets (e.g. Standard One, the Standard, and the Enterprise Edition). Oracle targets the small and medium business segment with its Standard One Edition. To compete with Microsoft and to get users to switch to Oracle in these segments, the Standard One Edition is priced on a par with the Basic Edition of MS SQL Server. Oracle software can be run on Microsoft Windows, Linux, Solaris and Unix servers.

All three companies are key player in the DBMS market but also powerhouses in other segments of the software industry. It will be interesting to observe how the

---

[9] See, for example, http://www.microsoft.com/sql/evaluation/compare/pricecomparison.mspx.
[10] On the competitive nature of Oracle's pricing, see http://www.computerworld.com/managementtopics/management/itspending/story/0,10801,61398,00.html.

market will develop. Will one of the firm win the race to lead the market due to complementary products it owns, e.g. ERP systems? Will it continue to be a close race and, if so, why? While we will not address these questions in detail, we hope to shed some light on the historical developments in the markets.

RESEARCH QUESTION AND METHODOLOGY

The main focus of this paper is the **switching and usage behaviour of individual firms.** We are particularly interested in the effect of a number of potentially complementary technologies, and the degree of substitutability or complementarity among different DBMS.

To do this, we will utilise a number of unique features of the ICT-LSE dataset: a) the panel nature of the data will allow tracking not only the choice of *which technology* to adopt, but also the choice of *when* to adopt. Using a panel with detailed firm characteristics will also reveal determinants of the timing and nature of technology choices. b) detailed information on a large number of technologies used enables me to isolate internal complement effects (ICE), which was not possible with existing, single-technology data.

BASELINE REGRESSIONS AND VARIABLES

Our first set of regressions will simply look at the static usage decisions at a site. We run a simple logit model and a random-effects logit panel regression where the dependent variable is 1 if a firm uses a particular DBMS at time t, and 0 otherwise. We then run regressions on the likelihood of a site to start using (to “switch” to) a specific DBMS at time t. Finally, we also look at the likelihood of a site abandoning their existing DBMS and starting use of another one (“competitive switching”).

We perform several robustness tests, including different sets and specifications of independent variables and controls, but also different regression models (e.g. logit regression) and running our basic regressions on early and late users of DBMS separately.

In order to uncover potential internal complement effects, we use the following covariates (Variable definitions and descriptive statistics can be found in the Appendix).

Enterprise Resource Planning Systems (ERP). The ERP market has been dominated by SAP for the last decade or so. The complementarities between ERP and DBMS are obvious, as ERP rely on vast amounts of data, which can be made available by DBMS. ERP are typically bigger in scope than DBMS. However, if ERP could fulfil all the functions of a DBMS equally well or better, using an ERP might make DBMS usage less likely. We use separate variables for SAP and other ERP to see if there is a specific effect from using the most frequently used ERP.

Operating Systems (OS). DBMS, like any other software, has to run on top of an operating system. As most software firms are multiproduct firms and compatibility with specific OS is a strategic variable by DBMS vendors (or vice versa), it will be interesting to see if usage of a particular OS has an impact on the usage of specific software.[11] This is of particular interest for the DBMS market, since both IBM and Microsoft had a significant presence in both markets, and Oracle had been designed to be "portable", i.e. working equally well on different OS. Again, we will uncover the existence and extent of internal complement effects with OS.

Connectedness. Sites that use multiple servers or are connected via a Wide Area Network (WAN) are prone to have higher and different demands on their DBMS. At the same time, it is possible that an ERP is a better substitute for DBMS for "connected" firms. Our regressions will help uncover the effects of connectedness in two forms – the number of servers, which captures the sheer data volume, and number of networked "locations" on site – which proxies for a specific type of use for a DBMS.

IT Expertise and Intensity. The effect of higher IT expertise and intensity can be twofold. First, IT intensive firms are more likely to customize their software, which would suggest that using a single, user-friendly DBMS is attractive. On the other hand, more IT-savvy organizations can "cope" more easily with multiple DBMS because their absorptive capacity is likely to be higher (see, e.g. Cohen and Levinthal, 1990, Matusik and Heeley, 2005). We are looking for the net effect of these two conflicting forces by including the number of IT employees (as a proxy for the general IT intensity) and the number of IT developers (as proxy for the expertise in programming) in our regressions.

[11] Kretschmer (2004) studies the reverse question and looks at the effect of office applications software on the usage patterns of PC Operating Systems.

Site Size. Larger firms may have different demands on their DBMS. Apart from the higher quantity of data, larger firms may store different data and have to regulate access and security more elaborately. Including site size and control variable will therefore help distinguish between DBMS that are particularly suited for larger sites and ones that can be used cost-efficiently in smaller sites.[12]

THE DATA

The dataset we use is built from a large ICT firm-level panel and matched firm characteristics, which will be described in more detail below.

Harte-Hanks ICT data

Harte-Hanks (HH) is a global company that collects IT data primarily for the purpose of selling on to large producers and suppliers of IT products (e.g. IBM, Dell etc). Their data is collected for roughly 16,000 sites in the UK over a period of 2000 to the present day.[13] Harte-Hanks surveys sites on a rolling basis with an average of 11 months between surveys. This means that at any given time, the data provides a "snapshot" of the stock of a firm's IT.

The fact that HH sells this data on to major firms like IBM and Cisco, who use this to target their sales efforts, exerts a strong market discipline on the data quality. If there were major discrepancies in the collected data this would be rapidly be picked up by HH's clients when they placed sales calls using the survey data, and would obviously be a severe problem for HH future sales. Because of this HH runs extensive internal random quality checks on its own data, enabling them to ensure high levels of data accuracy.

Cleaning Process

The data comes in yearly slices and had to be assembled as a panel. In its raw form, for each software application used in the firm a quantity is given, which would enable the calculation of actual market shares. However, these numbers are not reliable – some sites report a site license for a particular software program as quantity =1, while some sites will count the number of on-site PCs to derive software numbers. Therefore, quantity data has been dropped from the dataset. Further, sites that do not

---

[12] As discussed in the section on pricing and competition, DBMS vendors grant quantity discounts for larger firms and sites.

[13] In fact, Harte-Hanks has been collecting data since the early 1990s, but due to reporting and surveying inconsistencies we focus on the 2000 – 2004 time period.

report complete information on auxiliary characteristics like size, sic code, number of servers etc. have been dropped form the sample, as a key research question is the effect of such characteristics on technology adoption. Finally, for some estimations, we balance the panel in order to strip out the effects of failing firms (which would be less likely to upgrade their current technology) and entering firms (which have the latest vintage software and are unlikely to upgrade quickly). Our results with the balanced panel are therefore likely to overestimate the likelihood of switching. Finally, for some regressions we also restrict our sample to sites that have been using DBMS in the first year of our sample, 2000. This enables us to track shifts in market share rather than new additions to the user population.

RESULTS

We generate results via four different lenses. First, we take a look at the descriptive statistics and point out some noteworthy patterns of our data. We then run simple usage regressions to determine what makes usage of a particular database management system more likely. We then adopt a dynamic perspective and consider the decision to *switch into* a particular database management system. Finally, we consider the case in which *switching into* a DBMS implies *switching away* from another one. We finally combine all our results to gain a more complete picture of the nature of competition in the DBMS industry.

Some Stylized Facts

Descriptive statistics of our full dataset, the balanced panel, as well as the early and late user groups can be found in the Appendix. Figure 1a shows the usage shares of the main DBMS of all sites in our dataset.[14] We can see that overall usage is increasing, and in 2004 almost 90% of sites are using at least one DBMS. Figure 1b only considers sites that have been using a DBMS in the first year of our sample, 2000, and Figure 1c looks at the sites that have not been using any DBMS in 2000. We can see that the tendency to use multiple DBMS remains as Figure 1a and 1b show a similar picture. Note that Figures 1b and 1c make up the net effect shown in 1a.

[14] Note that usage shares are not synonymous with market shares, as a firm can use more than one DBMS, which implies that the shares can add up to more than 100%.

Figure 1a: Usage shares of major DBMS, full sample.

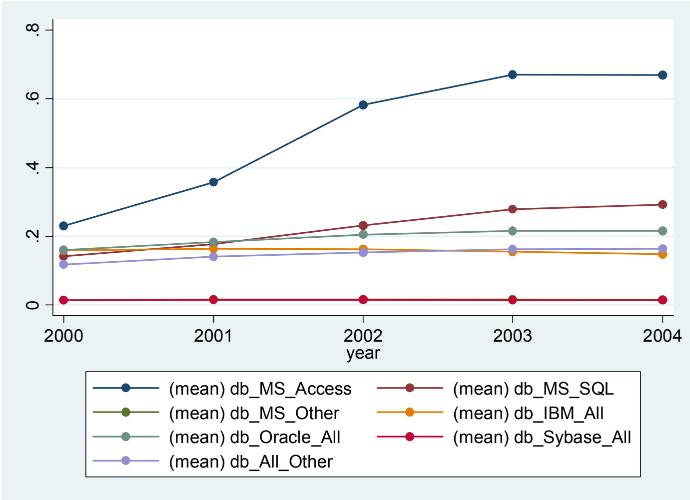


Figure 1b: Usage shares of major DBMS, early users.

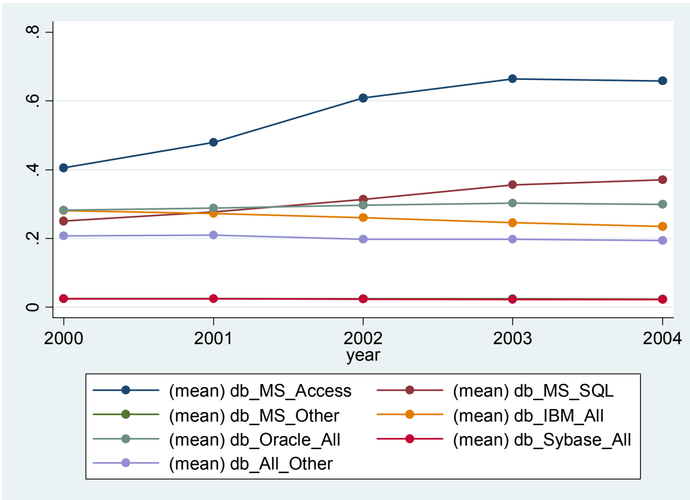
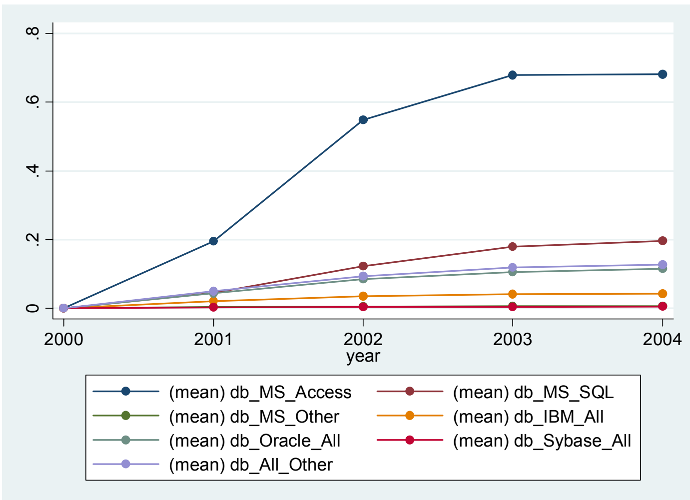


Figure 1c: Usage shares of major DBMS, late users.



It also becomes clear from Figure 1c that Microsoft Access has been most successful in attracting new users to their technology, with Microsoft SQL Server a distant second.

Table 1: Multiple DBMS Usage

Of Sites Using DBMS [%]	2000	2001	2002	2003	2004
Single Product	64%	59%	53%	48%	46%
Multiple Products	36%	41%	47%	52%	54%
Total	100%	100%	100%	100%	100%

We look at the patterns of individual sites in more detail in Table 1. We find that a considerable number of sites use multiple DBMS concurrently, and that this tendency is increasing over time. By 2004, more than half the sites in our sample use multiple DBMS.

Figure 2: Pairwise DBMS Combinations

Pair-wise use of DBMS [%][15]		MS Access	MS SQL Server	MS Other	IBM All	Oracle All	Sybase All
MS Access	**2000**	--	16.4%	1.8%	10.9%	13.8%	1.7%
	2004	--	26.9%	1.3%	13.5%	18.9%	1.3%
MS SQL Server	**2000**	--	--	1.7%	9.0%	13.6%	2.8%
	2004	--	--	1.5%	9.7%	20.8%	2.1%
MS Other	**2000**	--	--	--	0.6%	1.4%	1.1%
	2004	--	--	--	0.8%	1.3%	1.1%
IBM All	**2000**	--	--	--	--	7.6%	1.5%
	2004	--	--	--	--	8.3%	1.5%
Oracle All	**2000**	--	--	--	--	--	2.6%
	2004	--	--	--	--	--	2.4%

Table 2 gives pairwise combinations of DBMS in 2000 and 2004. We find that MS Access, MS SQL Server and Oracle are most commonly used in conjunction. That is, the likelihood that if one of the three is used, one of the other two will be used as well is highest (e.g. 26.9% for MS Access and MS SQL Server in 2004). Further, for these three DBMS, the tendency of simultaneous use has been increasing over time.

Our data suggests that different DBMS are not used exclusively at each site as commonly expected. This raises some interesting questions – are there specific pairs of DBMS that complement each other well? Are there firm characteristics that favour usage of particular DBMS?[16] What explains the development over time of these usage

[15] This has been constructed as $\frac{Joint\ use(A+B)}{[Joint\ use(A+B)+Single\ use(A)+Single\ use(B)]}$.

[16] Kretschmer (2004) analyzes the characteristics of firms using multiple operating systems and finds that task variety has an important effect on the propensity to use multiple operating systems, which

patterns? Is Access gaining on their rivals because they poach existing users, because they dominate the market for “new” consumers or due to their complementary offerings (e.g. DBMS for end-users versus system developers)?[17] Our three sets of regressions will attempt to answer some of these questions.

Usage Regressions

We first run logit regressions with a dummy variable on the LHS that equals one if the site uses the DBMS in question and zero if it does not. This simply attempts to uncover the circumstances that favour use of a particular DBMS. Tables 3a) to d) report our results for the four main DBMS groups in our sample – MS Access, MS SQL Server, IBM DBMS, and Oracle.[18] Column 1 pools all data and all years, column 2 reports the results of a panel regression, column 3-7 report results of yearly logit regressions, and column 8 reports results of the same covariates with a probit specification.

---

confirms the intuition given by Farrell and Saloner (1986). As DBMS are a specific application, we would expect that task variety plays a smaller role in sites’ usage decisions on DBMS.

[17] See Financial Times on how Microsoft is gearing its MS Office Pack towards BI and datamanagement (http://news.ft.com/cms/s/f7e3aa18-43ce-11da-b752-00000e2511c8.html).

[18] Results on the other DBMS are available from the author upon request.

Table 3a: MS Access Usage Regressions

Variable Name	(1) Coeff.	S.E.	(2) Coeff.	S.E.	(3) Coeff.	S.E.	(4) Coeff.	S.E.	(5) Coeff.	S.E.	(6) Coeff.	S.E.	(7) Coeff.	S.E.	(8) Coeff.	S.E.
ERP SAP	.198**	.051	.304**	.127	.108†	.147	.101†	.119	-.010†	.111	.045†	.114	.083†	.111	.122**	.032
ERP OTHER	.556**	.030	.695**	.071	.526**	.086	.521**	.070	.222**	.065	.173**	.066	.255**	.065	.346**	.018
OS WINDOWS	.556**	.154	.804**	.297	-.474*	.248	-.018†	.277	1.09**	.426	1.347**	.542	1.708**	.470	.340**	.093
OS OS/2	-.676**	.148	-.866**	.365	-.444†	.329	-.690**	.310	-.458**	.351	-.399†	.396	-.420†	.395	-.422**	.090
OS OS/400	-.114**	.031	-.131†	.084	-.112†	.077	-.240**	.070	-.142†	.069	-.045†	.074	.097†	.077	-.071**	.019
OS UNIX	-.067**	.026	-.037†	.067	.031†	.062	-.027†	.057	-.042†	.059	.079†	.063	.102†	.064	-.042**	.016
LOG(EMP)	-.105**	.016	-.167**	.042	-.140**	.039	-.067*	.036	-.075**	.038	.022†	.040	-.025†	.040	.065**	.010
LOG(SERVER)	.245**	.016	.341**	.039	.148**	.037	.156**	.033	.122**	.036	.084**	.040	.135*	.040	.151**	.010
LOG(NTWRK)	.156**	.020	.293**	.049	.198**	.047	.154**	.042	.145**	.046	.027†	.052	.045†	.052	.096**	.012
LOG(EFFINF)	-.173**	.020	-.266**	.050	-.099**	.047	-.142**	.043	-.171**	.045	-.137**	.050	-.107**	.050	-.107**	.012
LOG(EFFDEV)	-.094**	.019	-.220**	.050	-.091**	.047	-.109**	.041	-.048†	.044	-.062†	.048	-.132**	.049	-.057**	.012
CONSTANT	-.856**	.169	-1,833**	34.64	-.848**	.299	-.790**	.318	-.913**	.452	-.791†	.566	-1.128**	.502	-.526**	.102
Observations	28,873		28,873		6,163		6,189		5,679		5,358		5,484		28,873	
χ^2	1,077.49		3,980.19		111.83		171.66		91.91		59.89		98.73		1,114.99	
R^2	.029				.016		.022		.012		.009		.015		.029	

Notes:

- ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level.
- Regressions (2) – (8) use the same subset of variables. Estimation techniques and subsamples vary, as outlined below.

(1) Preferred regression: Logit regression on usage dummy usage of ERP SAP, other ERP, Windows OS, OS/2 OS, OS/400 OS, UNIX OS, log(employees), log(number of servers), log(network nodes), log(number of IT workers), and log(software developers).
(2) Preferred regression, random effects panel data logit regression.
(3) Preferred regression, year 2000 only.
(4) Preferred regression, year 2001 only.
(5) Preferred regression, year 2002 only.
(6) Preferred regression, year 2003 only.
(7) Preferred regression, year 2004 only.
(8) Preferred regression, probit specification.

Table 3b: MS SQL Server Usage Regressions

	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
Variable Name	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
ERP SAP	.150**	.055	.303**	.137	.112†	.162	.093†	.133	.144†	.119	.075†	.114	.052†	.109	.090**	.033
ERP OTHER	.416**	.032	.642**	.076	.464**	.099	.359**	.081	.239**	.070	.232**	.067	.357**	.064	.246**	.019
OS WINDOWS	.890**	.218	.918**	.372	.043†	.317	.617†	.374	.791†	.555	-.424†	.506	2.742**	.959	.508**	.118
OS OS/2	-.669**	.191	-.968**	.413	-.870†	.393	-.340*	.360	-.563†	.456	-.397**	.079	-.743†	.500	-.374**	.107
OS OS/400	-.292**	.036	-.485**	.093	-.156†	.087	-.223**	.081	-.292**	.079	-.392**	.067	-.344**	.078	-.172**	.021
OS UNIX	-.398**	.030	-.636**	.074	-.331†	.073	-.370†	.068	-.412**	.067	-.148**	.042	-.339**	.065	-.235**	.018
LOG(EMP)	-.174**	.019	-.222**	.047	-.177**	.046	-.142†	.043	-.168**	.042	.663**	.044	-.145**	.040	-.104**	.011
LOG(SERVER)	.599**	.019	1.075**	.045	.465**	.045	.448**	.042	.555**	.043	.152**	.054	.635**	.043	.355**	.011
LOG(NTWRK)	.228**	.023	.398**	.056	.290**	.055	.243**	.050	.211**	.052	-.039**	.051	.169**	.054	.138**	.014
LOG(EFFINF)	-.027†	.022	.001†	.056	.038**	.053	.036**	.049	.003†	.049	-.021†	.050	-.066†	.050	-.013†	.013
LOG(EFFDEV)	-.032†	.022	-.046†	.054	-.039*	.051	-.026**	.046	-.041†	.047	-1.668†	.191	-.012†	.049	-.018†	.013
CONSTANT	-2.987**	.232	-850.290**	35.845	-2.760**	.372	-3.019*	.416	-2.717**	.581	.075**	.114	-4.412**	.971	-1.781**	.127
Observations	28,873		28,873		6,163		6,189		5,679		5,341		5,484		28,873	
χ^2	2,867		2,646.27		458.07		459.18		491.14		553.09		559.41		3,112.93	
R^2	.097				.077		.073		.082		.095		.093		.097	

Notes:

- ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level.
- Regressions (2) – (8) use the same subset of variables. Estimation techniques and subsamples vary, as outlined below.

(1) Preferred regression: Logit regression on usage dummy usage of ERP SAP, other ERP, Windows OS, OS/2 OS, OS/400 OS, UNIX OS, log(employees), log(number of servers), log(network nodes), log(number of IT workers), and log(software developers).
(2) Preferred regression, random effects panel data logit regression.
(3) Preferred regression, year 2000 only.
(4) Preferred regression, year 2001 only.
(5) Preferred regression, year 2002 only.
(6) Preferred regression, year 2003 only.
(7) Preferred regression, year 2004 only.
(8) Preferred regression, probit specification.

Table 3c: IBM Server Usage Regressions

	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
Variable Name	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
ERP SAP	-.048†	.076	-.035†	.166	-.113†	.199	-.132†	.174	-.011†	.162	.460†	.636	-.058†	.162	-.025†	.041
ERP OTHER	.372**	.040	.497**	.089	.149†	.115	.367**	.094	.419**	.087	2.938**	.091	.400**	.085	.209**	.022
OS WINDOWS	.128†	.215	.404†	.420	-.092†	.363	-.057†	.391	.704†	.677	.601†	.094	.428†	.679	.065†	.117
OS OS/2	.780**	.186	1.152**	.436	.941**	.309	1.026**	.364	.813†	.477	.048†	.054	.158†	.621	.431**	.103
OS OS/400	2.974**	.040	5.801**	.104	3.059**	.086	3.020**	.086	2.996**	.089	-.026**	.054	2.863**	.089	1.751**	.022
OS UNIX	.670**	.040	.953**	.089	.732**	.088	.757**	.087	.712**	.091	-.086**	.068	.545**	.093	.374**	.021
LOG(EMP)	.0410	.022	.092*	.054	.088†	.047	.034†	.047	-.014†	.051	.004†	.069	.049†	.052	.022†	.012
LOG(SERVER)	-.009†	.022	-.048†	.050	-.008†	.049	-.022†	.050	-.016†	.052	.039†	.068	.023†	.053	-.001†	.012
LOG(NTWRK)	-.103**	.027	-.147*	.065	-.144**	.055	-.096†	.058	-.081†	.063	-3.436†	.828	-.106†	.066	-.056**	.015
LOG(EFFINF)	.010†	.029	-.026†	.065	.027†	.060	.004†	.060	.044†	.063	.460†	.636	-.043†	.068	.002†	.015
LOG(EFFDEV)	.045†	.028	.155*	.065	.025†	.060	.052†	.059	.053†	.063	2.938†	.091	.072†	.068	.030**	.015
CONSTANT	-2.440**	.232	2.064†	44.972	-2.362**	.405	-2.257**	.439	-2.905**	.706	.601**	.094	-2.693**	.712	-1.426**	.127
Observations	28,873		28,873		6,163		6,189		5,679		5,358		5,484		28,873	
χ^2	6,193.46		3,239.06		1,368.57		1,333.67		1,235.83		1,141.68		1,120.23		6,772.96	
R^2	.251				.263		.255		.254		.246		.238		.252	

Notes:

- ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level.
- Regressions (2) – (8) use the same subset of variables. Estimation techniques and subsamples vary, as outlined below.

(1) Preferred regression: Logit regression on usage dummy usage of ERP SAP, other ERP, Windows OS, OS/2 OS, OS/400 OS, UNIX OS, log(employees), log(number of servers), log(network nodes), log(number of IT workers), and log(software developers).
(2) Preferred regression, random effects panel data logit regression.
(3) Preferred regression, year 2000 only.
(4) Preferred regression, year 2001 only.
(5) Preferred regression, year 2002 only.
(6) Preferred regression, year 2003 only.
(7) Preferred regression, year 2004 only.
(8) Preferred regression, probit specification.

Table 3d: Oracle Server Usage Regressions

Variable Name	(1) Coeff.	S.E.	(2) Coeff.	S.E.	(3) Coeff.	S.E.	(4) Coeff.	S.E.	(5) Coeff.	S.E.	(6) Coeff.	S.E.	(7) Coeff.	S.E.	(8) Coeff.	S.E.
ERP SAP	.680**	.058	1.165**	.140	.701**	.153	.646**	.134	.587**	.127	.721**	.126	.159**	.073	.410**	.034
ERP OTHER	.163**	.036	.274**	.0818	.085†	.110	.134†	.088	.100†	.078	.142†	.075	-1.736*	.509	.093**	.021
OS WINDOWS	.430**	.179	1.107**	.358	-.340†	.261	.098†	.284	1.186†	.609	2.044*	1.035	-.637**	.092	.269**	.104
OS OS/2	-1.218**	.198	-1.639**	.426	-.702*	.354	-.858*	.366	-1.552**	.446	-1.824	.597	.647**	.069	-.702**	.111
OS OS/400	-.647**	.041	-1.171**	.103	-.597**	.095	-.711**	.093	-.636**	.091	-.670**	.094	.075**	.049	-.377**	.023
OS UNIX	.612**	.031	1.103**	.075	.683**	.070	.563**	.067	.568**	.069	.650**	.071	.343**	.046	.352**	.018
LOG(EMP)	-.010†	.022	.053†	.0494	-.040†	.049	-.037†	.048	-.043†	.049	.022†	.049	.280†	.061	-.004†	.012
LOG(SERVER)	.260**	.020	.558**	.044	.223**	.044	.168**	.042	.229**	.045	.277**	.048	.205**	.057	.155**	.011
LOG(NTWRK)	.389**	.027	.675**	.059	.356**	.058	.455**	.057	.454**	.060	.376**	.062	.031**	.053	.224**	.015
LOG(EFFINF)	.205**	.025	.467**	.059	.235**	.055	.193**	.052	.194**	.055	.223**	.058	-4.304**	.228	.121**	.014
LOG(EFFDEV)	.046**	.023	.022†	.057	.009†	.052	.068†	.049	.080†	.052	.060†	.054	.159†	.073	.030*	.014
CONSTANT	-4.721**	.203	.124**	.019	-3.744**	.328	-4.385**	.348	-5.519**	.647	-6.465**	1.062	-1.736**	.509	-2.793**	.117
Observations	28,873		28,873		6,163		6,189		5,679		5,358		5,457		28,873	
χ^2	4,602.00		2,750.66		811.57		877.32		928.30		971.09		967.63		5,076.59	
R^2	.177				.153		.161		.179		.196		.190		.178	

Notes:

- ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level.
- Regressions (2) – (8) use the same subset of variables. Estimation techniques and subsamples vary, as outlined below.

(1) Preferred regression: Logit regression on usage dummy usage of ERP SAP, other ERP, Windows OS, OS/2 OS, OS/400 OS, UNIX OS, log(employees), log(number of servers), log(network nodes), log(number of IT workers), and log(software developers).
(2) Preferred regression, random effects panel data logit regression.
(3) Preferred regression, year 2000 only.
(4) Preferred regression, year 2001 only.
(5) Preferred regression, year 2002 only.
(6) Preferred regression, year 2003 only.
(7) Preferred regression, year 2004 only.
(8) Preferred regression, probit specification.

A consistent result is that ERP and DBMS are positively correlated in their usage. This would suggest that they are indeed complements, so that data generated in one can be used in the other. The individual coefficients on ERP_SAP and ERP_OTHER suggest that Microsoft DBMS are particularly complementary to non-SAP ERP since the coefficient is positive in all specifications, while it is only sometimes significant and of smaller magnitude for SAP ERP. IBM follows a similar pattern – although the coefficient on SAP ERP is consistently negative (but insignificant). This suggests that IBM DBMS and SAP ERP are not complements, if anything. Use of Oracle, on the other hand, is positively correlated to use of SAP ERP in all specifications, which confirms the notion of Oracle's greatest strength of "unlimited scalability".[19]

The other potential internal complement effect is connected to the use of different operating systems as IBM and Microsoft are present in both markets and Oracle is unattached to any operating system. We find that internal complement effects indeed play a role, as IBM DBMS are more likely to be used in conjunction with OS/2 and OS/400, and Microsoft DBMS are more likely to be used with Windows. Interestingly, Oracle seems to have a degree of complementarity with both Windows and Unix, as shown by the coefficients on both these variables – conversely, the coefficients in the Oracle usage regressions for OS/2 and OS/400 are consistently negative. IBM DBMS also appear complementary to Unix, since the coefficient is consistently positive and significant in all regressions.

The number of servers (LOG_SERVER) and the number of network nodes (LOG_NTWRK) has a consistent (and mostly significant) positive effect on usage of Microsoft DBMS and Oracle – results for IBM DBMS are less conclusive, although the number of network nodes, LOG_NTWRK, has a negative and significant effect in most regressions. Sites with a high degree of interconnectedness are therefore less likely to use an IBM DBMS.

LOG_EFFINF and LOG_EFFDEV increase the likelihood of Oracle being used – the generality of Oracle suggests that more expertise is needed to customize it to a particular location and circumstances. On the other hand, high general IT expertise and programming capacities lead to a lower likelihood of MS Access being used. The

[19] SAP is most effective for large organizations with multiple sites of different sizes, which implies that questions of scalability across sites are particularly important in an SAP environment.

intuition here is that MS Access is a relatively generic and user-friendly program that can be used "off the shelf" or adjusted at relatively little cost.

Size has a negative effect on the use of Microsoft and a positive one on IBM, and has no effect on use of Oracle. This would be consistent with an interpretation that IBM systems are most attractive to larger organizations, while Microsoft DBMS appeal to smaller organizations, with Oracle being relatively general (again confirming Oracle's perceived strength of easy scalability).

Switching Regressions

For our switching regressions, we construct a dummy variable that takes the value 1 if a site starts using a particular DBMS in a particular year. We use a hazard rate specification and report results in Table 4.[20] Column 1 reports the results from our balanced panel, column 2 gives the results for the subset of firms that have been using a DBMS in the first year of our sample (2000), and column 3 only includes non-users in 2000. We find that our results are qualitatively similar, but we also highlight the differences in our discussion.

[20] Results from a standard random effects logit panel regression are available from the author upon request.

Table 4a: MS Access Switching Regressions

	Balanced Panel	Early User Panel	Late User Panel
ERP SAP	.933† (.091)	.937† (.114)	.962† (.158)
ERP OTHER	.617** (.038)	.607** (.050)	.639** (.0586)
OS WINDOWS	1.518† (.327)	1.048† (.399)	1.751* (.460)
OS OS/2	.948† (.220)	1.145† (.335)	.776† (.296)
OS OS/400	.882* (.046)	.853* (.058)	1.112† (.101)
OS UNIX	1.074† (.047)	1.086† (.068)	1.084† (.067)
LOG(EMP)	1.030† (.026)	1.036† (.036)	1.040† (.040)
LOG(SERVER)	.934* (.025)	.942† (.035)	.952† (.036)
LOG(NTWRK)	.974† (.030)	.958† (.042)	.976† (.041)
LOG(EFFINF)	.945† (.032)	.944† (.047)	.953† (.045)
LOG(EFFDEV)	1.076* (.036)	1.093† (.051)	1.044† (.051)
p	2.632** (.041)	2.643** (-.058)	2.631** (.057)
WALD χ^2	13,891.09	8,056.23	5,164.24
Observations	21,551	16,608	5502
Note: ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level. Standard errors in parentheses.			

Table 4b: MS SQL Server Switching Regressions

	Balanced Panel	Early User Panel	Late User
ERP SAP	.660** (.089)	.674* (.115)	.654* (.139)
ERP OTHER	.572** (.046)	.683** (.066)	.441** (.062)
OS WINDOWS	1.016† (.363)	1.268† (.737)	.909† (.410)
OS OS/2	1.075† (.364)	1.610† (.578)	.236† (.237)
OS OS/400	.913† (.071)	1.018† (.098)	1.060† (.146)
OS UNIX	1.171* (.074)	1.296** (.108)	1.202† (.114)
LOG(EMP)	1.105* (.043)	1.152** (.060)	1.055† (.061)
LOG(SERVER)	.878** (.033)	.835** (.042)	1.023† (.053)
LOG(NTWRK)	1.024† (.050)	.976† (.065)	1.140* (.074)
LOG(EFFINF)	1.080† (.052)	1.083† (.070)	1.048† (.071)
LOG(EFFDEV)	.990† (.046)	.996† (.061)	1.00** (.069)
p	1.860** (.045)	1.941** (.515)	1.760** (.065)
WALD χ^2	105.74	58.16	96.48
Observations	25,697	18,465	7,791
Note: ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level. Standard errors in parentheses.			

Table 4c: IBM Server Switching Regressions

	Balanced Panel	**Early User Panel**	**Late User**
ERP SAP	.547† (.229)	.533† (.317)	.591† (.350)
ERP OTHER	.454** (.103)	.269** (.116)	.662† (.179)
OS WINDOWS	.275* (.142)	.115** (.070)	.754† (.761)
OS OS/2	1.474† (1.070)	3.019† (2.253)	0.000† (.002)
OS OS/400	1.952* (.352)	1.513† (.442)	3.983** (.907)
OS UNIX	1.786** (.301)	1.888* (.511)	1.911** (.408)
LOG(EMP)	1.257* (.124)	1.461** (.231)	1.088† (.141)
LOG(SERVER)	.697** (.068)	.827† (.127)	.705** (.085)
LOG(NTWRK)	.740** (.087)	.696† (.134)	.819† (.117)
LOG(EFFINF)	1.118† (.146)	1.236† (.255)	1.082† (.176)
LOG(EFFDEV)	1.153† (.150)	.940† (.182)	1.277† (.216)
p	1.519** (.101)	1.661** (.602)	1.398** (.121)
WALD χ^2	80.23	40.76	62.60
Observations	28,366	19,984	8.981

Note: ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level. Standard errors in parentheses.

Table 4d: Oracle Server Switching Regressions

	Balanced Panel	Early User Panel	Late User
ERP SAP	.742† (.134)	.660† (.171)	.815† (.208)
ERP OTHER	.564** (.067)	.670** (.100)	.453** (.093)
OS WINDOWS	.657† (.272)	.388† (.197)	.902† (.525)
OS OS/2	.356† (.254)	.734† (.525)	0.000† (.001)
OS OS/400	.702** (.085)	.959† (.141)	.651† (.152)
OS UNIX	1.210* (.108)	1.225† (.154)	1.664** (.208)
LOG(EMP)	1.102† (.063)	1.187* (.094)	1.031† (.085)
LOG(SERVER)	.750** (.038)	.810** (.057)	.848* (.057)
LOG(NTWRK)	1.160** (.081)	.995† (.099)	1.339** (.121)
LOG(EFFINF)	1.390** (.094)	1.108† (.112)	1.672** (.141)
LOG(EFFDEV)	1.004† (.062)	1.150† (.107)	.925† (.072)
p	1.617** (.058)	1.780** (.087)	.384** (.052)
WALD χ^2	207.76	52.93	261.90
Observations	27,145	19,376	8,343
Note: ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level. Standard errors in parentheses.			

We first note that we find a consistent and significantly (in all but one regression) negative effect of ERP_OTH on the likelihood of switching, which seems surprising at first. Our usage regressions have shown that there appear to be significant complementarities between DBMS and ERP. We believe that this result is due to the fact that ERP are typically much wider in scope and more expensive to install and customize (the sheer scale of the SAP consulting industry should confirm this!). That is, once an ERP is in place and up and running, it is unlikely that a new DBMS will be purchased, which would trigger another round of adjustments and customizations. That is, the lumpy nature of ERP investment and the strong complementarities between ERP and DBMS imply that once a working combination is in place, the willingness to switch is low. In addition, the likelihood of adopting Microsoft SQL Server on top of SAP is significantly reduced as both applications are relatively specific in their programming environment, but are both designed to be relatively flexible, so that a site using SAP is unlikely to require another flexible DBMS.

Operating Systems have a relatively inconsistent effect on the propensity to switch DBMS. Windows and Access, both Microsoft's flagship products in their respective fields, appear to be complementary, while using Unix makes adoption of all three other DBMS more likely. OS/400, one of IBM's Operating Systems, positively affects the likelihood of a switch to IBM DBMS, and negatively affects the likelihood of switching to Oracle and (with limited support) MS Access, which again implies the existence of internal complement effects.

The degree of connectedness via servers has a consistent negative effect on switching DBMS and is significant in most regressions, whereas highly networked sites are more likely to switch to MS SQL Server or Oracle and less likely to switch to IBM, which is consistent with the usage regressions in the previous section.

Expertise in general IT or programming has, as in our usage regressions, a negative impact on the likelihood of switching to MS Access and a positive impact on the likelihood of switching to Oracle. It is insignificant for our other DBMS regressions.

We finally note some of the important differences in our results between the early and late samples. First, in the sample of late users, usage of Windows only has a positive effect on switching to Microsoft Access. While this is not unexpected, this also indicates that new adopters of a DBMS are likely to opt for Microsoft Access if they use Windows OS. As Windows has held a 90% market share for several decades, this implies that Microsoft's dominance is unlikely to be broken by new users of DBMS – this is confirmed by Table 1c), where we can see that late adopters have, in the time period of our sample, converged to the MS Access' overall usage share of about 60%. Second, the positive effect of size on switching to IBM and Oracle is not significant for the late users subsample. This would suggest that the advantages of these DBMS for larger sites vanish if sites are not using any DBMS before. Thus, a large site that is considering installing a DBMS from scratch is not significantly more likely to choose IBM or Oracle. This could be due to the aggressive pricing policies of Microsoft for larger clients. Finally, we note that the number of network nodes only has a significant positive effect on the likelihood of adoption of SQL Server and Oracle for late adopters. A "greenfield investment" by a highly networked company is thus more likely to fall to SQL Server and Oracle, while a "top-up investment" holds no particular advantage for these DBMS.

Competitive Switching Regression

Finally, we analyze a particular form of switching. Competitive switching is assumed to occur if a firm has been using any other DBMS and abandons these in order to use another one – our dependent variable is an appropriate dummy variable. Since this has been occurring mainly in the direction of Microsoft Access, we only report switching to Access.[21] We analyze competitive switching by running a Weibull hazard regression to identify the timing effects of switching (reported in Table 5).

Table 5: MS Access Competitive Switching Regressions

	Coeff.	**S.E.**
ERP SAP	1.031†	.622
ERP OTHER	.508†	.197
OS WINDOWS	890,964e+09†	1.34e+09
OS OS/2	0.000†	.003
OS OS/400	.873†	.314
OS UNIX	.886†	.275
LOG(EMP)	1.050†	.174
LOG(SERVER)	.712†	.130
LOG(NTWRK)	.642*	.129
LOG(EFFINF)	1.144†	.260
LOG(EFFDEV)	1.061†	.261
p	1.421	.166
Wald χ^2	30.50	
Observations	28,731	
Note: ** indicates 1% significance, * indicates 5% significance, † indicates significance at the 10% level.		

We find that most variables are insignificant, expect for use of a non-SAP ERP, the number of network nodes and the number of servers. All three variables have a significant and negative effect on the likelihood of switching to MS Access. On the one hand, this is due to the specific strengths of MS Access in smaller and less data-intensive organizations, but it also suggests that in the presence of a large networked organization and/or a large-scale enterprise system, a drastic change of a particular application becomes less likely. That is, the switching costs of a networked firm or an ERP-using firm are likely to include anticipated adjustment cost both in

[21] There have been 187 competitive switches to MS Access, 119 to MS SQL, 14 to IBM DBMS and 39 to Oracle.

communication software and in ERP software, which in turn will decrease the likelihood of switching.

INTERPRETATION AND CONCLUSION

The software industry has been a fruitful ground for the study of network technologies for reasons of data availability and the obvious potential for network effects. This paper is a first look at the competitive landscape of a particular software market – the market for Database Management Systems. Apart from capturing the dynamics of competition in this industry, this paper attempts to uncover several interesting features that have not previously been studied in network industries: First, we capture the fact that DBMS may not be pure substitutes in the sense that consumers have unit demand for any DBMS. Kretschmer (2004) finds that operating systems (OS) may not be used exclusively even on a single site, with different OS fulfilling different tasks. Similarly, we show that in the DBMS market there has been a tendency in the last years to use multiple programmes concurrently – interestingly, while simultaneous use of MS Access and MS SQL Server can be expected since they are written to be interoperable, Oracle has been another beneficiary of this tendency, since the sites in our sample are increasingly using Oracle in conjunction with Microsoft DBMS. Second, we take into account the effect of complementary products on site – that is, indirect network effects on the site level rather than economy-wide level, or internal complement effects. We find that there exist significant internal complement effects between operating systems and DBMS of the same vendor, and between Enterprise Resource Planning systems (ERP) and DBMS. Our regressions also confirm that investment into ERP are typically a more lumpy and long-term investment that seems to guide the use of DBMS in the future – in other words, once an ERP-DBMS system has been set up and fine-tuned, the constellation is unlikely to be changed by adopting a new DBMS. The notion of lumpy investments and the precarious balance of interdependent computer environments is also confirmed in the most drastic form of adoption – abandoning an old system while adopting a new one. We find that current use of an ERP and a widely linked network decrease the likelihood of switching to the dominant supplier, Microsoft.

As mentioned, this paper is only a first step toward a deeper understanding of issues of *multiple usage* of (allegedly) competing, but differentiated products and the existence and strength of *internal complement effects*. More work is needed to uncover the precise nature of complementarities, and the cross-effects of switching of one product on the propensity to switch another complementary product too. Breuhan (1997) has shown that the likelihood of switching to a competing vendor increases if a new generation of the incumbent product is introduced. Along similar lines, it would be interesting to see if the likelihood of switching is increased if a new generation of a complementary product is i) introduced in the market, and ii) adopted by the organization. Further, technological and competitive aspects of the products we study have not been utilized in detail. While we have some information on the general strengths and weaknesses of the DBMS we study, it would be interesting to go to the level of product features and their impact on usage and adoption behaviour, as well as some information on dynamic pricing strategies and their impact. Finally, we have not considered other firm characteristics, such as location, industry, performance, asset stocks, etc. In future work, we expect to link these characteristics in to our current dataset.

BIBLIOGRAPHY

Arthur, W.B. (1989). Competing Technologies, Increasing Returns, and Lock-in by Historical Events. *Economic Journal* 99, 116-139.

Åstebro, T. (2004). Sunk Costs and the Depth and Probability of Technology Adoption. *Journal of Industrial Economics* 52, 381-399.

Bayus, B. and V. Shankar (2003). Network Effects and Competition: An Empirical Analysis of the Home Video Game Industry. *Strategic Management Journal* 24, 375-384.

Breuhan, A. (1997) Innovation and the Persistence of Technological Lock-in. Ph.D. dissertation, Department of Economics, Stanford University.

Brynjolfsson, E. and C. Kemerer (1996). Network Externalities in Microcomputer Software: An Econometric Analysis of the Spreadsheet Market. *Management Science* 42, 1627-1647.

Cabral, L. and T. Kretschmer (forthcoming), Standards Battles and Public Policy. In: S. Greenstein, V. Stango (Eds.), *Standards and Public Policy*, forthcoming.

Cohen, W. And Levinthal, D. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35, 128-152.

Farrell, J., G. Saloner (1986): Standardization and Variety. Economics Letters 20, pp. 71-74.

Gandal, N. (1994). Hedonic price indexes for spreadsheets and an empirical test for network externalities. *Rand Journal of Economics* 25, 160-170.

Gandal, N., Greenstein, S. and Salant, D. (1999). Adoptions and Orphans in the Early Microcomputer Market. *Journal of Industrial Economics* 47, 87-105.

Gandal, N., M. Kende, and R. Rob (2000). The Dynamics of Technological Adoption in Hardware/Software Systems: The case of Compact Disk Players. *Rand Journal of Economics* 31, 43-61.

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

Koski, H. and T. Kretschmer (2005). Entry, Standards and Competition: Firm Strategies and the Diffusion of Mobile Telephony. *Review of Industrial Organization*. 26, 89-113.

Kretschmer, T. (2004). Upgrading and Niche Usage of PC Operating Systems. *International Journal of Industrial Organization* 22, pp. 1155-1182.

Kretschmer, T. (2005). Competition, Inertia, and Network Effects. Mimeo, *LSE*.

Lerner, J. and Tirole, J. (2002). Some Simple Economics of Open Source. *Journal of Industrial Economics* 2, 197-234.

Liebowitz, S. and S. Margolis (2001). *Winners, Losers, & Microsoft: Competition and Antitrust in High Technology*. Oakland, CA: Independent Institute.

Matusik, S, Heeley, M. (2005). Absorptive Capacity in the Software Industry: Identifying Dimensions That Affect Knowledge and Knowledge Creation Activities. *Journal of Management* 31. 549-572.

Ohashi, H. (2003). The Role of Network Effects in the US VCR Market, 1978–1986. *Journal of Economics and Management Strategy* 12, 447-494.

Rogers, E. (2003). *Diffusion of Innovations* (3rd Ed.). New York, Simon & Schuster.

Rohlfs, J. (2001). *Bandwagon Effects in High-Technology Industries*. MIT Press, Cambridge.

Sarnikar, S. (2002). The Effect of Java on Network Benefits in Computer Software Markets. Mimeo, *University of Arizona*.

APPENDIX

Variable Definitions

Variable	Definition
MS Access	Dummy equals 1 if site uses Microsoft Access; else 0.
MS SQL	Dummy equals 1 if site uses Microsoft SQL Server; else 0.
MS Other	Dummy equals 1 if site uses other Microsoft DBMS; else 0.
IBM All	Dummy equals 1 if site uses any IBM DBMS (incl. Informix and Lotus applications); else 0.
Oracle All	Dummy equals 1 if site uses any Oracle DBMS; else 0.
Sybase All	Dummy equals 1 if site uses any Sybase DBMS; else 0.
All Other	Dummy equals 1 if site uses a DBMS from vendor other then IBM, Microsoft, Oracle or Sybase; else 0.
ERP SAP	Dummy equals 1 if site uses ERP software from SAP; else 0.
ERP OTHER	Dummy equals 1 if site uses ERP software from vendor other then SAP; else 0.
OS WINDOWS	Dummy equals 1 if site uses Windows Operating System; else 0.
OS OS/2	Dummy equals 1 if site uses IBM OS/2 Operating System; else 0.
OS OS/400	Dummy eqals 1 if site uses IBM OS/400 Operating System; else 0.
OS UNIX	Dummy equals 1 if site uses any Unix Operating System; else 0.
LOG(EMP)	(log) Total number of employees on sites.
LOG(SERVER)	(log) Total number of servers on sites.
LOG(NTWRK)	(log) Total number of sites connected via Wide Area Network.
LOG(EFFINF)	(log) Total number of IT employees on sites.
LOG(EFFDEV)	(log) Total number of IT development employees on sites.

Summary statistics for standard panel

Variable	Obs	Mean	Std. Dev.	Min	Max
MS Access	73,911	.487	.500	0	1
MS SQL	73,911	.219	.414	0	1
MS Other	73,911	.0147	.120	0	1
IBM All	73,911	.158	.364	0	1
Oracle All	73,911	.194	.395	0	1
Sybase All	73,911	.0137	.116	0	1
All Other	73,911	.146	.353	0	1
ERP SAP	73,911	.059	.235	0	1
ERP OTHER	73,911	.195	.397	0	1
OS WINDOWS	73,911	.988	.111	0	1
OS OS/2	73,911	.006	.079	0	1
OS OS/400	73,911	.157	.363	0	1
OS UNIX	73,911	.311	.463	0	1
LOG(EMP)	76,889	318.754	824.030	0	150,002
LOG(SERVER)	69,067	9.407	32.905	1	3,572
LOG(NTWRK)	76,850	261.998	1014.977	0	120,155
LOG(EFFINF)	76,794	12.426	52.377	0	2,500
LOG(EFFDEV)	76,720	3.702	19.652	0	2,000

Summary statistics for balanced panel

Variable	Obs	Mean	Std. Dev.	Min	Max
MS Access	71,289	.484	.450	0	1
MS SQL	71,289	.220	.414	0	1
MS Other	71,289	.015	.122	0	1
IBM All	71,289	.161	.368	0	1
Oracle All	71,289	.196	.397	0	1
Sybase All	71,289	.014	.118	0	1
All Other	71,289	.145	.352	0	1
ERP SAP	71,289	.059	.235	0	1
ERP OTHER	71,289	.195	.396	0	1
OS WINDOWS	71,289	.989	.105	0	1
OS OS/2	71,289	.006	.079	0	1
OS OS/400	71,289	.159	.366	0	1
OS UNIX	71,289	.316	.465	0	1
LOG(EMP)	71,289	317.797	834.935	0	150,002
LOG(SERVER)	66,826	9.423	33.177	1	3572
LOG(NTWRK)	71,282	265.766	1,015.808	0	120,155
LOG(EFFINF)	71,194	12.585	52.907	0	2,500
LOG(EFFDEV)	71,120	3.742	19.696	0	2,000

Summary statistics for early users

Variable	Obs	Mean	Std. Dev.	Min	Max
MS Access	41,478	.554	.497	0	1
MS SQL	41,478	.309	.462	0	1
MS Other	41,478	.024	.152	0	1
IBM All	41,478	.260	.439	0	1
Oracle All	41,478	.293	.455	0	1
Sybase All	41,478	.022	.148	0	1
All Other	41,478	.201	.401	0	1
ERP SAP	41,478	.067	.250	0	1
ERP OTHER	41,478	.216	.412	0	1
OS WINDOWS	41,478	.991	.098	0	1
OS OS/2	41,478	.007	.082	0	1
OS OS/400	41,478	.212	.409	0	1
OS UNIX	41,478	.355	.479	0	1
LOG(EMP)	41,478	381.874	1,023.028	0	150,002
LOG(SERVER)	39,576	11.689	37.774	1	3,572
LOG(NTWRK)	41,476	326.201	759.895	0	42,000
LOG(EFFINF)	41,449	15.715	54.369	0	2500
LOG(EFFDEV)	41,449	4.893	22.522	0	2000

Summary statistics for late users

Variable	Obs	Mean	Std. Dev.	Min	Max
MS Access	32,433	.402	.490	0	1
MS SQL	32,433	.103	.305	0	1
MS Other	32,433	.003	.058	0	1
IBM All	32,433	.026	.160	0	1
Oracle All	32,433	.067	.249	0	1
Sybase All	32,433	.002	.050	0	1
All Other	32,433	.074	.262	0	1
ERP SAP	32,433	.048	.213	0	1
ERP OTHER	32,433	.169	.374	0	1
OS WINDOWS	32,433	.984	.126	0	1
OS OS/2	32,433	.006	.076	0	1
OS OS/400	32,433	.086	.281	0	1
OS UNIX	32,433	.257	.437	0	1
LOG(EMP)	35,411	244.820	488.236	0	15,000
LOG(SERVER)	29,491	6.345	24.590	1	1,800
LOG(NTWRK)	35,374	186.720	1,245.209	0	120,155
LOG(EFFINF)	35,345	8.569	49.664	0	2,000
LOG(EFFDEV)	35,345	2.304	15.513	0	600