JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **26**: 122–150 (2011)
Published online 30 September 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1207

AN EMPIRICAL MODEL OF MAINFRAME COMPUTER INVESTMENT

SUNG-JIN CHO*
Department of Economics, Seoul National University, Seoul, South Korea

SUMMARY
This paper introduces a dynamic model of investment decisions in mainframe computer systems. I estimate and test the model using detailed micro data from a company in the telecommunications industry. The model accounts for 'technological depreciation' which distinguishes computers from other investment goods where physical depreciation is typically a key factor underlying replacement investment decisions. The company increased its installed mainframe computer capacity by over 30-fold over the 10-year sample period. Part of this growth was undoubtedly due to the huge increase in performance and the corresponding drop in the per unit cost of processing capacity of mainframes, a consequence of 'Moore's law'. However, there was also tremendous growth in the need for computers for billing, account processing and other tasks, due to the rapid growth in the telecommunications industry over this same time period. I estimate the unknown parameters of the investment model using a nonlinear least squares–nested fixed-point algorithm (NLS-NFXP), which solves the Bellman equation underlying the dynamic model of investment and replacement of mainframe computers by nonlinear least squares. I demonstrate that it is feasible to estimate this model on an ordinary PC, whereas standard discretization approaches to solving the firm's optimal investment policy might not even be feasible using supercomputers. I show that the estimated model fits the data very well, and accurately captures the large growth in installed mainframe capacity, the timing and magnitude of replacement investment, as well as periodic upgrades of existing mainframe units. I use the model to decompose how much of the 30-fold increase in mainframe computer capacity is due to 'Moore's law' (i.e. the huge drop in the unit cost of installed mainframe capacity), and how much is due to the growth in demand for services of mainframes, due to the rapid growth in demand for telecommunications services (particularly cell phone accounts) by the firm's customers.

1. INTRODUCTION

Despite the importance of computers in the 'information economy', relatively little is known about the factors influencing mainframe investment, particularly with regard to the timing of upgrade and replacement decisions. With the environment of rapid technological progress and steadily declining costs, consumers and firms must decide whether they should upgrade or replace their existing computer systems now, or wait and purchase a faster/cheaper system in the future.

This paper presents a dynamic programming model of how a firm chooses between keeping, upgrading or replacing its legacy computer systems in the face of uncertainty regarding the demand for its services and the timing and extent of future cost reductions of new computer systems. I estimate the model using a detailed set of data on the computer holdings of one of the world's largest telecommunications companies.

An initial analysis of the data leads to several conclusions. First, the average interval between successive upgrades or replacements has become shorter over the last 10 years, perhaps reflecting the increased rate of technological progress in computing equipment over the same period. Second, although the average working life of computers was only around 6 years at the beginning of the

* Correspondence to: Sung-Jin Cho, Department of Economics, Seoul National University, Gwanak 599, Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea. E-mail: sungcho@snu.ac.kr

sample period, it decreased to approximately 5 years by the end of the period. Third, I show that as increased demand for services served by computers begins to exceed their processing capacity, the firm is more likely to expand its capacity via an upgrade of existing computers rather than outright purchases when the existing computers are relatively new. However, as the age of computers approaches the length of the replacement cycle, the firm is more likely to replace them. Taken together, these facts support the notion that rapid technological progress and economic development affect a firm's replacement and upgrade policy.

Analysis of data on the mainframe investment for a major telecommunications firm suggests that one of the major forces driving computer systems replacement decisions is technological depreciation. Figure 1 shows the average non-capacity-adjusted prices paid by the firm for mainframe computers, which can be thought of as nominal prices in terms of computer capacity. In contrast, Figure 2 shows the average capacity-adjusted prices paid by the firm for mainframe computers. According to Figure 1, the overall price paid for mainframe computers does not decrease over this time period. However, the capacity-adjusted price paid for mainframes decreases sharply and continuously (Figure 2). Based on this observation, we can conjecture that although the price of mainframes did not decrease much over the period, their capacity increased by large amounts, leading to a large decrease in the cost per unit of capacity.

This phenomenon can be attributed to Moore's law. Moore's law states that the processing capacity of central processing units (CPUs) doubles every 18 months. In the storage industry density has been doubling every 12 months, an even faster rate than that of CPU development. The period of my data starts in 1989 and ends in 1999, during which time processing capacity improved from 1 M transistors per CPU (486 DX processor) to over 24 M transistors per CPU (Pentium III processor). It should be noted that there was a tremendous improvement in computer technology between 1984 and 1997, and that the experts in computer technology expected significantly more rapid development between 1997 and 2009.

The other main force behind the large increase in the firm's mainframe capacity over this period is the massive growth in the demand for its mainframe-based services, namely account processing, billing and so forth. The growth in demand is another outgrowth of Moore's law since the miniaturization of circuitry and electronics led to an increase in the demand for cell phones.

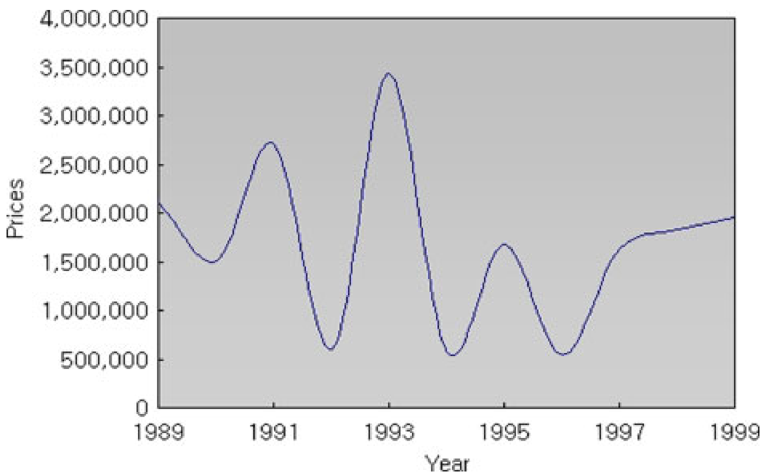

Figure 1. Average purchasing prices of mainframes (non-capacity-adjusted). This figure is available in color online at wileyonlinelibrary.com/journal/jae

DOI: 10.1002/jae

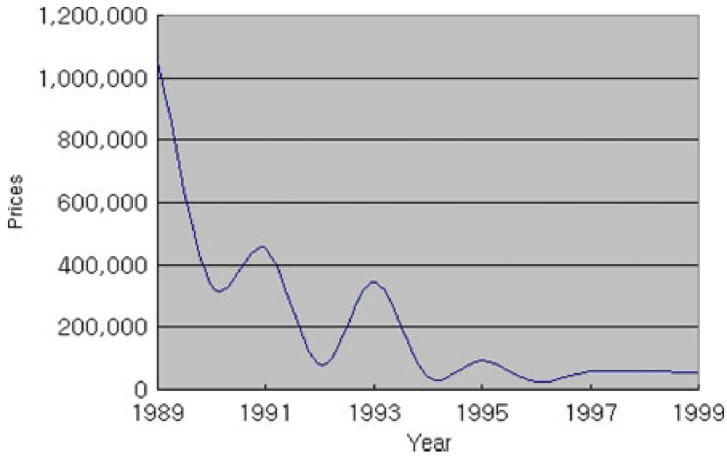

Figure 2. Average purchasing prices of mainframes (capacity adjusted). This figure is available in color online at wileyonlinelibrary.com/journal/jae

The proliferation of cell phone accounts is one of the primary factors underlying the increased demand for the firm's mainframe services over this period.

This paper offers two unique contributions in terms of methodology. First, it introduces a stochastic optimal investment model that explains replacement and upgrade actions and applies it to technologically deteriorating systems. Heretofore, there has not been a model capable of explaining firms' investment behavior regarding the replacement and upgrade of technologically depreciating systems. Based on the unique features of computer systems, especially technological depreciation, I develop a stochastic dynamic programing (DP) model to determine whether the facts about investment behavior resulting from this empirical analysis can be rationalized as an optimal investment strategy for the firm.

Second, this is the first study to apply a nonlinear least squares–nested fixed-point algorithm (NLS-NFXP), which is a combination of the nested fixed-point algorithm and the parametric approximation method to a high-dimensional fixed-point problem as an estimation technique. The parametric approximation method, which is used to solve the DP problem, greatly reduces the computational burden of solving the infinite-horizon version in which decisions are taken at monthly intervals. This method also allows the two key state variables–current demand and the price per unit of new capacity–to assume a continuum of possible values. It also converts the contraction fixed-point problem into a nonlinear least squares problem. The acceleration in solution time, compared with standard methods by the discretization of state space, is sufficiently large to make it feasible to estimate the unknown parameters of the model based on maximum likelihood.

This paper also makes unique empirical contributions. I use a new and unique dataset of computer investments by one of the world's largest telecommunications companies. The data provide evidence that the firm's replacement and upgrade behaviors fully reflect the technological depreciation of its computers.

Using the estimated model, I am able to predict how investment in installed mainframe capacity would have changed in the absence of Moore's law. The model predicts that, if the cost and quality of mainframe computers were held fixed at their 1989 values, installed mainframe capacity would have increased by approximately 11- to 13-fold (depending on the type of system analyzed), in response to the sevenfold increase in demand over the decade of the sample. On the other hand, if we hold the demand for computing services fixed at their 1989 values, the model predicts that a

DOI: 10.1002/jae

large drop in per unit costs of mainframe computers (a result of Moore's law) would have resulted in a 17- to 21-fold increase in installed mainframe capacity. When we account for both cost and demand factors simultaneously, the model accurately predicts the huge 30- to 33-fold increase in mainframe computer capacity over this decade.

The paper is organized as follows: In Section 2 I present a summary of related research. Section 3 explains the data and explanatory facts of the data. In Section 4 I present the formal analysis, developing a stochastic dynamic programming model to determine whether the stylized facts regarding replacement and upgrade behavior can be rationalized as an optimal investment strategy for this firm. In Section 5 I estimate the model using NLS-NFXP. I then conduct several simulations to illustrate the estimation results. Section 6 provides some concluding remarks.

2. SUMMARY OF RELATED RESEARCH

Previous research on systems replacement has focused on the replacement of bus engines (Rust, 1987) and aircraft engines (Kennet, 1994). Rust's (1987) seminal work on systems replacement provides a general template for approaching this topic. He does this by formulating a regenerative optimal stopping model for bus engine replacement to describe the behavior of the superintendent of maintenance at the Madison Metropolitan Bus Company. The proposed optimal stopping rule is a strategy for deciding when to replace current bus engines, and is given as a function of observed and unobserved state variables. Rust (1987) provides a general framework that can be used to analyze replacement behavior in various fields. In fact, it is the first research that uses a 'bottom-up approach' for modeling replacement investment. Second, the paper develops a nested fixed-point algorithm for estimating dynamic programming models of discrete choices. It should be emphasized, however, that computers differ from those regarding engines in several respects. First, the major forces that drove large increases in mainframe capacity over this period are completely different from those for bus and aircraft engines. While bus and aircraft engines are replaced because of physical depreciation, natural wear and tear or mechanical failure, computer systems are usually replaced as a result of technological depreciation. Research on engine replacement demonstrates that the primary reason for replacing engines is to prevent future failure, with capacity improvement only a secondary reason. As a result, the state variables in research on engine replacement are the hours of operation and the history of engine shutdowns, which are both measurements of physical depreciation. In contrast, though the prevention of future failure can be a reason to replace or upgrade computer systems, the main reasons would be to improve performance and meet the service demands. In case of computer systems, replacement resulting from physical depreciation accounts for a relatively small fraction of replacements. Thus the physical depreciation of state variables mentioned above for engines may not be appropriate in a model of computer systems replacement.

Second, upgrading is an alternative to replacement when attempting to improve computer performance. In the case of bus or aircraft engines, there is no upgrade choice.[1] For computer systems, upgrading is sometimes a better choice than replacement. Therefore, the replacing of computer systems requires a more complicated decision process than replacing engines. With a computer system, the possible decisions are whether to replace it, upgrade it, buy an additional system, or keep the current system as it is. These are the main choices. Based on these main choices, there is a separate set of sub-choices. For example, a replacement decision requires

[1] Some people might consider engine maintenance for better performance to be an upgrade choice. However, I assume that engine maintenance is a 'do nothing' form of behavior since, due to the nature of engines, it is difficult to achieve improved engine performance without replacing it.

DOI: 10.1002/jae

subsequent choices regarding the capacity and/or brand of the new computer system. This system of decision making can be described using a multiple discrete-choice model.

There are several potentially important factors that are unique to the replacement decisions regarding mainframe computer systems, including the introduction of new software,[2] system services provided by vendors,[3] and fleet replacement.[4] Despite the aforementioned differences between replacement of engines and computers, the algorithm developed by Rust is very useful in solving problems that typically arise in investigating replacement behavior, and the results of his paper have been widely applied since its publication.

Despite its significant role in replacement research, Rust's model was not intended for computer systems. There have been, however, several articles related to investment in computer systems, namely Hendel (1999) and Greenstein and Wade (1997). Hendel presents a multiple–discrete-choice model for the analysis of differentiated, durable goods undergoing a continuous process of technological change. Hendel develops a model of PC purchasing behavior to deal with the multiple–discreteness of PC demand; firms spreading their purchases over various brands of computers through block purchases. Hendel's model, combined with his new data on PC holdings, permits demand estimation at the micro level. The model is very useful in explaining the optimal replacement behavior with regard to PCs, since block purchasing is one of the most important features of PC replacement.

Greenstein and Wade (1997) investigate the product life cycle in the commercial mainframe market. In particular, they examine the entry and exit behavior of mainframe computers in the market using the hazard and Poisson models. The hazard model is used to estimate the probability of product exit, while the Poisson model is used to estimate the probability of product introduction. Additionally, this paper indicates many important market structures which may cause the entry or exit of products, including cannibalization, vintage and degree of competitiveness. A series of articles by Bresnahan and Greenstein (1997 and 1999) investigate the structural changes in the mainframe computer market with regard to technological changes.

The literature regarding investment in computers does not deal with the demand-side factors influencing the replacement of computers. Greenstein and Wade (1997) and Bresnahan and Greenstein (1999) focus on the supply-side behavior of computer markets, while this study of the replacement of computer systems is based on demand-side choices. Hendel focuses on demand-side choices in purchasing decisions regarding personal computers, where brand choice is an important decision. My data on mainframe computer investment decisions, including both replacement and upgrade decisions, suggest that when firms replace their current mainframe computer systems they tend to keep the same brand, since that makes it easier to obtain service. Thus my model disregards brand choice.[5] Like Hendel, we still form the model implicitly as a multiple–discrete-choice model. Actual replacement decisions are based on the current stock of computer systems. There are various tasks performed by the firm, and each task requires a fixed number of mainframe computer systems. Thus an aggregation of tasks of the firm and replacement choices over these

[2] The introduction of new software, such as a new operating system (OS), is one of the main reasons to replace computers, since a new OS may require a more advanced system or larger capacity to work properly. For example, each new OS has a minimum requirement for the computer's hardware specifications and this minimum requirement tends to increase over time, as newer operating systems become available.

[3] Unlike engine replacements, service support by vendors may play an important role in computer replacement decisions. Since vendors usually do not support old systems without an additional service contract, maintaining old computer systems may be more costly than replacing them with new computers

[4] We may examine to what extent mainframe replacement behavior is carried out individually versus on a 'fleet replacement' basis due to the costs of training administrative staff. In many cases, block purchases of new computers can benefit the firm in the form of a quantity discount. These features can be considered in a future extension as richer data become available.

[5] This assumption can be relaxed in future research.

aggregated tasks can be viewed as choice of the number of computers. The simultaneous choice for replacement timing is accompanied by a choice regarding the aggregated tasks.

3. THE DATA

3.1. Summary of the Data

I have obtained data from one of the largest telecommunications companies in the world. The company handles over 60% of all telephone services in its particular market. It also offers several other telecommunication services, including cellular PCS (Personal Communications Service), Internet, cable and satellite communication services. As of 1998, the company had 864 hosts (including workstations) and about 39,000 PCs. These hosts and PCs are spread out across 400 regional headquarters and regional offices. All regional headquarters operate independently and own their computer systems, and there is some difference in terms of computer capacity across these offices. In most cases, each regional headquarters independently makes decisions about the maintenance of and investment in its mainframe computers.

The computer systems in this company can be divided into two categories according to their use: (i) research and (ii) management and/or delivering services. Since computers used for research use are purchased and replaced on a project basis, their maintenance activities do not reflect technological depreciation.[6] In this paper, I consider only those computer systems used for management or delivering services. I also do not include the replacement of PCs in my model, since in PC replacement there is no upgrade activity, but rather only block purchases or replacements.

The time frame of the dataset starts in 1989 and ends in 1999. The data prior to 1989 are incomplete, though some computer systems have a history starting as early as 1977. For the 1989–1999 period, I have the full history of upgrades and replacements for 105 of the company's computer systems. The data consist of purchase dates, purchase prices, specifications, upgrade and replacement dates, upgrade and replacement prices, and other details on each replacement or upgrade such as system specifications. Monthly data on the firm's number of customers are also available.

Price data for CPU, hard drive, memory and other hardware were obtained from several computer databooks,[7] online computer resources[8] and manufacturers' web sites.[9]

3.2. Exploratory Investigation of the Data

I divided all computer systems in the sample into two categories based on their CPU benchmark standards, which are Transaction Processing Council (TPC) standards and Million Instructions per Second (MIPS) standards.[10] Different mainframe computers in the company report only one of the two standards. Since my dataset consists of various computer systems and dates, I divide computer systems into different task groups within each standard for the following two reasons. First, if the company previous had a MIPS type computer to do a certain task, when a replacement decision was made, it is assumed to always replace a MIPS computer with a new computer using the MIPS standard. Thus, once a certain system's brand is designated as serving a given task,

[6] Mainframe computers used for research use have a finite horizon with regard to research projects. My model assumes an infinite horizon.
[7] SIA annual data books.
[8] CNET.com, ZDnet.com, PC World, etc.
[9] Intel, AMD, MIPS, TPC, SUN, Motorola, Honeywell, Fujitsu, Unisys, Tandem, Samsung, Micron, Seagate, IBM, etc.
[10] Currently, the MIPS standard is in the process of being merged into the TPC standard, which includes transactions per minute (TPM) and transactions per second (TPS).

DOI: 10.1002/jae

later replacements are from the same, or similar, system brand. Second, it is very difficult to convert the MIPS standard into the TPC standard. The total sample consists of 48 MIPS and 57 TPC standard mainframe computer systems. All mainframe computer systems are associated with specific tasks. For example, MIPS standard mainframe computers perform the following tasks: billing development, billing management, general management, new customer info-system and super high-speed printing tasks. TPC standard computer systems perform business information management, customer development, total document, pre-billing, line management and material information tasks.

Table I illustrates the average, minimum and maximum costs of three activities–new purchase, upgrade or replacement–in terms of the two CPU standards. As expected for both standards, the cost of upgrade is less than the cost of a new purchase or replacement. According to computer industry databooks, the cost per unit capacity decreases over time. If, for example, the cost in the base year of 1982 was 100, then it was only 1 in 1998. Based on this information, the firm has increased the capacity of its computer systems tremendously, since the average price of replacement is the same or higher than the average costs of outright new purchases. This phenomenon can also be confirmed in the several computer industry databooks, which show that the cost of high-end computer systems, such as mainframe computers, has not decreased, and that they have in fact slightly increased over time. Figure 3 illustrates upgrade and replacement schedules associated with several of the important tasks mentioned above.

The other striking fact from Table I, Figure 3 and my available data is that the firm restricted itself to replacing mainframes with new, higher-capacity mainframes, rather than choosing to buy multiple higher-powered workstations.[11] This phenomenon seems to be a puzzle, since other firms in the IT industry were substituting away from mainframes to workstations over this period. One explanation for this empirical puzzle is that there is some degree of hardware/software complementarity. In other words, the firm has little choice but to replace mainframes with mainframes because of the large personnel switching costs involved in retraining the company employees to use other software that could more easily integrate different types of computer hardware platforms.[12]

The intervals between replacements are generally much longer than those between upgrades. The average intervals between replacements are 49.8 months and 55.7 months for the TPC standard and the MIPS standard, respectively. The maximum intervals between replacements are 53 months

Table I. Costs for three activities in the sample

Activity	Cost	MIPS	TPC
New purchase	Average	$572,919.9	$968,191.1
	Min.	$41,917.7	$20,440.1
	Max.	$4,893,545.6	$4,633,600.4
Replacement	Average	$1,082,499.4	$899,340.8
	Min.	$16,752.8	$14,854.3
	Max.	$7,160,791.3	$3,377,322.2
Upgrade	Average	$263,123.9	$435,181.4
	Min.	$2,645.12	$3,251.5
	Max.	$3,176,710.1	$2,283,130.1

Note: The reason for the big differences between minimum and maximum costs in the various activities is that the data consist of a wide variety of computer systems, from those with relatively little capacity to larger-capacity mainframe computers. This variety results in a significant gap between the two costs.

[11] This fact is embedded in the model.
[12] I will put aside this puzzle for future research.

DOI: 10.1002/jae

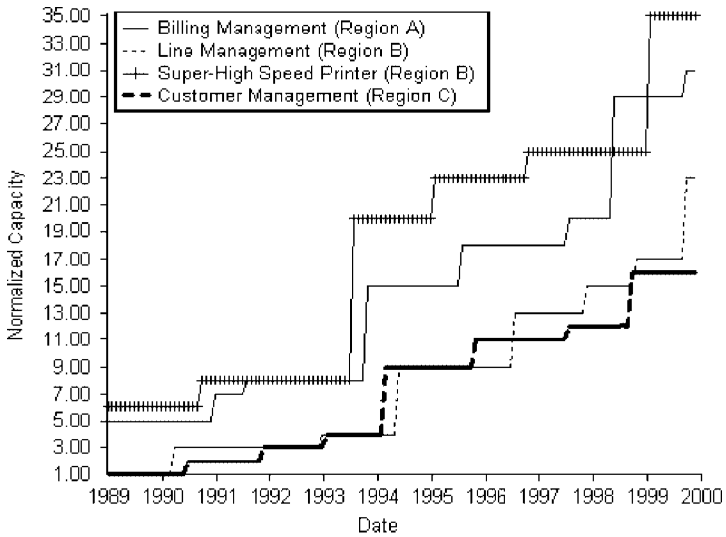

Figure 3. Upgrade and replacement schedule for computers associated with certain important tasks

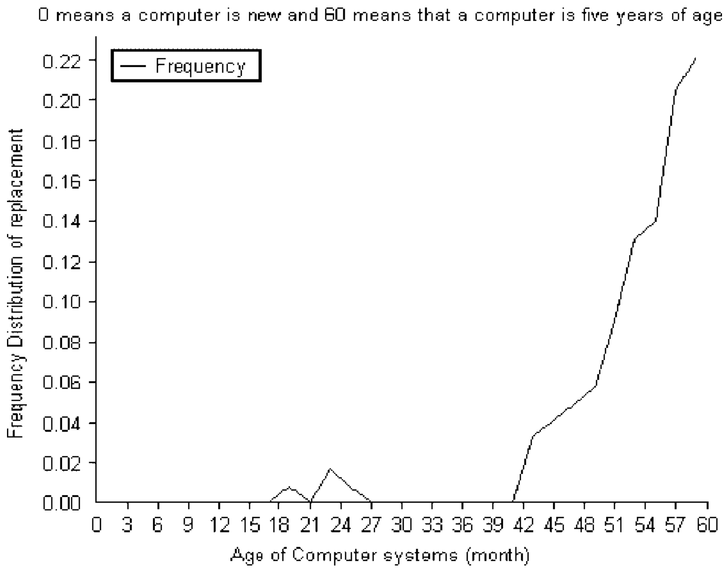

Figure 4. Frequency distribution of replacements in terms of the computer's age

and 61 months for the two standards. In 1989, corporate policy restricted the lifespan of mainframe computers to 6 years. Recently, this was shortened to 5 years, a reflection of the rapid pace of technological process. Figure 4 shows the replacement frequency of computer systems in the firm.

The other notable fact is that the interval between consecutive actions for computer systems assigned to certain tasks shortens as the sample progresses.[13] Part of this growth was undoubtedly due to the huge increase in performance and the corresponding drop in the per unit cost of

[13] Obviously, there are four combinations of consecutive actions: (i) upgrade–replacement; (ii) replacement–upgrade; (iii) upgrade-upgrade; (iv) replacement-replacement.

DOI: 10.1002/jae

processing capacity of mainframes, a consequence of 'Moore's law'. However, there was also tremendous growth in the need for computers for billing, account processing and other tasks, due to the rapid growth in the telecommunications industry over this same time period.

Figure 5[14] shows that cost per capacity decreases rapidly from 1994 to 1999. The demand for the services provided by the company also grows tremendously during this period. More frequent upgrades and replacements are seen from 1995 to 1997, when the demand for services increases by the greater amounts. After mid 1998 there is very little upgrade/replacement observed, since demand decreases significantly as a result of the economic recession. Figure 6[15] shows the trend of total demand.

The trends in average capacity are illustrated in Figures 7 and 8.[16] Both figures show that capacities increased rapidly from 1994 to 1998, during which time cost per capacity and demand change rapidly. However, since the reduction in cost per capacity is much larger than the increases in demand, we would conjecture that the effect of cost per capacity on capacity increases are much larger than the effect of demand.

The average frequency of upgrades for an individual computer system is 2.3 times. The maximum frequency of consecutive upgrades is four times. This is because each computer has limited slots for upgrades. Once the upgrade slots are full, the system needs to be replaced in order to increase capacity or to meet growing demand. The average frequency of replacements at each task level is approximately two, though some tasks undergo three or more replacements. Meanwhile, there are some tasks which do not undergo any replacements.

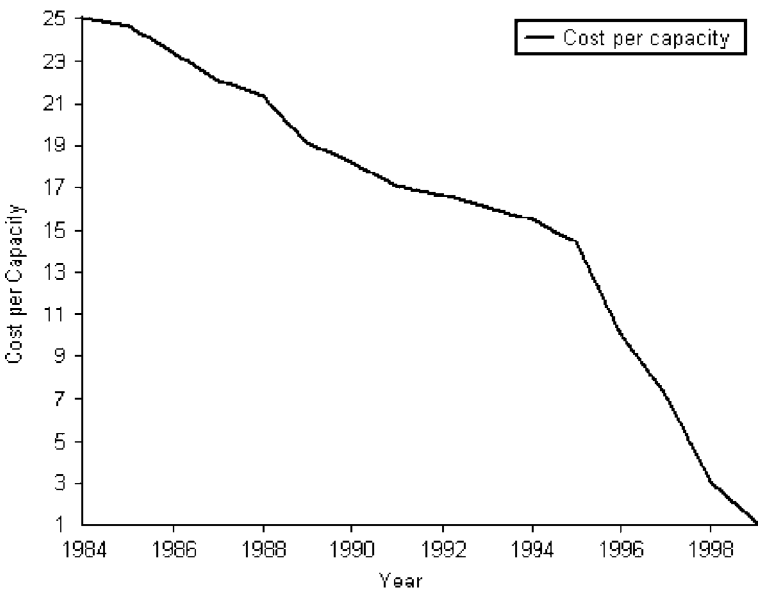

Figure 5. 'Real price' of semiconductors (all values are normalized)

[14] *Source*: SIA Annual Databook.
[15] A detailed explanation of the unit of demand is in Section 5.1.
[16] In both Figures 7 and 8, the *y*-axis represents the weighted average of mainframe computer capacity. The three most important components of computers are the CPU, memory and the hard disk. Each of these three components is weighted as follows: CPU 0.5, memory 0.25 and hard disk 0.25. These weighted average capacities are then discretized for simplification. (More details are given in a later section.) These weights were confirmed by several systems administrators in the company. Figures 7 and 8 show average capacities in terms of MIPS and TPC standards.

DOI: 10.1002/jae

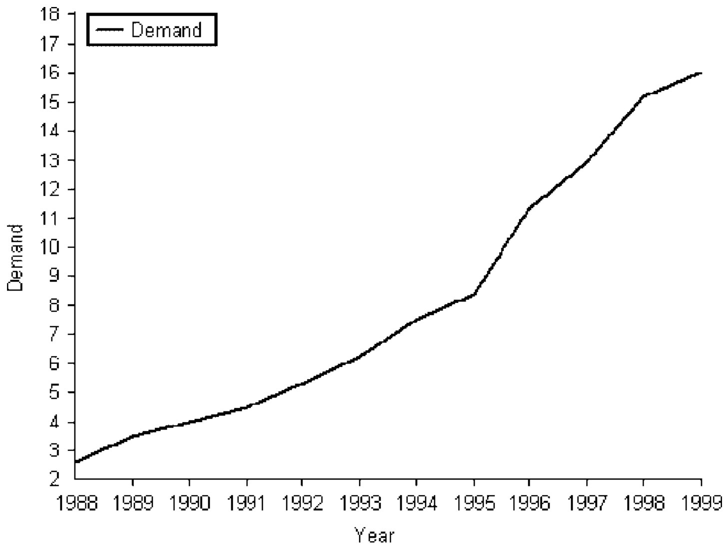

Figure 6. Trend of total demand (all values are normalized)

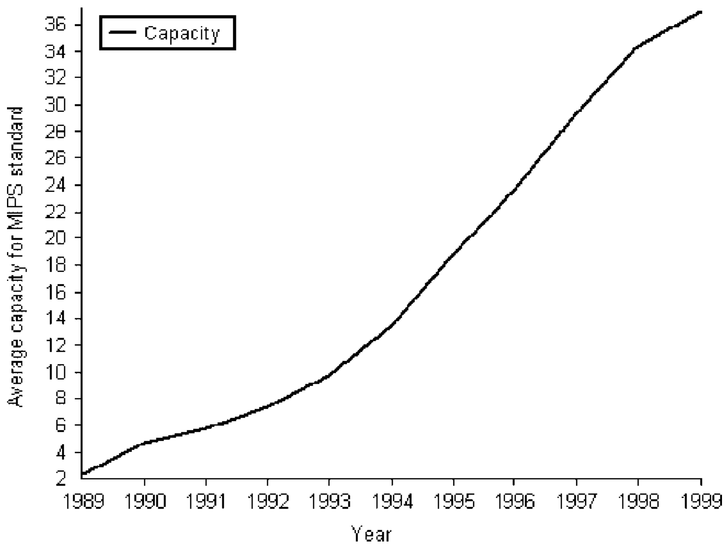

Figure 7. Trend in capacity (MIPS standard) (all values are normalized)

4. THE MODEL

This section develops a stochastic dynamic programming model in order to explain the observed pattern of replacement and upgrade observed in the data and to determine whether the pattern can be explained as an optimal strategy for the firm. The stochastic DP model consists of a vector of *state variables* X_t, *control variables* a_t, a *profit function* $\pi(X, a)$, a *discount factor* β and a *Markov transition density* $p(X'|X, a)$, representing the stochastic law of motion for the states of computer systems. I assume that the state variable X_t can be partitioned into two components: $X_t = (x_t, \varepsilon_t)$, where x_t is an *observed state vector* and ε_t is an *unobserved state vector*. The system administrator observes both components of X_t, but the econometrician observes only x_t. The system

DOI: 10.1002/jae

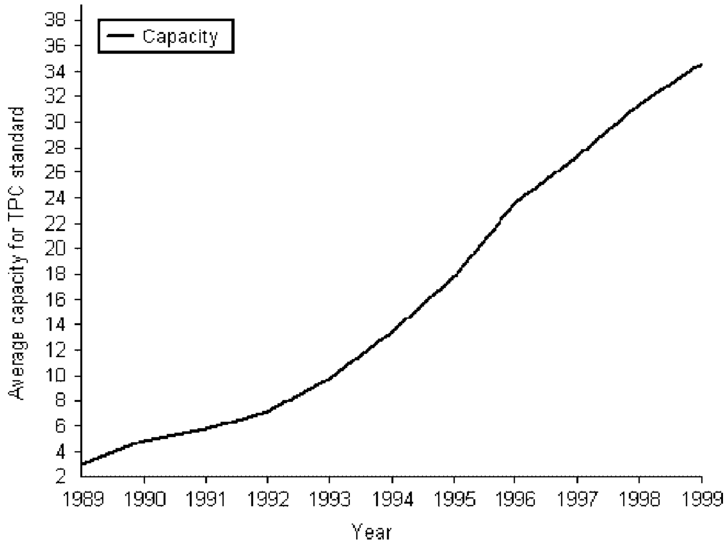

Figure 8. Trend in capacity (TPC standard) (all values are normalized)

administrator weighs the consequences of various operating decisions given the states of various computer systems and attempts to perform the best actions. I assume that the result of this decision process can be summarized by a vector of current net benefits (or costs, if negative) corresponding to each operating decision.

My final objective is to explain the data by deriving a stochastic process $\{a_t, X_t\}$ with an associated likelihood function $L(a_1, \ldots, a_n, X_1, \ldots, X_n, \theta)$ formed from the solution to a particular optimal stopping problem. A simplified discussion of the model for the 'replacement/upgrade decision rule' can be presented as follows. In the absence of the stochastic i.i.d. shocks, the rule should be an age/cost/demand-dependent threshold rule, e.g. for a given mainframe age, when demand exceeds an initial threshold there is an upgrade. If demand exceeds a second higher threshold there is a replacement of the system. Similarly, when the cost per unit to replace a mainframe computer is sufficiently close to the cost paid for the existing unit, there is inaction, but when the cost per unit of replacement falls sufficiently far to below the cost per unit paid a replacement occurs.

4.1. Choice and State Variables

Suppose that, in every month of the year, a system administrator investigates the status of each computer system and decides whether to upgrade it, replace it or keep it. Thus the choice set is $A_t = \{0, U, 1\}$, where $(a_t = 0)$ represents keeping the system unchanged, $(a_t = U)$ is an upgrade, and $(a_t = 1)$ is a replacement of the system. When the choice is to replace, the system administrator needs to choose the capacity of the new system, i.e. there are n sub-choices of capacities, $K_1, \ldots, K_n$. Each K_r, where $r = 1, \ldots, n$, is a capacity choice for replacement.

The final choice set is as follows: a: $A = \{0, U, K_1, \ldots, K_n\}$, i.e. keep $= 0$, upgrade $= u$ and replace $= (K_1, \ldots, K_n)$.

I assume that two state variables−the capacity and the age of a current computer system−are discrete. Two continuous variables are also included: the demand for services and the cost per capacity in the computer market.

DOI: 10.1002/jae

The observed state set in the model is x: $x_t = \{d_t, k_t, g_t, c_t\}$, where d_t = demand for service provided by each mainframe, k_t = current capacity of the computer system, g_t = age of each computer system and c_t = real cost per capacity, which is determined by the market for computer capacity. The two state variables g_t and k_t explain internal states of computers, while d_t and c_t represent external states of the computer systems.

The aggregate demand for the firm's services, D_t, consists of the sum of the individual demands, $d_{t,j}$, with $d_{t,j} = \xi_j D_t$, where $d_{t,j}$ is the demand for a task j at time t, where $0 < \xi_j < 1$.[17] In order to calculate a fraction ξ_j, for a demand, $d_{t,j}$, that a specific task serves, I sum the capacities of all computer systems at time t and assume that a proportion of the capacity of a certain system corresponds to the fraction of the demand for the system.[18] The aggregate demand for services is assumed to follow an AR(1) process, i.e. $\ln(D_t) = \alpha + \rho \ln(D_{t-1}) + \upsilon_t$ with $\upsilon \sim$ i.i.d. $N(0.\overline{\omega}^2)$. In our empirical results, we find that $|\rho| < 1$ so the demand process is stationary[19] and $\ln(D_t)$ is normally distributed with mean $\frac{\alpha}{1-\rho}$ and variance $\frac{\overline{\omega}^2}{1-\rho^2}$.

The real cost per capacity, c_t, is bounded below zero and evolves as follows:

$$c_{t+1} = \left\{ \begin{array}{c} \delta_t c_t \text{ with probability } 1-b \\ c_t \text{ with probability } b \end{array} \right\} \tag{1}$$

where δ_t has a truncated normal distribution with mean μ and variance ν^2 with a range of $0 < \delta_t < 1$. The parameter b represents a probability of cost per capacity.

Therefore, we have the following probability: $p(c_{t+1} \le z \mid c_t) = (1-b) \times p\{\delta_t c_t \le z\} + b \times I(c_t \le z)$.

The age variable, g_t, represents the age in months of each computer system. Recall that this firm has a predetermined rule for replacement based on the age of each system and I keep track of the age of each system.

4.2. Profit Function

I assume that each mainframe computer system is specifically associated with a certain task. Further, there is exactly one computer system per task. Also, the purchase of additional computers as an alternative to replacement is prohibited.[20] For example, the UNISYS 1100/72H2 mainframe usually works for the billing development task, the IBM RS/6000 SP system serves the marketing support task, and the DIGITAL AlphaServer 2100 4/275 is assigned to the line management task. In fact, these systems serve their own specific tasks by themselves. Furthermore, when the firm should replace these systems, the firm usually replaces them with the same series as the original systems. Therefore, it is appropriate to assume that each computer's profit is the same as its corresponding task's profit.

The profit function for any individual task, $\pi(d_t, k_t, g_t, a_t, \theta_1)$ is as follows:

$$\pi(d_t, k_t, g_t, a_t, \theta_1) = R(q(k_t, g_t), d_t, B, \theta_1, a_t) - C(k_t, g_t, d_t, c_t, \theta_1, \varepsilon) \tag{2}$$

[17] In fact, I observe only D_t, not $d_{t,j}$.

[18] Thus I have $\xi_j = \frac{k_{t,j}}{\sum_i k_{t,i}}$.

[19] If telecommunications follow the typical S-curve of diffusions, eventually demand flattens out (and thus becomes stationary) even though initially at the beginning of the diffusion of the technology (e.g. cell phones) the demand is growing rapidly. If the initial condition for this process is not from the ergodic distribution, the actual process for demand will not be stationary, but will asymptotically converge to a stationary process.

[20] This is based on what I observed from the data.

DOI: 10.1002/jae

where $R(q(k_t, g_t), d_t, \theta_1, B, a_t)$ is a revenue function and $C(k_t, g_t, d_t, c_t, \theta_1, \varepsilon)$ is a cost function. Adjusted capacity at time t, $q(k_t, g_t)$, is represented in the revenue function as a function of capacity k_t and age variables g_t. This adjusted capacity illustrates how capacity contributes to the revenue function. For example, the contribution of capacity to profit will decline as a computer gets old. B in the revenue function is a shadow price, which can be interpreted as an average rate of use of a certain computer. This can be calculated as (total demand)/(total capacity) at each time period. The θ_1 in the revenue and cost functions represents a set of unknown parameters.

k_t, which is the capacity at time t, in both revenue and cost functions has the following form depending on choices a_t, such as $k_t = \left\{ \begin{array}{cc} k_{t-1} & a_t = 0 \\ k_{t-1} + h & a_t = U \\ K_r & a_t = K_r \end{array} \right\}$, where a_t is the decision at time t and h is a capacity increase by upgrade, with $h = 1, 2$. I limit the h to 1 or 2 because, as discussed earlier, there is a limit for upgrades. One example of this is the limited availability of upgrade slots in mainframes. When there is a replacement, the new mainframe has a new capacity K_r, which is different from and superior to both k_{t-1} and $k_{t-1} + h$.

The revenue function can have either a flexible or a restrictive functional form. The flexible functional form is $B \times F(q(k_t, g_t), d_t, \theta_1, a_t)$, where the function F can be linear, a square root, quadratic, cubic[21] or a mixed form. The restrictive form is the 'minimum' function, G, such that $R(d_t, k_t, g_t, \theta_1, B, a_t) = B \times G(\min(q(k_t, g_t), d_t), g_t, d_t, \theta_1, a_t)$. The intuition behind a minimum function is that each mainframe computer system operates with the minimum capacity necessary to meet the current demand for its task.

The cost function has the following structure, dependent on the available choices:

$$\begin{aligned} & C(k_t, g_t, d_t, c_t, a_t, \theta_1, \varepsilon) \\ &= \left\{ \begin{array}{ll} m(d_t, k_t, g_t, \theta_1) + \varepsilon(0) & a_t = 0 \\ m(d_t, (k_t - h), g_t, \theta_1) + UC((k_t - k_{t-1}), c_t, \theta_1) + \varepsilon(U) & a_t = U \\ F(k_t, \theta_1) + r(k_t, c_t, \theta_1) - s(k_{t-1}, c_t, \theta_1) + \varepsilon(K_r) & a_t = K_r \end{array} \right\} \end{aligned} \quad (3)$$

In the cost function, $m(d_t, k_t, g_t, \theta_1)$ at $a_t = 0$ represents a maintenance cost for 'keep' and 'upgrade' decisions, since each mainframe computer system should receive regular maintenance to perform its task without interruption. $UC((k_t - k_{t-1}), c_t, \theta_1)$ in the cost for upgrade decision illustrates an upgrade cost for new capacity. In the case of a replacement cost function, $F(k_t, \theta_1)$ and $r(k_t, c_t, \theta_1)$ at $a_t = K_r$ are the fixed cost of replacement and a variable replacement cost respectively, while $s(k_{t-1}, c_t, \theta_1)$ is the value of a scrapped computer. It is this firm's policy to assume that any scrapped computer systems have no resale value. This is, in fact, not the case since these systems maintain a small resale value on the open resale market. Since $s(k_{t-1}, c_t, \theta_1)$ is included in the cost function, it is expected to have a negative sign. I assume that there is no maintenance cost for replacement.

I incorporate unobserved state variables, $\varepsilon(a)$, by assuming that unobserved costs $\{\varepsilon(0), \varepsilon(U), \varepsilon(K_r)\}$,[22] follow a specific stochastic process, which is an *i.i.d. multivariate extreme value distribution*, i.e. $q(\varepsilon|X) = \Pi_{j \in A(X)} \exp\{-\varepsilon(j)\} \exp\{-\exp\{-\varepsilon(j)\}\}$.

[21] Table III.

[22] $\varepsilon(0)$ is an unobserved cost of keeping a computer system, such as a managerial cost to prevent systems failures, a cost for service contracts, or certain other tolerance costs resulting from not replacing or upgrading. A positive value for $\varepsilon(0)$ could be interpreted as unobserved system overload, which indicates that a corresponding computer system should be upgraded or replaced. Alternatively, a positive $\varepsilon(0)$ could represent the expiration of a service contract or an unobserved component failure that requires the corresponding computer to be repaired. A negative value of $\varepsilon(0)$ could be interpreted as a report from a system administrator that a computer system has enough capacity to cover the current demand and is working smoothly. $\varepsilon(U)$ is an unobserved cost associated with upgrading computer systems. A negative value of $\varepsilon(U)$ could indicate that an upgraded computer system has plenty of upgrade slots, and that there are enough computing components to upgrade, whereas a positive value could be interpreted as the corresponding computer

DOI: 10.1002/jae

4.3. Bellman Equation and Likelihood Function

The optimal value function V_θ for each task is defined by $V_\theta(x, \varepsilon) = \max_{a \in A}[\pi(x_t, a, \theta_1) + \varepsilon_t(a) + \beta EV_\theta(x_t, \varepsilon_t, a)]$, where $EV_\theta = \int_y \int_\eta V_\theta(y, \eta) p(\mathrm{d}y, \mathrm{d}\eta | x_t, \varepsilon_t, a, \theta_0)$.

Then, as an optimal policy rule, a stationary decision rule is defined as

$$a_t = z(x_t, \varepsilon_t, \theta) := \operatorname*{arg\,max}_{a \in A(x_t)} [\pi(x_t, a_t, \theta) + \varepsilon_t(a) + \beta EV_\theta(x_t, a_t, \varepsilon_t)] \tag{4}$$

and $z(x_t, \varepsilon_t, \theta)$ is the optimal control.

4.3.1 Markov Transition Probability

I follow Rust (1987) in making the standard simple assumption that the transition probability φ can be factored as $\varphi(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, a_t, \theta_0) = p(x_{t+1} | x_t, a_t, \theta_0) q(\varepsilon_{t+1} | x_{t+1})$, which Rust calls a 'Conditional Independence Assumption (CI)', where θ_0 is a vector of unknown parameters characterizing the transition probability for the observable part of the state variables. From the setup of choice variables, θ_0 is defined as $\theta_0 = \{a, \rho, \mu, \nu, b\}$.

In order to reach $p(x_{t+1} | x_t, a_t)$, I assume that all state variables are independent of one another. Therefore, $p(x_{t+1} | x_t, a_t) = p(x^1_{t+1} | x^1_t, a_t) \times p(x^2_{t+1} | x^2_t, a_t) \times p(x^3_{t+1} | x^3_t) \times p(x^4_{t+1} | x^4_t)$, where $x^1_t = k_t$, $x^2_t = g_t$, $x^3_t = d_t$ and $x^4_t = c_t$.

However, because of the assumption that deterministic evolution of capacity and age variables depends on replacement choices, I can focus only on $p(d_{t+1} | d_t)$ and $p(c_{t+1} | c_t)$.

4.3.2 Policies of the Actions

With this assumption of ε, we can rewrite V_θ as follows:

$$V_\theta(x, a) = \Bigg\{ \pi(x, a, \theta) + \beta \int_y \sigma \log \left[\sum_{a' \in \{0, U, K_1 \ldots K_n\}} \exp[(\pi(y, a', \theta_1) + \beta EV_\theta(y, a'))/\sigma] \right] p(\mathrm{d}y | x, a, \theta_0) \Bigg\} \tag{5}$$

where σ is a standard deviation of ε_t.

Conditional choice probabilities $P(a_t | x, \theta)$ are then given by

$$P(a = 0, \text{keep} | x, \theta) = \frac{\exp\{(\pi(x, \theta_1, a = 0) + \beta EV_\theta(x, a = 0))/\sigma\}}{\sum_{a'} \exp[(\pi(x, a', \theta_1) + \beta EV_\theta(x, a'))/\sigma]} \tag{6}$$

$$P(a = U, \text{upgrade} | x, \theta) = \frac{\exp\{(\pi(x, \theta_1, a = U) + \beta EV_\theta(x, a = U))/\sigma\}}{\sum_{a'} \exp[(\pi(x, a', \theta_1) + \beta EV_\theta(x, a'))/\sigma]} \tag{7}$$

$$P(a = K_r, \text{replace} | x, \theta) = \frac{\exp\{(\pi(x, \theta_1, a = K_r) + \beta EV_\theta(x, a = K_r))/\sigma\}}{\sum_{a'} \exp[(\pi(x, a', \theta_1) + \beta EV_\theta(x, a'))/\sigma]} \tag{8}$$

---

system having a limited number of upgradeable slots. $\varepsilon(K_r)$ is also interpreted as an unobserved cost when the action of replacement occurs. A positive value for $\varepsilon(K_r)$ could be interpreted as a price increase of a backup system during a replacement period, whereas a negative value could be interpreted as a price decrease of a backup system. I am unable to identify these unobserved costs in this dataset. I have also implicitly assumed that the stochastic processes $\{x^m_t, \varepsilon^m_t\}$ are independently distributed across different computer systems, m, except for the two state variables demand for services, d_t and cost per unit capacity, c_t.

DOI: 10.1002/jae

4.3.3 Log-Likelihood Function

We then have the two partial log-likelihood functions at time t as follows:

$$L_t^1 = \ln(P(a_t|x_t, \theta)) \text{ and } L_t^2 = \ln(p(x_t|x_{t-1}, \theta_0)) \quad (9)$$

where L_t^1 is a log-likelihood function of the conditional-choice probability and L_t^2 is a log-likelihood function of the transition probability. Thus we have the total log-likelihood function as follows:

$$L(a_1, \ldots, a_T, x_1, \ldots, x_T|x_0, a_0, \theta) = \sum_{t=1}^{T} \ln(P(a_t|x_t, \theta)) + \sum_{t=1}^{T} \ln(p(x_t|x_{t-1}, \theta_0)) \quad (10)$$

5. ESTIMATION

The general method to solve the fixed-point problem is a discretization of observed state variables. When the observed state variable is continuous, the required fixed point is in fact an infinite dimensional object. Therefore, in order to solve the fixed-point problem, it is necessary to discretize the state space so that the state variable takes on only finitely many values. But there are limits regarding this method: (i) 'curse of dimensionality'; (ii) the limits it places on our ability to solve high-dimensional DP problems. However, the discretization method may not be applicable to computer replacement research to solve the fixed-point problem, because the aforementioned problems seriously affect the calculation time of a nested fixed-point algorithm, because the nested fixed-point algorithm uses the fixed-point algorithm outside of the maximum likelihood estimation.

Instead, we use the parametric approximation,[23] which parametrizes the value functions with state variables, which is a practical, efficient and numerically stable method for estimating certain structural models lacking closed-form solutions with high-dimensional state space. After incorporating the parametric approximation, which eliminates the need to discretize the continuous state variables, the estimation requires the nested fixed-point algorithm. This algorithm is intended to find parameters that maximize the likelihood functions subject to the constraint that the function EV_θ is the unique fixed point. This estimation procedure can be called a nonlinear least squares–nested fixed-point estimation (NLS-NFXP).

The revenue and cost equations were estimated with both restrictive and flexible functional forms. To be more precise, one restrictive functional form and several flexible functional forms were estimated; linear, square root, quadratic, cubic and mixed forms. Among these functional forms, the cubic form gives the best estimation results. In fact, all additional terms, from linear to quadratic, and from quadratic to cubic forms, are statistically significant at the 95% confidence level.

The parameters for state variables, θ_0, and parameters for revenue and cost functions, θ_1, are estimated separately. First, the parameters for state variables are estimated. Estimates of θ_1 are calculated based on these estimates of θ_0.

5.1. Nonlinear Least Squares–Nested Fixed-Point Estimation

The estimation procedure by NLS-NFXP is a two-step procedure. First, outside of the system, θ_0, parameters for state variables are estimated separately from the structural parameters; next, inside of the system, the structural parameters θ_1 are estimated by the nested fixed-point algorithm.

[23] Further information regarding the parametric approximation can be provided on request; alternatively, it may be found at http://plaza.snu.ac.kr/ sungcho/

DOI: 10.1002/jae

In other words, inside of the maximum likelihood estimation the above nonlinear least squares estimation (NLS) is performed and fixed points, EV_θ, are calculated. Based on the fixed points, the maximum likelihood estimation is performed.[24]

5.1.1 *Results of Estimation*

Parameters for demand and cost per capacity. For simplicity, I estimate the parameters $\theta_0 = \{\alpha, \rho, \mu, \nu, b\}$, which govern the transition probabilities for demand and cost per capacity separately from the parameters of the profits function. As I mentioned earlier, an individual demand, $d_{t,j}$, for a task j is a fraction of aggregated demand, D_t, such that $d_{t,j} = \xi_j D_t$. In order to calculate a fraction ξ_j, I sum all the capacities of the computer systems at each time t and assume that a proportion of the system's capacity corresponds to a fraction of the system's demand. The parameters for D_t are estimated by maximum likelihood estimation.

The parameters of cost per capacity, c_t, are obtained by the maximum likelihood estimation method.[25] Table II presents the estimation results for D_t and c_t.

Structural estimates of revenue and cost functions. Tables V and VI are the results of a structural estimation in terms of the cubic functional forms in Table III.[26] The tables report the structural parameter estimates computed by maximizing the likelihood function L_f^1 in equation (9) using the nested fixed-point algorithm. I present structural estimates for the unknown parameters for the cubic specifications suggested in Tables V and VI.[27] The estimation results for $\beta = 0.999$[28] correspond to a dynamic model in which the present value of current and future profit streams is maximized by the investment decisions of the firm.

Most parameters of the revenue and cost functions are estimated precisely and have the expected sign. In Table VI parameters θ_{42}, θ_{43} and θ_{44} (except the constant term) of the revenue function for scrapped computers are insignificant at the 95% level in several functional forms. This can be explained by two possible factors. First, the proposed functional form could be misspecified. Second, any scrapped computer has a constant lump-sum value regardless of its remaining capacity. According to several interviews with system administrators of the firm, the second assumption seems more reasonable, since the firm does not care about the value of scrapped computer systems, and they in fact donate old, replaced computer systems to charity.

Table II. Parameter estimates for state d_t and c_t

d_t		c_t	
Parameters	Estimate	Parameters	Estimate
α	1.0229 (0.128)	b	0.759 (0.094)
ρ	0.9405 (0.038)	μ	9.127 (0.065)
ϖ^2	0.0245 (0.005)	ν^2	8.794 (0.017)
Likelihood	−42.837	Likelihood	−51.237
Obs. size	62	Obs. size	62

Note: Standard errors in parentheses.

[24] The Berndt, Hall, Hausman and Hall (BHHH) algorithm is used, along with numerical derivatives.

[25] The log-likelihood function of c_t is $L_f^2(c_1, \ldots, c_T|\theta_0) = \sum_{t=1}^{T} \ln(P(c_t|c_{t-1}, \theta_0))$.

[26] I also estimated a structural estimation for the 'minimum' function as a restrictive functional form.

[27] When I tried the estimation additionally with $\beta = 0.99$ and $\beta = 0.95$, there was no distinguishable difference.

[28] I fix $\beta = 0.999$

Table III. Cubic functional forms used in the model as a flexible form

	Specifications
Revenue	
Keep	$\theta_{11} + \theta_{12}(q(f_t(k, a_t), g_t, \theta_{35}) \times d_t \times \gamma) + \theta_{13}(q(f_t(k, a_t), g_t, \theta_{35}) \times d_t \times \gamma)^2$ $+\theta_{14}(q(f_t(k, a_t), g_t, \theta_{35}) \times d_t \times \gamma)^3$
Upgrade	$\theta_{21} + \theta_{22}(h \times \gamma \times d_t) + \theta_{23}(h \times \gamma \times d_t)^2 + \theta_{24}(h \times \gamma \times d_t)^3$ $\theta_{25}(q((f_t(k, a_t) - h), g_t, \theta_{35}) \times \gamma \times d_t) + \theta_{26}(q((f_t(k, a_t) - h), g_t, \theta_{35}), g_t) \times \gamma \times d_{t,j})^2$ $+\theta_{27}(q((f_t(k, a_t) - h), g_t, \theta_{35}) \times \gamma \times d_t)^3$
Replacement	$\theta_{31} + \alpha_{32}(f_t(k, a_t) \times \gamma \times d_t) + \theta_{33}(f_t(k, a_t) \times \gamma \times d_t)^2 + \theta_{34}(f_t(k, a_t) \times \gamma \times d_t)^3$
Cost	
Keep	$I(f_t(k, a_t) \geq d_t)\{\theta_{51} + \theta_{52}(f_t(k, a_t) \times m_t) + \theta_{53}(f_t(k, a_t) \times m_t)^2 + \theta_{54}(f_t(k, a_t) \times m_t)^3\}$ $+I(f_t(k, a_t) < d_t)\{\theta_{51} + \theta_{52}(f_t(k, a_t) \times m_t) + \theta_{53}(f_t(k, a_t) \times m_t)^2$ $+\theta_{54}(f_t(k, a_t) \times m_t)^3 + \theta_{55}(l \times (d_t - q(f_t(k, a_t), g_t, \theta_{35}))$ $+\theta_{56}(l \times (d_{t,j} - q(f_t(k, a_t), g_t, \theta_{35}))^2 + \theta_{57}(l \times (d_t - q(f_t(k, a_t), g_t, \theta_{35}))^3\}$
Upgrade	$\theta_{61} + \theta_{62}(f_t(k, a_t) \times m_t) + \theta_{63}(f_t(k, a_t) \times m_t)^2 + \theta_{64}(f_t(k, a_t) \times m_t)^3$ $+\theta_{65}(c_t \times f_t(k, a_t) \times cp) + \theta_{66}(c_t \times f_t(k, a_t) \times cp)^2 + \theta_{67}(c_t \times f_t(k, a_t) \times cp)^3$
Replacement	$\theta_{71} + \theta_{72}(cp \times f_t(k, a_t)) + \theta_{73}(cp \times f_t(k, a_t))^2 + \theta_{74}((c_t \times cp) \times f_t(k, a_t))$ $+\theta_{75}((c_t \times cp) \times f_t(k, a_t))^2 + \theta_{76}((c_t \times cp) \times f_t(k, a_t))^3$
Scrap	$\theta_{41} + \theta_{42}((k_t \times c_t)) + \theta_{43}((k_t \times c_t))^2 + \theta_{44}((k_t \times c_t))^3$

Notes:
cp is a scale parameter for cost functions from calibration ($cp = 2.013$); m_t is a unit maintenance cost which is increasing with g_t, such as $m_t = m(g_t, \theta_{81}) = \theta_{81} \times \sqrt{g_t}$.
The assumptions imposed on these functional forms, $f_1(k_t, g_t, rm)$ and um_t are due to the large number of unknown parameters. These assumptions for $f_1(k_t, g_t, rm)$ and um_t can be released in further research.
Flexible form. Table III shows cubic functional forms of profit and cost functions of the firms associated with keep, upgrade, replacement and scrap decisions.
$q(f_t(k, a_t), g_t, \theta)$ illustrates how capacity contributes to the revenue functions, such as an adjusted capacity. For example, the contributions of capacity will decline as the computer gets old. However, in the revenue function for replacement, the replacement capacity K_r will fully contribute to the revenue function for replacement. $q(f_t(k, a_t), g_t)$ is assumed to be a simple function which is increasing in $f_t(k, a_t)$, and decreasing in g_t, such as $q(f_t(k, a_t), g_t, \theta) = (\theta_{35} \times f_t(k, a_t))/\sqrt{g_t}$.
γ is components of a set of unknown parameter, θ_1, which is a measure of unit value for telecommunication services the firm provides. Also, it can be interpreted as an average value of unit demand for aggregated services. l, as a component of θ_1, is a unit labor charge per capacity in order to compensate for the shortage of the current adjusted capacity, $(d_t - q(f_t(k, a_t)))$. When demand exceeds the current capacity, the firm usually hires more labor to make up the shortage of the amount of $[l \times (d_t - q(f_t(k, a_t)))]$.

Figures 9–12[29] show the three policies (keep, upgrade and replace) and their profit functions, plotted against various costs per capacity, in the case where demand is lower than the current capacity with age fixed. In Figures 9 and 10 the profit functions of upgrade fall slightly, as cost per capacity increases due to the amount of upgrade. However, since replacement requires changing the current system as a whole, the cost of replacement will increase tremendously as cost per capacity increases. Thus, as cost per capacity increases, the likelihood of replacement falls and eventually reaches zero. When the cost per capacity is high enough, the best choice for keeping up with current demand becomes the choice of upgrade. Figures 11 and 12 show the same profit functions for older computer systems. As the computer gets older, replacement becomes preferable to upgrade. However, as cost per capacity increases, the probability of replacement falls and the probability of upgrade rises, making it the best choice.

Figures 13–16 show the three polices (keep, upgrade and replace) and three profit functions of old computer systems at various levels of demands for the system's tasks, with capacity and cost per capacity fixed. In Figures 13 and 14, as demand increases the profit functions for keep, upgrade and replacement are expressed as smoothly increasing curves. However, each policy behaves

[29] All figures are based on estimated parameters for cubic functional forms.

DOI: 10.1002/jae

Table IV. Explanation of a set of parameters, θ_1 in Table III

Parameters θ_1	Function	Verification of parameters
Revenue		
$\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}$	Keep	Revenue from old components
$\theta_{21}, \theta_{23}, \theta_{23}, \theta_{24}$	Upgrade	Revenue from upgraded components
$\theta_{25}, \theta_{28}, \theta_{27}$	Upgrade	Revenue from old components
$\theta_{31}, \theta_{32}, \theta_{33}, \theta_{34}$	Replacement	Revenue from new components
θ_{35}	Keep, upgrade	Adjusted capacity, $q(f(k, a), g_t, \theta)$
γ	All	An average value of unit demand
Cost		
$\theta_{41}, \theta_{42}, \theta_{43}, \theta_{44}$	Scrap	Scrapped value of computer
$\theta_{51}, \theta_{52}, \theta_{53}, \theta_{54}$	Keep	Maintenance cost for keep
$\theta_{55}, \theta_{56}, \theta_{57}$	Keep	Make up cost for shortage
$\theta_{61}, \theta_{62}, \theta_{63}, \theta_{64}$	Upgrade	Maintenance cost for upgrade
$\theta_{65}, \theta_{66}, \theta_{67}$	Upgrade	True upgrade cost
$\theta_{71}, \theta_{72}, \theta_{73}$	Replacement	Fixed cost[a]
$\theta_{74}, \theta_{75}, \theta_{76}$	Replacement	Variable cost[b]
θ_{81}	Keep, upgrade	Unit maintenance cost for keep and upgrade, $m(g_t, \theta_{81})$
l	Keep	Unit labor charge per capacity

[a] Fixed cost for replacement is invariable with respect to cost per unit capacity.
[b] Variable cost for replacement is variable with respect to cost per unit capacity.
Table IV explains the detail of a set of parameters, θ_1 associated with Table III.

differently. Until the point where the capacity is slightly above current demand, the choice of 'keep' is more likely to occur with decreasing likelihood. Beyond the point of current demand 'upgrade' will be more likely to occur for maximization of profits if the system is relatively new. However, for a relatively old system the choice of replacement outperforms the choice of upgrade and thus replacement is more likely to occur. Figures 15 and 16 show how cost per capacity affects the profit functions and policy rules. When cost per capacity is higher (Figures 15 and 16), upgrade becomes the most profitable choice for more levels of demand than when costs are lower. However, this situation changes when demand is much greater than current capacity.

5.2. Simulation Based on Estimation Results

Based on the estimated parameters, several simulations are performed to generate simulated data for comparison with real data. I simulate the life of a computer system in order to examine decisions regarding upgrade and replacement. In order to investigate the frequencies of replacement and upgrade, all computer systems of the firm–not just one–are simulated. However, when examining capacity evolution, the whole life of a certain task is simulated. The actual decision process is assumed to have randomness, i.e. any decision may be made, even though there is a most profitable choice among the three options in each period.

Frequency of upgrade. Figure 17 compares simulated data and actual data in terms of frequency of upgrade for computers at various ages. Generally, the shape and tendency of upgrade frequency resemble each other. Most upgrade activities occur approximately between 1.5 and 2 years of age and at 3.5 years of age. There are two reasons for this. First, according to Moore's law, computer processing capacity doubles every 18 months. Therefore, by upgrading its computer systems, the firm makes continuous efforts to keep up with technological progress in order to lower its operating costs. Second, the firm expands its services in order to meet rapidly growing demand through more frequent upgrades. The two humps in the frequency of upgrade can be intuitively explained in the model. When the firm purchases a new mainframe, each

DOI: 10.1002/jae

Table V. Structural parameter (θ_1) estimates for flexible form (cubic) ($\beta = 0.999$)

Revenue		MIPS		TPC	
Function	Parameters	Estimate	SE	Estimate	SE
Keep	θ_{11}	13.209	(2.125)	12.031	(1.045)
	θ_{12}	1.446	(0.028)	1.746	(0.294)
	θ_{13}	1.813*	(1.147)	1.659	(0.314)
	θ_{14}	1.124	(0.235)	1.056	(0.125)
Upgrade	θ_{21}	13.774	(1.238)	12.256	(1.001)
	θ_{22}	1.114	(0.136)	1.573	(0.354)
	θ_{23}	1.901	(0.243)	1.817	(0.347)
	θ_{24}	1.298	(0.021)	1.169	(0.185)
	θ_{25}	1.741*	(1.045)	2.001	(0.019)
	θ_{26}	0.589	(0.002)	1.035	(0.147)
	θ_{27}	2.184	(0.029)	3.206*	(3.267)
Replacement	θ_{31}	12.203	(1.562)	12.322	(1.511)
	θ_{32}	2.301	(0.037)	2.540	(0.124)
	θ_{33}	1.037*	(0.772)	1.163*	(1.217)
	θ_{34}	3.321	(0.056)	3.776	(0.194)
	θ_{35}	1.024	(0.014)	1.109	(0.037)
	γ	3.524	(0.002)	3.167	(0.351)
	B[a]	0.887	(0.031)	0.965	(0.019)

* Not significant at 95% level.
[a] The shadow price B was calculated separately.

new mainframe is expected to last for 5 years, which means, on average, coverage of 14.4% growth in demand for the firm's services per year. In the model, each upgrade is restricted to one or two units of capacity increase, approximately equivalent to a maximum 10% increase in unit capacity. This firm sets its threshold tolerance levels of computing capacity at 60% of total capacity for each mainframe. Thus, as time passes, its tolerance level drops as service demand increases. Usually, after a new mainframe is used for 2 years, the tolerance level decrease by 20% on average, which is a little higher than threshold. But, with the effect of a huge reduction in the cost of mainframes, the firm upgrades its mainframe at approximately 2 years to speed up its computing efficiency. Then, at 1.5 or 2 years after each upgrade, the computing level shrinks to 60% or 50% of capacity, which is equal to or less than the threshold level (60%). At this moment, the mainframe gets the last upgrade before replacement, which makes for a maximum of 2 years of use. If we loosen the assumption of upgrade capacity in the model, the upgrade frequency becomes sporadic. In comparison with the actual data, the upgrade frequency in the simulated data is slightly higher, but the difference is minimal and acceptable.

Frequency of replacement. Figure 18 compares simulated data and actual data in terms of frequency of replacement for various ages of computers. In general, the shapes and trends of replacement resemble each other. Most replacement activities occur between 4 and 5 years of age. Even though the frequency of replacement in the simulated data is slightly higher than in the actual data, the difference is minimal and acceptable. One noticeable fact is that several replacement activities occur in the period between 1 and 1.5 years of computer age in both datasets.

This situation has two compelling explanations. First, as is evident in Figure 17, computers tend to be upgraded for the first time at approximately 2 years of their age. However, replacement is more beneficial to the firm than upgrade in some cases because of costs, efficiency or unexpected increases in demand for the computer's services. When those computers require increased capacity,

DOI: 10.1002/jae

Table VI. Structural parameter (θ_1) estimates for flexible form (cubic) ($\beta = 0.999$)

Cost		MIPS		TPC	
Function	Parameters	Estimate	SE	Estimate	SE
Scrap	θ_{41}	16.269	(1.649)	17.185	(1.197)
	θ_{42}	1.532*	(2.487)	1.683*	(1.248)
	θ_{43}	0.191*	(1.549)	0.252*	(2.301)
	θ_{44}	1.245*	(2.432)	2.421*	(4.579)
Keep	θ_{51}	5.102	(1.032)	5.514	(1.154)
	θ_{52}	1.338	(0.026)	1.514*	(2.042)
	θ_{53}	0.254	(0.032)	0.212	(0.061)
	θ_{54}	0.248	(0.003)	0.401	(0.105)
	θ_{55}	0.265	(0.063)	0.315	(0.021)
	θ_{56}	0.951	(0.106)	0.759	(0.015)
	θ_{57}	0.417	(0.053)	0.699	(0.113)
Upgrade	θ_{61}	4.008	(0.887)	4.256	(0.984)
	θ_{62}	0.954	(0.058)	1.023	(0.254)
	θ_{63}	0.362	(0.032)	0.309	(0.046)
	θ_{64}	0.304	(0.079)	0.412	(0.094)
	θ_{65}	0.591	(0.017)	1.254	(0.008)
	θ_{66}	0.831*	(0.719)	0.954	(0.124)
	θ_{67}	1.127	(0.018)	1.551	(0.187)
Replacement	θ_{71}	4.518	(1.056)	4.341	(0.608)
	θ_{72}	2.231*	(1.143)	2.145*	(1.449)
	θ_{73}	0.767	(0.014)	0.697	(0.177)
	θ_{74}	2.732	(0.516)	2.198	(0.397)
	θ_{75}	1.815	(0.059)	1.254	(0.005)
	θ_{76}	2.218	(0.218)	1.758	(0.122)
	θ_{81}	0.998	(0.157)	0.972	(0.201)
	l	1.551	(0.059)	2.485	(0.038)
	Likelihood	−5991.64		−6487.51	
	Obs. size	5760		6840	

* Not significant at 95% level.

they are replaced rather than upgraded. Second, the firm tends to replace computer systems that fail to accomplish their given tasks within an initial testing period.[30] In general, the simulated data from the estimated parameters show a more frequent tendency to replace computer systems than do the actual data.

Total expenditures of the firm. Figure 19 compares actual data and simulated data in terms of total expenditures of the firm. In general, simulated total expenditure tracks the actual total expenditure well enough to ensure the model's usefulness. In addition, note that the firm's total expenditure does not increase over time, even though there are tremendous increases in computer capacities, as shown in Figures 7 and 8. This is because real cost per capacity decreases considerably over time, even though the firm increases its computer system's capacities greatly. Therefore, we can confirm the fact that the expenditure-decreasing effect of real cost per capacity surpasses the expenditure-increasing effect of computer system capacity.

Evolution of installed mainframe capacity. Figures 20 and 21 compare the evolution of capacities in the actual data and the simulated data in terms of TPC and MIPS standards. In both figures the first graph line, representing 'actual capacity', presents actual capacities from the data under the TPC and MIPS standards. The second graph line, representing 'simulated capacity by total

[30] In general, a 'lemon' is a computer that experiences failures within 1.5 years.

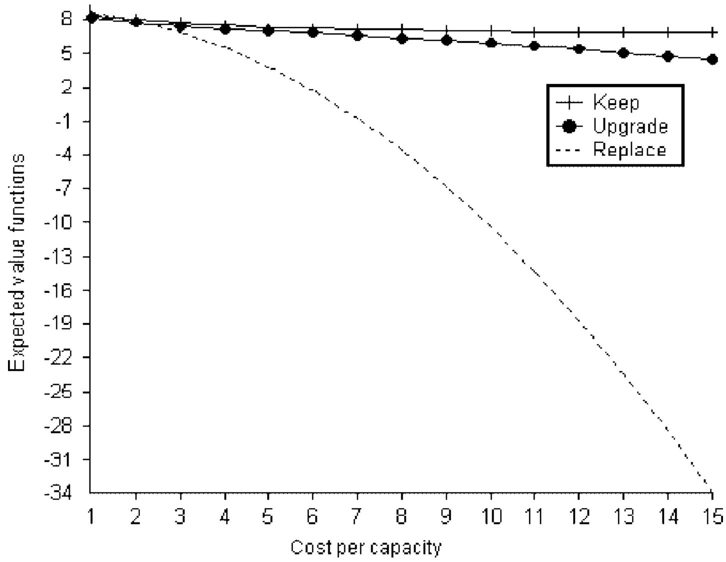

Figure 9. Profit functions of keep, upgrade, and replacement decisions for relatively new computers

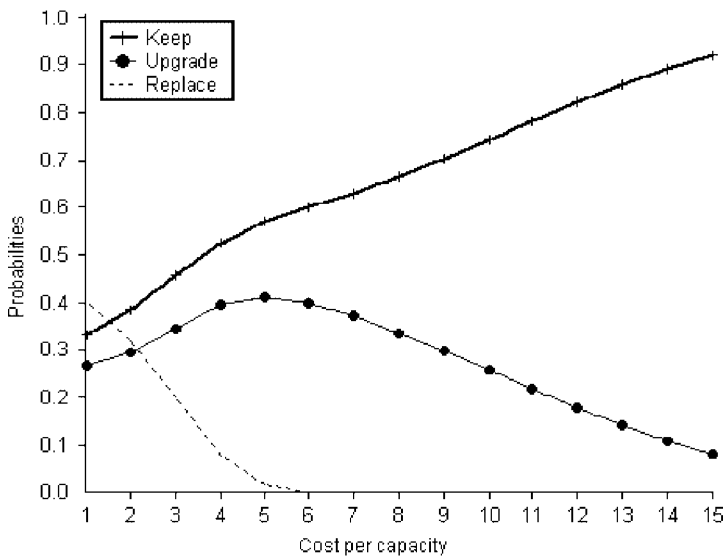

Figure 10. Three policy rules at various costs per capacity for relatively new computers

effect', shows simulated capacities under the same conditions as the actual data. The third graph line, representing 'simulated capacity by technology progress', illustrates simulated data without the effect of a demand increase, in which case only technological progress affects the firm's decision to keep, upgrade or replace current mainframe computers. The fourth graph line, 'simulated capacity by demand', shows simulated capacities when only demand affects the firm's decision.

The last two graph lines in each figure separately explain how technological progress and demand increases affect the capacity of computer systems. In Figure 20, where computer capacity is represented by the TPC standard, the actual capacity and simulated total capacity evolve in a

DOI: 10.1002/jae

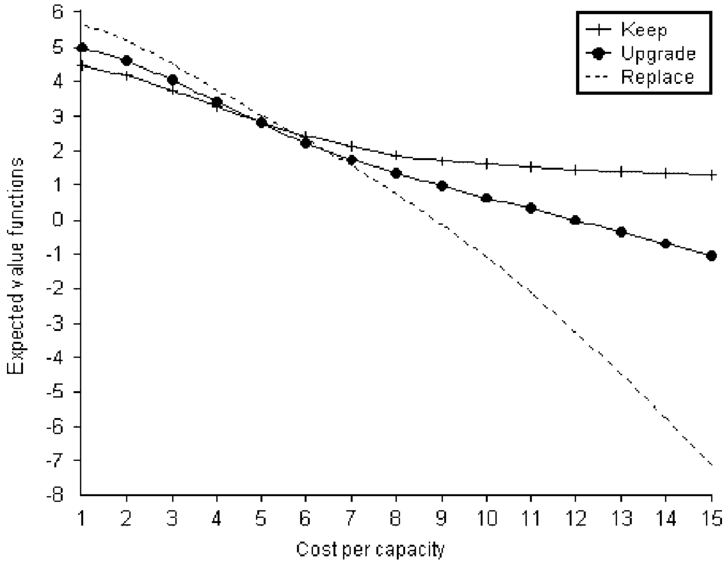

Figure 11. Profit functions of keep, upgrade and replacement decisions for relatively old computers

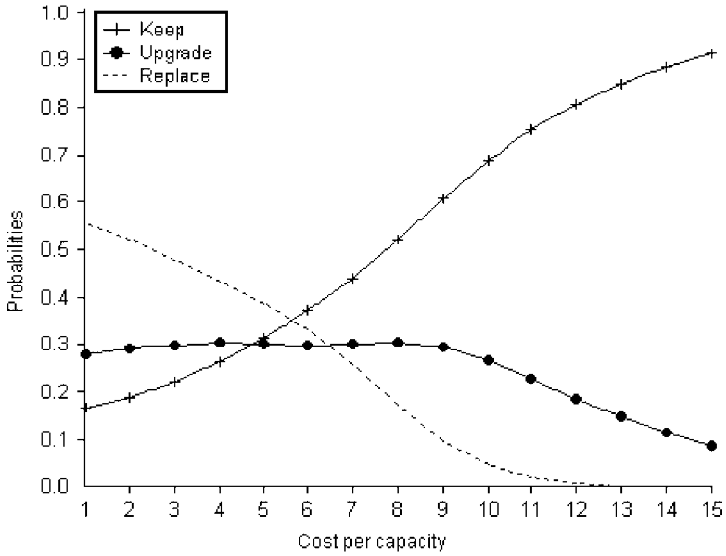

Figure 12. Three policy rules with various costs per capacity for relatively old computers

very similar manner. The simulated data track the actual data very accurately in the case of the TPC standard. In the case of the MIPS standard computer systems (Figure 21), actual capacity and simulated capacity evolve at a similar pace at the beginning of the time period, but as time goes on there is a slight discrepancy between the two capacities; in fact, the actual capacity lies above the simulated capacity. Nevertheless, the general evolving tendencies of the two capacities are similar to each other.

Separating the effects due to technological progress and demand for services shows clearly that technological progress plays a more significant role than demand for services does in the case of

DOI: 10.1002/jae

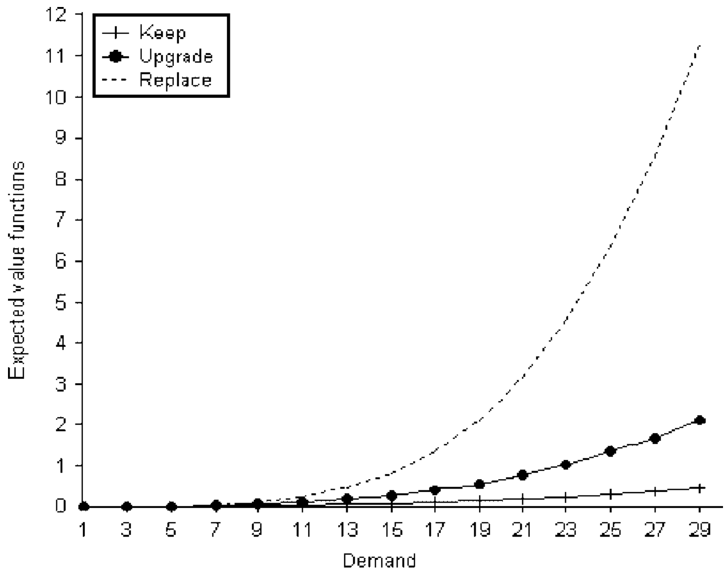

Figure 13. Profit functions of keep, upgrade and replacement in terms of relatively low cost per capacity

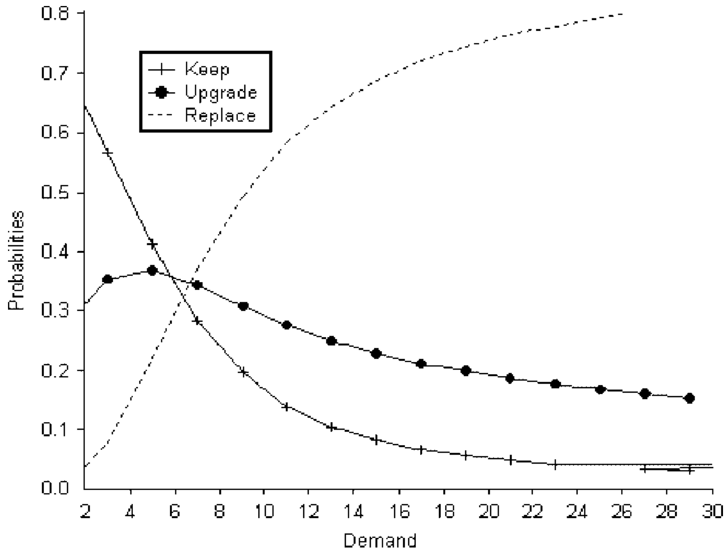

Figure 14. Three policy rules at various demand levels in terms of relatively low cost per capacity

TPC standard computer systems. This confirms our conjecture that technological progress is the first and main cause of mainframe computer system replacements and upgrades in the company. In the case of MIPS, the difference between the two effects of technological progress and demand increase is minimal, even though technological progress improves computer capacity slightly more than does the demand for services.

Note that for both Figures 20 and 21 the combined capacities from adding two simulated capacities based on technological progress and demand at each point of date exceed the simulated capacities by total effect. The simple addition of the two simulated capacities shows overlapping

DOI: 10.1002/jae

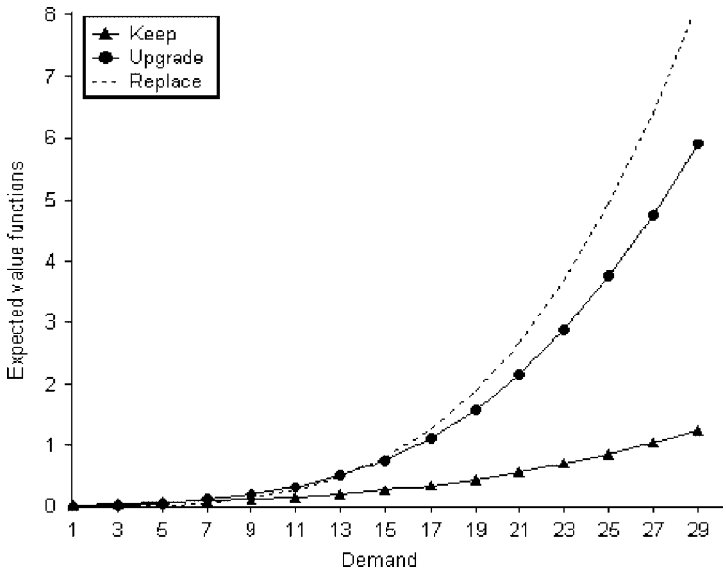

Figure 15. Profit functions of keep, upgrade and replacement in terms of relatively high cost per capacity

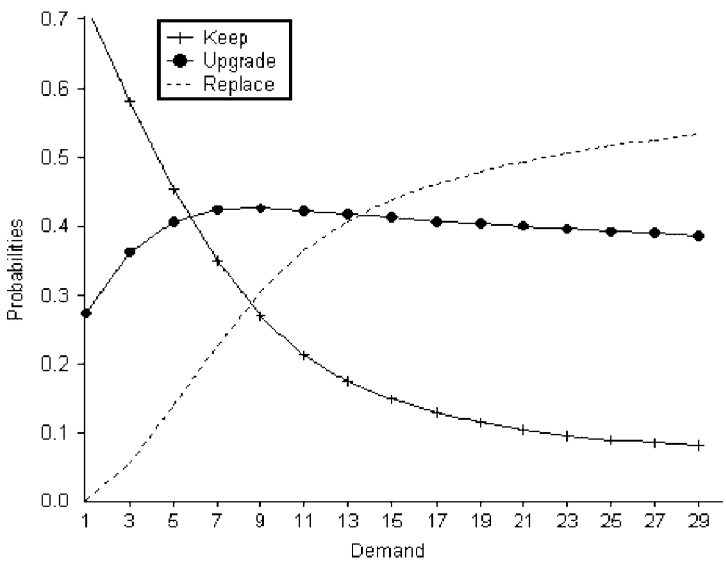

Figure 16. Three policy rules at various demand levels in terms of relatively high cost per capacity

capacities, which means there are idle and wasted capacities in the sum, since each effect should account for increases in capacity both individually and simultaneously. However, the simulated capacity by total effect is obtained by the optimization decision by considering both effects–technological progress and demand–simultaneously, rather in isolation from each other, taking into account that increased capacity can respond to technological and demand changes simultaneously. This illustrates a fairly expected and reasonable situation.

Policy of upgrade and replacement. Figures 22 and 23 present three simulated policies in two different situations. Figure 22 illustrates the situation in which the cost per capacity decreases

DOI: 10.1002/jae

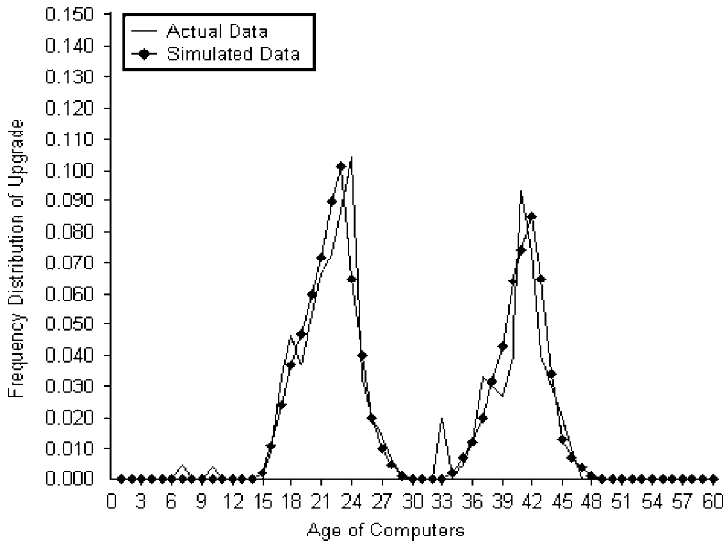

Figure 17. Comparison between simulated data and actual data in terms of frequency of upgrade

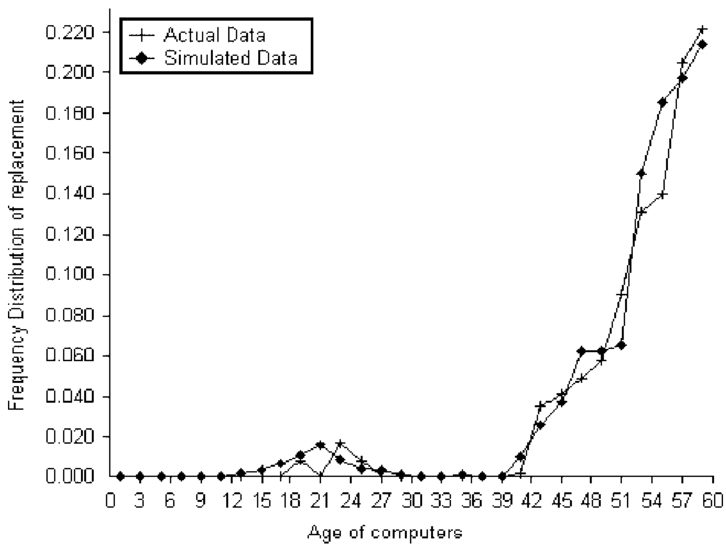

Figure 18. Comparison between simulated data and actual data in terms of frequency of replacement

rapidly with relatively small starting capacity and demand. Figure 23 shows the situation in which the cost per capacity decreases relatively slowly with a large demand for capacity.

The differences between Figures 22 and 23 are as follows. Figure 22 shows relatively higher likelihoods of keep and replacement than those of Figure 23. This is because small capacity requires relatively low maintenance costs. Also, as the computer gets older, replacement will be more profitable than upgrade, because of the relatively small cost per capacity. The situation is different in Figure 23. In the initial phase, keeping is the proper choice, but the likelihood of keeping is higher than that of Figure 22 since large capacity means there is no need for upgrade

DOI: 10.1002/jae

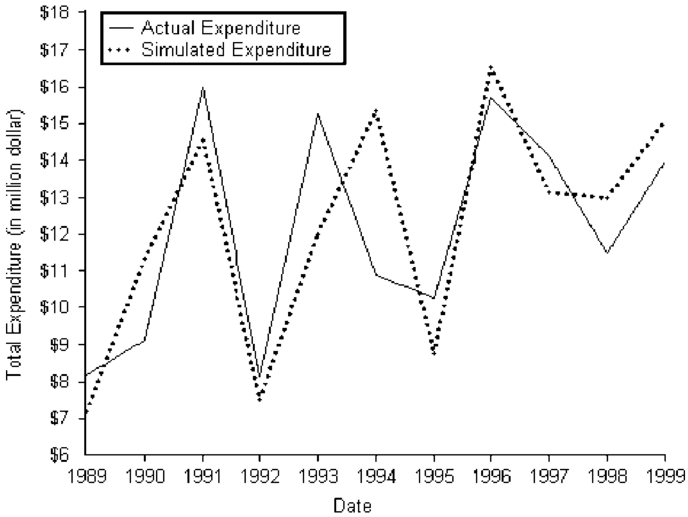

Figure 19. Comparison between actual total expenditure and simulated total expenditure

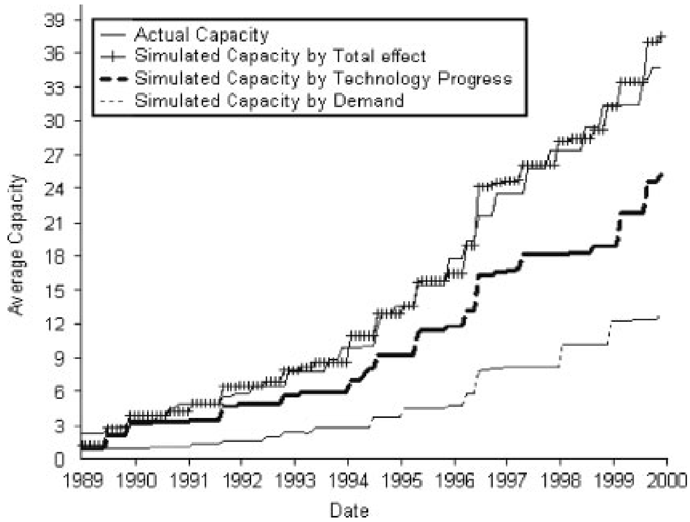

Figure 20. Comparison between actual capacity and simulated capacity (TPC case)

or replacement. Moreover, as shown in Figure 23, upgrade is more profitable than replacement over time, because of the relatively high cost per capacity.

As a result, the proposed model seems to explain the investment strategy of the firm regarding replacement and upgrade in a fairly reasonable way. Also, the simulations confirm that the firm follows an optimal investment strategy to replace and upgrade its computer systems by keeping track of the rapid development of computer technology and demand for its services.

6. CONCLUSION

In this study a proposed stochastic dynamic programming model was developed to test whether the explanatory facts of investment behavior could be rationalized as an optimal investment strategy

DOI: 10.1002/jae

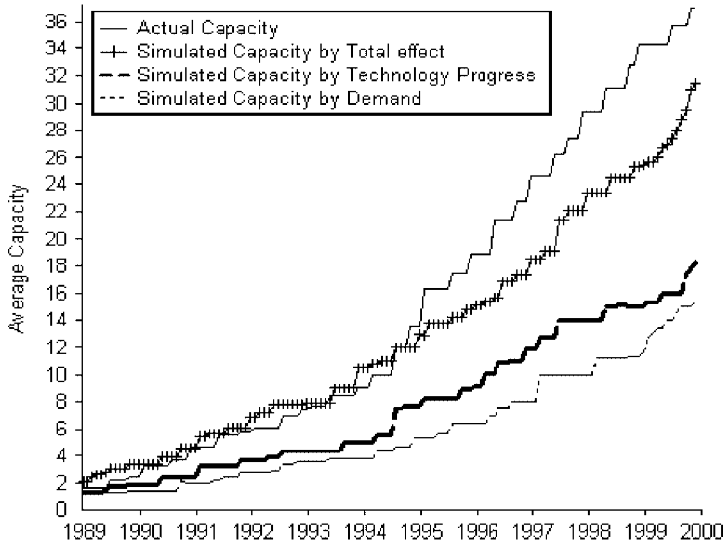

Figure 21. Comparison between actual capacity and simulated capacity (MIPS case)

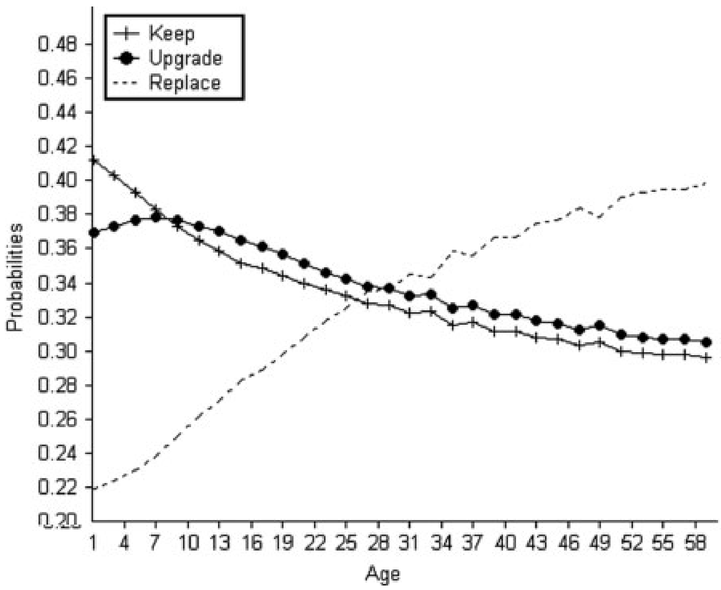

Figure 22. Simulated policy rules of keep, upgrade and replacement with rapidly declining cost per capacity in terms of relatively small starting demand

for the firm in question. In the model, the firm can choose among three main actions at each time period: to keep a computer system, to upgrade it, or to replace it. Contingent on the replacement decision, there are n sub-choices regarding capacity. The state variables include the processing capacity of the current system, the level of demand for this processing capacity, the age of the current system and the current market price of a standardized unit of processing capacity. The technological depreciation and the relative performance of each computer system are measured by composite measures of all four state variables in the model. The model depends on the unknown

DOI: 10.1002/jae

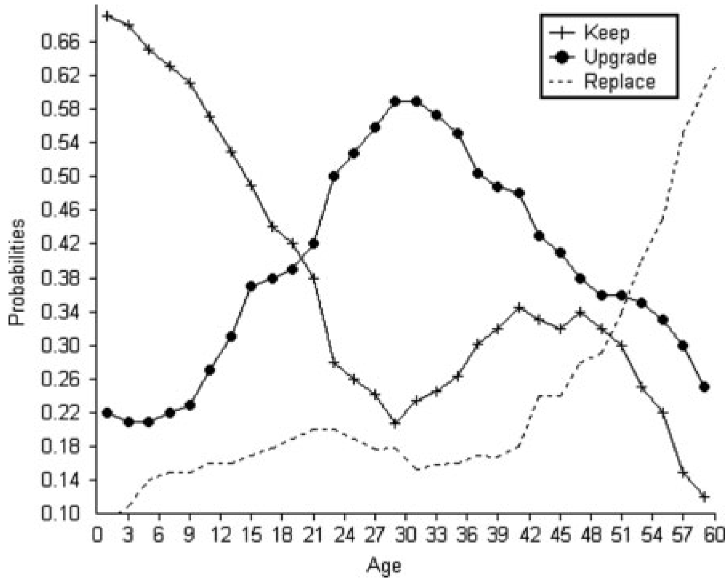

Figure 23. Simulated policy rules of keep, upgrade and replacement with slowly declining cost per capacity in terms of relatively large starting demand

primitive parameters that specify the firm's profit function and its expectation of future values of the state variables, with its expectation of future reductions in the price of computing capacity playing a critical role in the model's predictions of the optimal length of the replacement cycle.

I use a parametric approximation method to solve the DP problem. This greatly reduces the computational burden involved in solving the infinite-horizon version of the model where decisions are taken at monthly intervals and the two key state variables–current demand and the price per unit of new capacity–are allowed to assume a continuum of possible values. The parametric approximation procedure converts the contraction fixed-point problem into a nonlinear least squares problem. The acceleration in solution time is sufficiently large to make it feasible to estimate the unknown parameters of the model based on maximum likelihood.

The estimation results support the stylized facts observed from the data in general, allowing for a better understanding of the replacement behavior in the era of rapidly developing computer technology. In particular, the likelihood of an upgrade or replacement increases with the age of the current system and decreases with the current price of computing capacity. These results imply that the intervals between successive replacements or upgrades tend to shorten over time as the cost of computing decreases. The decision to keep or increase capacity through upgrade or replacement depends on the expectation of future demand and future cost per capacity. The simulated data based on the estimation results achieve the main objective of answering how well the proposed model answers the following question: 'How much of the 30-fold increase in mainframe processing capacity is due to the huge reduction in the cost of capacity, and how much is due to the growth in demand for capacity due to the comparably rapid growth in demand for services?' The model also confirms that the firm does not use an arbitrary rule of thumb in deciding to upgrade and replace its mainframe computers so rapidly. Rather, the firm appears to have a very sophisticated understanding of the impact of technological progress resulting from Moore's law and is taking advantage of this progress to significantly reduce its operating costs and provide better service to its customers.

DOI: 10.1002/jae

ACKNOWLEDGEMENTS

I thank the editor and Steven Berry and Martin Pesendorfer for their advice and encouragement. I am also indebted to the provider of the computer data, who wishes to remain anonymous. All errors are my own.

REFERENCES

Berndt E, Hausman J, Hall B, Hall R. 1976. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* **3**: 653–665.

Bertsekas D. 1976. *Dynamic Programming and Stochastic Control*. Academic Press: New York.

Bresnahan TF. 1986. Measuring the spillovers from technical advance: mainframe computers in financial services. *American Economic Review* **76**(4): 742–755.

Bresnahan TF, Greenstein S. 1997. Technical Progress in Computing and in the Uses of Computers. *Brookings Papers on Economic Activity, Micro*.

Bresnahan TF, Greenstein S. 1999. Technological competition and the structure of the computer industry. *Journal of Industrial Economics* **47**: 1–40.

Gandal N, Greenstein S, Salant D. 1999. Adoptions and orphans in the early microcomputer market. *Journal of Industrial Economics* **47**: 87–105..

Greenstein SM, Wade JB. 1997. *Dynamic modeling of the product life cycle in the commercial mainframe computer market, 1968–1982*. Working paper 6124, NBER.

Hendel I. 1999. Estimating multiple–discrete choice models: an application to computerization returns. *Review of Economic Studies* **66**: 423–446.

Hernandez-Lepka O, Lasserre JB. 1996. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer: New York.

Jorgenson DW. 1973. The economic theory of replacement and deprecation. In *Econometrics and Economic Theory*, Sellykaerts W (ed.). Macmillan: New York: 189–221.

Jorgenson DW, Stiroh KJ. 1999. Information technology and growth. *American Economic Review*, **89**(2): 109–115.

Kennet DM. 1994. A structural model of aircraft engine maintenance. *Journal of Applied Econometrics* **9**: 351–368.

McFadden D. 1973. Conditional logit analysis of qualitative choice behavior. In *Frontiers of Economics*, Zarembka P (ed.). Academic Press: New York; 105–142.

McFadden D. 1981. Econometric models of probabilistic choice. in *Structural Analysis of Discrete Data*, Manski C, McFadden D (eds). MIT Press: Cambridge, MA; 198–272.

Montemerlo MS, Love JC, Opiteck GJ, Goldhaber-Gordon D, Ellenbogen JC. 1996. Technologies and designs for electronic nanocomputers. In *MITRE MSR Program*, July 1996.

Perska-Juliussen K, Juliussen E. 1997. *8th Annual Computer Industry Almanac*. Computer Industry Almanac Inc.: Arlington Heights, IL.

Rust J. 1985. Stationary equilibrium in a market for durable assets. *Econometrica* **53**: 733–806.

Rust J. 1986. When is it optimal to kill off the market for used durable goods? *Econometrica* **54**: 65–86.

Rust J. 1987. Optimal replacement of GMC bus engine: an empirical model of Harold Zurcher. *Econometrica* **55**: 999–1033.

Rust J. 1988. Maximum likelihood estimation of discrete control processes. *SIAM Journal of Control and Optimization* **26**: 1006–1024.

Rust J. 1995. *Nested Fixed Point Algorithm Document Manual*, Vol. 5, November 1995. (http://gemini.econ.umd.edu/jrust/nfxp.html).

SIA Annual Databook. 1998. Semiconductor Industry Association: San Jose, CA.

SIA Annual Databook. 1999. Semiconductor Industry Association: San Jose, CA.