

1993

An econometric analysis of the desktop computer hardware market

Brian D. Lynch

University of Wisconsin- La Crosse

Frank M. Lin

California State University, San Bernardino

Winston T. Lin

State University of New York at Buffalo

H. Raghav Rao

State University of New York at Buffalo

Follow this and additional works at: <http://scholarworks.lib.csusb.edu/jiim>

 Part of the [Management Information Systems Commons](#)

Recommended Citation

Lynch, Brian D.; Lin, Frank M.; Lin, Winston T.; and Rao, H. Raghav (1993) "An econometric analysis of the desktop computer hardware market," *Journal of International Information Management*: Vol. 2: Iss. 2, Article 2.

Available at: <http://scholarworks.lib.csusb.edu/jiim/vol2/iss2/2>

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in Journal of International Information Management by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

An econometric analysis of the desktop computer hardware market

Brian D. Lynch

University of Wisconsin - La Crosse

Frank M. Lin

California State University at San Bernardino

Winston T. Lin

H. Raghav Rao

State University of New York at Buffalo

ABSTRACT

The purpose of this study was to compare six competing econometric models which depict the relationship between hardware characteristics and machine cost for the desktop computer market. The Box-Cox methodology and multiyear data were used to facilitate this comparison. The analysis validated that the Box-Cox methodology is a viable means for evaluating competing model formulations within the field of information systems research. The results were consistent with past research that suggested a double natural log model formulation for representing the functional relationships among variables when modeling machine cost as a function of hardware attributes. Further, the more complex power transformation model formulations suggested by the Box-Cox methodology did not significantly outperform the more traditional and simpler double natural log model. More specifically, the results indicated that variables related to primary memory and microchip technology have the largest impact on machine cost. Additionally, variables related to machine connectivity, machine expandability, and year of observation were also found to be significant explanatory variables for machine cost.

INTRODUCTION

Desktop computers are and will be having a significant impact on large U. S. corporations. Expenditures for desktop computers can become a major component of an organization's information systems budget, and hence the decision process regarding desktop acquisition is being taken seriously. For instance, studies of purchasing situations at large organizations (e.g., Fortune 500 firms), show that companies pay close attention to cost-benefit and pricing analyses when making hardware and software investment decisions (Reichert et al., 1988; Remmlinger and Waldmann, 1988). A well formulated, statistically significant computer attribute cost impact model could be used as a screening mechanism for identifying desktop alternatives that would be worthy of in-depth consideration by purchasers (e.g., screening out those machines identified as overpriced by the model, while giving a more detailed examination to those machines identified as underpriced by the model). The growing prevalence of downsizing (i.e., transferring work from larger machines to desktops) further highlights the importance of the desktop as a key corporate computer resource for the 1990s.

Historically, the terms microcomputer, desktop, and personal computer (PC) have been used interchangeably, because these types of machines were primarily designed to be used by one person at a time (Laudon & Laudon 1988; Long, 1989). However, recent technological developments have allowed "microcomputers" to be configured and marketed in different ways (e.g., as part of a larger network or as a laptop computer or as a workstation). The research presented here is aimed specifically at machines positioned by their vendors in the desktop segment of the market. A desktop computer is an effective computing resource that performs well as a word processor, spreadsheet, database and number-cruncher. The increasing importance of desktop machines is underscored by the fact that desktop unit sales went from 5.9 million in 1983 to 9.4 million in 1988 (Hillkirk, 1989).

Prior research related to the economic impact of computer characteristics has focused on price or cost, or price/MIPS, or some index of computer performance as the dependent variable. Independent variables typically focus on hardware characteristics (e.g., random access memory (RAM), direct access secondary storage (DASD), processing speed, etc). Other characteristics are sometimes examined, such as year of introduction, year of observation, IBM compatibility, etc. (Grosch, 1953; Cale et al., 1979; Ein-Dor, 1985; Kang et al., 1986; Mendelson, 1987; Ein-Dor and Feldmesser, 1987, 1989; Kang, 1989; Kang and Pick, 1989; Lynch et al., 1990; Dave & Fitzpatrick, 1991).

Traditionally, econometric analyses related to the impact of computer characteristics on computer cost (or price) or computer performance are usually based on linear or log model formulations with log formulations being the most common. The log approach represents Cobb-Douglas type formulations (Kang, 1989; Lynch et al., 1990). Kang (1989, p. 587) and Lynch et al. (1990, p. 121) provide detailed reviews of past research in this area. When prior research suggests that competing model formulations exist, Box and Cox's (1964) transformation methodology provides a proven rigorous statistical approach for comparing the suggested model formulations.

In the recent study by Lynch et al. (1990), the Box-Cox methodology was used to investigate the microcomputer market for one observation or data year - 1987. The essence of their study was the application of the Box-Cox methodology for choice of a model which best describes the relationships between attributes of microcomputers and their costs.

They utilized a special form of the Box-Cox methodology that power transforms both the dependent and independent variables by the same value. This special form of the Box-Cox methodology includes a double-logarithmic or Cobb-Douglas model and a linear model, as special cases. However, a more thorough examination of the potential model formulations would extend the analysis to include a generalized Box-Cox transformation form in which all variables (dependent and independent) in a model are subject to different power transformations, which includes the above-mentioned special form as a special case. Such an advance in methodology is now possible in view of the availability of computer programs to carry out the necessary empirical work.

This paper details an econometric investigation of the desktop market in this direction. Additionally, the econometric analysis of the relationship between desktop hardware characteristics and cost is extended to take into account multiple time horizons. As a result, the model considered herein accounts for advances in desktop technology due to time. The generalized Box-Cox methodology is used to identify and evaluate competing model

formulations. In particular, a key objective of the current research is to compare and contrast model formulations suggested by prior research (e.g., linear, double natural log, and Box-Cox power transformations).

The paper proceeds as follows. First is a brief overview of the generalized Box-Cox methodology. The next part describes the data and variable selection process. Model formulation and evaluation are discussed next. Then there is a detailed analysis of the results. Finally, there is a summary of the major findings and implications of the results and suggested areas for future research.

THE GENERALIZED BOX-COX METHODOLOGY

The Box-Cox method removes some of the subjectivity from the specification of the functional form (Spitzer, 1982). The strengths of the Box-Cox method were summarized by Spitzer (1978, p. 488):

- a. The transformations obtained are results of the estimation, not a priori subjectivity
- b. The transformations obtained include almost all of those commonly used by econometricians; and
- c. The estimation process itself generates a ranking value which can discriminate the effectiveness of alternative models.

Box and Cox (1964, p. 211) state that linear multiple regression is usually justified by assuming: (i) simplicity of structure for the dependent variable; " . . . (ii) constancy of error variance; (iii) normality of distributions; (iv) independence of observations." If any of the assumptions (i) - (iii) are violated, then nonlinear transformations of the data may improve matters. For example, Bozdogan and Ramirez (1986) utilize the Box-Cox methodology to transform multivariate data to "near" normality.

In their original paper, Box and Cox (1964) were mainly seeking power transformations for the response (i.e., dependent) variable. However, they also discussed simultaneous power transformations of dependent and independent variables. Box and Tidwell (1962) sought power transformations of the independent variables only. In practice, the Box-Cox and the Box-Tidwell transformations have become intertwined with most researchers simultaneously power transforming both the dependent and independent variables (Lin et al., 1992).

Researchers have come to view the simultaneous power transformations of the dependent and independent variables as variations of the Box-Cox methodology. For example, Spitzer (1982, p. 307) states that, "The power transformation introduced by Box and Cox (1964) and given by

$$(1) \quad Y^{(\lambda)} = (y^\lambda - 1)/\lambda \quad \lambda \neq 0$$

$$y^{(\lambda)} = \ln y \quad \lambda = 0$$

has been extensively used in recent years. Transformed variables can be included in a linear function so that generalized models of the form

$$(2) \quad y^{(\lambda)} = \beta_1 + \beta_2 X_2^{(\lambda_2)} + \dots + \beta_K X_K^{(\lambda_K)} + \varepsilon$$

can be specified and estimated. On occasion, neither a priori reasoning nor theory clearly dictates the functional form (transformation) which an additive model should assume." Spitzer (1982, p. 308) also argues that several different variations of Model (2) are possible:

$$(3) \quad y^{(\lambda)} = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon'$$

$$(4) \quad Y^{(\lambda)} = \beta'' + \beta_2'' X_2^{(\lambda)} + \dots + \beta_k'' X_k^{(\lambda)} + \epsilon''$$

$$(5) \quad y^{(\lambda_1)} = B_1''' + \beta_2'' X_2^{(\lambda_2)} + \dots + \beta_k'' X_k^{(\lambda_2)} + \epsilon'''$$

(Lin et al., 1992).

Like Spitzer (1982), we start with X_2 , as opposed to X_1 , to make the notation more manageable and easy to interpret. Each form puts certain restrictions on the power transformation of the variables. In model (3), only the values of the dependent variable are power transformed. In model (4), the dependent and all independent variables are transformed by the same value of λ . Model (5) transforms all independent variables by the same λ_2 value, while the dependent variable is transformed by the λ_1 value, different from λ_2 . Model (2), which is considered as the most general case, allows different power transformations on all variables (dependent and independent). Models (3), (4) and (5) are just special cases of Model (2). Model (4), the one used in Lynch et al., (1990), is probably the most widely used in practical and theoretical research applications (cf. Lin et al. 1992).

The λ parameters are estimated by the maximum likelihood method, a procedure used in, for example, Ehrlich, 1977; Levanbach and Cleary, 1984; and Lynch et al., 1990. The maximum likelihood value, L_{\max} , is calculated for different values of the λ parameters per the following function:

$$(6) \quad L_{\max} = - (n/2) \ln (RSS/n) + (\lambda - 1) \sum \ln(y),$$

where λ denotes the power transformation for the dependent variable, n is the total number of observations, and RSS denotes the sum of the squared residuals. Typical ranges of λ values are -2 to +2, or -1.5 to +1.5 (Levanbach and Cleary, 1984). The λ value [or set of λ values when multiple λ s are estimated as in Models (2), (4) and (5)] that maximizes the L_{\max} function is the λ (or set of λ s) used to transform the data. The λ value that maximizes the L_{\max} function is referred to as λ^* . When multiple λ s are estimated, each estimated λ would have its own λ^* value.

DATA AND VARIABLE SELECTION

Computerworld has become one of the standard data sources for research related to the econometric modeling of computer hardware characteristics. The data utilized in this study was gathered from this source. Specifically, the data were collected from articles summarizing the results of surveys of desktop hardware vendors that were conducted by *Computerworld* in 1987, 1988 and 1989 (Xenakis, 1987; Ryan, 1988; Xenakis, 1989). Only observations of IBM or IBM compatible machines were used in this study, including IBM's Personal System Model line. IBM compatibility has become the standard for desktops used in business (*Purchasing*, 1986). Lynch et al. (1990), in their study of the microcomputer market, also focused exclusively on IBM and IBM compatible machines. Even Apple Computer has been aware of its need to address the IBM compatibility issue for long-term viability (Meth, 1987). If Apple (or any vendor) wants to maintain or increase market share, their machines must address the issue of IBM compatibility, or more specifically, compatibility with the market.

A literature search was conducted to identify significant research related to the econometric modeling of desktop hardware characteristics. The work of Dave and Fitzpatrick (1991), Lynch et al. (1990), and an earlier benchmarking study by Sircar and Dave (1986) were the key studies identified. The General Accounting Office (Lewis and Crews, 1985) noted that benchmarking is no longer considered a cost-effective mechanism to evaluate machines. The relative absence of significant prior research necessitates the careful identification, consideration and statistical testing of potential variables for inclusion in any model formulation. The variables identified by Dave and Fitzpatrick (1991) and by Lynch et al. (1990) provide some guidance.

The variables included in Dave and Fitzpatrick's (1991) study are described in Table 1a, while Lynch et al.'s (1990) variables are described in Table 1b. Lynch et al. analyzed only data related to the 1987 observation year. Variables identified by researchers modeling the hardware characteristics of larger machines also offer some guidance (Grosch, 1953; Cale et al., 1979; Ein-Dor, 1985; Kang et al., 1986; Mendelson, 1987; Ein-Dor & Feldmesser, 1987 & 1989; Kang, 1989; Kang and Pick, 1989). Variables used in previous research (e.g., chip, serial ports, expansion slots, RAM, DASD, different formulations of speed, etc.) were closely scrutinized for their impact on machine cost. Various model formulations were examined. For example, various linear and double natural log regression models were run. Additionally, Pearson correlations and factor analyses were used to identify possible interaction terms and multicollinearity problems. Finally, nine independent variables that were consistently highly significant were used for a more in-depth statistical analysis of the data utilizing the generalized Box-Cox methodology.

Table 1a. Description of variables used by Dave and Fitzpatrick (1991)

<u>Variable Identifier</u>	<u>Description</u>
PRICE	Machine price in U. S. dollars (dependent variable)
CLOCK	Clock speed in megahertz
RAM	Random access memory (primary memory) in megabytes
HARDDISK	1 if hard disk is included with system, 0 otherwise
DY ₁	1 if year of introduction was 1988, 0 otherwise
DY ₂	1 if year of introduction was 1989, 0 otherwise
DC ₁	1 if machine contains 80286 chip, 0 otherwise
DC ₂	1 if machine contains 80386 or 80386SX chip, 0 otherwise
MAX	1 if configuration represents the maximum system configuration, 0 otherwise

NOTES:

1. MAX was used to differentiate between the minimum system configuration (i.e., MAX = 0) and the maximum system configuration (i.e., MAX = 1) for machines that offered a range in price and configurations.
2. Their data source was *PCWeek*.

Table 1b. Description of Variables Used by Lynch, Rao, and Lin (1990)

<u>Variable Identifier</u>	<u>Description</u>
COST	Machine price in U. S. dollars (dependent variable)
ATXT ¹	1 if machine is classified as AT machine, 0 otherwise
C80386SX ²	1 if C80386SX chip is present, 0 otherwise
DUMEXP	1 if machine comes with 8 expansion slots, 0 otherwise
SPMIN	Minimum machine speed setting in megahertz
SERPORTS	Number of serial ports that come with the machine

NOTES:

1. Machines were classified as an AT or an XT type machine. Thus, if ATXT equal zero, then the machine is an XT machine. More specifically, AT machines primarily used the C80286 chip, while XT machines primarily used the C8088 chip.
2. Lynch et al. (1990) stated that their model contained a dummy variable (i.e., C80386) that captured the presence or absence of the C80386 chip. In actuality, their dummy variable measured the presence or absence of the C80386SX chip. Undoubtedly, this specification error was due to the fact that the 1987 *Computerworld* survey did not need to differentiate between the C80386 and C80386SX (i.e., only C80386SX chips were present). Thus, the C80386SX chips were incorrectly identified as C80386 chips. In 1988 and 1989, as machines started using both the C80386 and C80386SX chips, *Computerworld's* surveys differentiated between the C80386 and C80386SX chips.

The data set consists of 707 observations with 335, 194, and 178 observations from years 1987, 1988, and 1989, respectively. Table 2 describes the ten variables (i.e., nine independent and one dependent) used in the analysis presented here. Table 3 shows the means and standard deviations for the variables for the full data set and for the subsets of the full data set related to each observation year.

Table 2. Description of Variables

<u>Variable Identifier</u>	<u>Description</u>
C80386	1 if the C80386 chip is present, 0 otherwise
C80386SX	1 if the C80386SX chip is present, 0 otherwise
C80286	1 if the C80286 chip is present, 0 otherwise
C8088	1 if the C8088 chip is present, 0 otherwise
RAMLOW	Minimum random access memory (RAM) in kilobytes that comes with the computer
EXPSLOTS	Number of expansion slots that come with the computer
SERPORTS	Number of serial ports that come with the computer
YEAR88	1 if the observation is from 1988, 0 otherwise
YEAR89	1 if the observation is from 1989, 0 otherwise
COST	Price of computer in U. S. dollars (dependent variable)

Table 3. Means and Standard Deviations for Variables

	Full Data Set		1987 Observations		1988 Observations		1989 Observations	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
C80386	.167	.373	.000	.000	.258	.439	.382	.487
C80386SX	.085	.279	.090	.286	.026	.159	.140	.348
C80286	.481	.500	.510	.501	.479	.501	.427	.496
C8088	.175	.381	.296	.457	.113	.318	.017	.129
RAMLOW	839.065	490.757	640.955	233.956	906.201	443.602	1138.742	687.935
EXPSLOTS	6.424	2.405	7.260	1.867	5.552	2.773	5.802	2.360
SERPORTS	1.149	1.007	.943	.991	1.351	1.292	1.315	.466
YEAR88	.274	.447	.000	.000	1.000	.000	.000	.000
YEAR89	.252	.434	.000	.000	.000	.000	1.000	.000
COST	3010.581	2381.405	1749.033	1136.750	4142.902	2712.760	4150.742	2536.302
	N = 707		N = 335		N = 194		N = 178	

Microcomputer chips represent the evolution of the central processing unit (CPU) technology for desktops. Four dummy variables were used to measure the CPU or "chip" that comes with a machine. These variables are labeled as C80386, C80386SX, C80286, and C8088. The C80386 is the most technologically advanced, followed by the C80386SX, C80286, and C8088, respectively. The chips differ in terms of the byte size used for processing and communication purposes. The C80386 and C80386SX chips are both 32-bit chips. However, the C80386 is designed to communicate with a 32-bit communication bus, while the C80386SX is designed to communicate with a 16-bit communication bus. The C80386SX was developed to accommodate the 16-bit communication bus architecture which existed at the time of its development. The C80286 is a 16-bit chip and the C8088 is an 8-bit chip.

Table 3 shows that the C80386 chip was not present in any of the 1987 observations, while the C80386SX was present in 30 machines (i.e., .09 * 335) or nearly 10% of the 1987 observations. By 1989, the C80386 or C80386SX were present in over 52% (i.e., 38.2 + 14.0) of the observations. Contrarily, the C8088 was present in 29.6% or 99 of the 1987 observations, while it was only present in 1.7% or 3 of the 1989 observations. The prevalence of the C80286 also declined during the 1987 to 1989 observation period, but not nearly as drastically as the C8088. The C80286 went from being present in 51.0% or 171 of the 1987 observations to 42.7% or 76 of the 1989 observations.

The RAMLOW variable measures the minimum amount of RAM in kilobytes that comes with a machine. RAM is used by a machine to execute programs. Table 3 shows that the mean RAMLOW increased from 839 kilobytes to 1139 kilobytes from 1987 to 1989. This makes intuitive sense, since the technological advancement of the CPUs would make it practical to process larger and more complex programs which would require more RAM to operate efficiently.

The EXPSLOTS variable measures the number of expansion slots that come with a machine. The mean for EXPSLOTS is 7.260, 5.552, and 5.802 for years 1987, 1988, and 1989, respectively. The increasing use of local area networks (LANs) is one possible explanation for the decline in the mean number of expansion slots. LANs enable machines to share resources (e.g., printers, etc.). Thus, the need to enhance an individual machine by inserting an expansion board into an expansion slot would be decreased.

The SERPORTS variable measures the number of serial ports that come with a machine. The mean for SERPORTS is .943, 1.351, and 1.315 for years 1987, 1988, and 1989, respectively. Two possible reasons for the increase in the mean number of serial ports are the increasing prevalence of "mouses" and modems. A mouse is a hand operated man-machine interface device. Mouses are attached to machines via a serial port. Modems enable a machine to connect with other computers using available phone lines. External modems are typically attached to a machine using a serial port. On the other hand, internal modems come on integrated circuit boards which are inserted into an expansion slot within a machine.

Kang (1989) used year of observation (i.e., year of measurement) to control for price changes over the years and year of introduction to control for the effect of machine age on computer price, while Dave and Fitzpatrick (1991) used only year of introduction to control for the effect of machine age on price. Year of introduction was not reported in the *Computerworld* data being analyzed here. Given the high number of vendors surveyed, it was determined that it would be too costly to contact them to ascertain year of introduction information for each machine. Further, vendors frequently reconfigure their older machines in order to be competitive with the market. Thus, over time, machines using older chip technology would probably increase the absolute amount of other machine attributes (e.g., DASD memory) and/or increase the level of technology represented by other attributes (e.g., a more advanced monitor along with the necessary software/hardware to take advantage of the monitor's capabilities). Thus, the analysis conducted here focuses on year of observation only and ignores the impact of year of machine introduction. Year of observation is captured by the dummy variables YEAR88 and YEAR89. YEAR88 equals one if the observation is from the 1988 survey, and equals zero otherwise. YEAR 89 equals one if the observation is from the 1989 survey, and equals zero otherwise. If YEAR 88 and YEAR 89 both equal zero, then the observation is from the 1987 survey.

Finally, the COST variable measures the cost of a machine. List price is used as a surrogate measure (Kang, 1989; Lynch et al., 1990) to measure machine COST. The price range, from minimum to maximum price as published by *Computerworld*, represents the low-end and high-end configurations for a given model. The COST analysis is based on the mean price (i.e., mid-point of the price range) of each machine (Kang, 1989; Lynch et al., 1990). The mean prices are \$1,749, \$4,143, and \$4,150 for years 1987, 1988, and 1989, respectively. The increase in mean price after 1987 probably reflects the shift to C80386 chip technology (and away from C8088) and an increase in other system attributes (e.g., RAM and SERPORTS) to better leverage the capabilities of the newer chip.

MODEL FORMULATION AND EVALUATION

Once the issue of variable selection was resolved satisfactorily, six step-wise multiple regression models were evaluated to ascertain the most appropriate data transformation. These included a linear model of the untransformed data, a double natural log transformation, and four models conforming to Box-Cox transformation equations (2)-(5). The linear and double natural log models represent special cases of Box-Cox transformation (2). The results of these analyses are summarized in Tables 4, 5, and 6.

Table 4 presents the six models with their associated coefficients. As noted in Table 4, Model I is the linear model; Model II is the double natural log model; and Models III-VI are the transformed models as per the Box-Cox method. For Models II-VI, the six dummy variables (i.e., C80386, C80386SX, C80286, C8088, YEAR88, YEAR89) are not transformed. Since the relative range (i.e., 0 or 1) of these variables are not very great, transformation would not have a big effect on the linearity of the regression (Box & Cox, 1964). This approach is also similar to Dave and Fitzpatrick (1991). They used a double natural log model formulation where dummy variables were not logged. Model VI, which is the most general case of the Box-Cox transformations, allows for different power transformations for the dependent and three of the independent variables (i.e., the non-dummy independent variables) in the model. For Models III-VI, λ values ranging from -1.5 to +1.5 were examined to identify the appropriate λ 's. The LIMDEP computer program was used to identify the λ 's (Greene, 1989).

Table 4. The Models

MODEL I - THE LINEAR MODEL

$$\text{COST} = 1505.15 + 3777.49 \text{ C80386} + 1.18 \text{ RAMLOW} - 175.73 \text{ EXPSLOTS} \\ - 277.38 \text{ C8088} + 790.30 \text{ YEAR88} + 1919.57 \text{ C80386SX} \\ + 276.86 \text{ SERPORTS} + 804.75 \text{ C80286} - 96.48 \text{ YEAR89}$$

MODEL II - THE DOUBLE NATURAL LOG MODEL

$$\text{LN(COST)} = 6.028 + 0.996 \text{ C80386} - 0.311 \text{ C8088} + 0.025 \text{ LN(SERPORTS)} \\ + 0.842 \text{ C80386SX} + 0.440 \text{ YEAR88} + 0.335 \text{ C80286} \\ + 0.194 \text{ LN(RAMLOW)} + 0.201 \text{ YEAR89} \\ - 0.023 \text{ LN(EXPSLOTS)}$$

MODEL III - THE BOX-COX MODEL ($\lambda^* = 0.04$)

$$\frac{(\text{COST}^{0.04} - 1)}{0.04} = 8.3175 + 1.4208 \text{ C80386} - 0.4108 \text{ C8088} + 0.5791 \text{ YEAR88} \\ + 1.1500 \text{ C80386SX} + 0.2214 \text{ YEAR89} + 0.0003 \text{ RAMLOW} \\ + 0.1295 \text{ SERPORTS} + 0.5038 \text{ C80286} - 0.0602 \text{ EXPSLOTS}$$

MODEL IV - THE BOX-COX MODEL ($\lambda^* = 0.09$)

$$\frac{(\text{COST}^{0.09} - 1)}{0.09} = 8.2284 + 2.0546 \text{ C80386} - 0.5688 \text{ C8088} + 0.0581 \frac{(\text{SERPORTS}^{0.09} - 1)}{0.09} \\ + 1.6771 \text{ C80386SX} + 0.8576 \text{ YR8088} + 0.2401 \frac{(\text{RAMLOW}^{0.09} - 1)}{0.09} \\ + 0.6674 \text{ C286} + 0.3699 \text{ YR89} - 0.0651 \frac{(\text{EXPSLOTS}^{0.09} - 1)}{0.09}$$

Table 4 (cont'd)

 MODEL V - THE BOX-COX MODEL ($\lambda_1^* = 0.01, \lambda_2^* = 0.55$)

$$\begin{aligned} \frac{(\text{COST}^{0.01} - 1)}{0.01} &= 7.3177 + 0.0063 \frac{(\text{RAMLOW}^{0.55} - 1)}{0.55} - 0.3090 \text{C8088} + 1.1000 \text{C80386} \\ &+ 0.4272\text{C88} + 0.9137 \text{C80386SX} + 0.1432 \frac{(\text{SERPORTS}^{0.55} - 1)}{0.55} \\ &+ 0.3976 \text{C80286} - 0.0822 \frac{(\text{EXPSLOTS}^{0.55} - 1)}{0.55} \end{aligned}$$

 MODEL VI - THE BOX-COX MODEL ($\lambda_1^* = 0.01, \lambda_2^* = 0.39, \lambda_3^* = 0.97, \lambda_4^* = 0.70$)

$$\begin{aligned} \frac{(\text{COST}^{0.01} - 1)}{0.01} &= 7.5401 + 1.1025 \text{C80386} - 0.3210 \text{C8088} + 0.1225 \frac{(\text{SERPORTS}^{0.39} - 1)}{0.39} \\ &+ 0.9225\text{C80386SX} + 0.4173 \text{YEAR88} + 0.0003 \frac{(\text{RAMLOW}^{0.97} - 1)}{0.97} \\ &+ 0.4018 \text{C80286} - 0.0677 \frac{(\text{EXPSLOTS}^{0.70} - 1)}{0.70} + 0.1321\text{YEAR89} \end{aligned}$$

Bartlett's test for heteroscedasticity was performed for all six models (Pindyck and Rubinfeld, 1981, pp. 147-148). This analysis showed that heteroscedasticity was present at the .01 percent significance level for the linear model, but was not a serious problem for the other models. This further supports the use of data transformation.

Table 5 presents the t-statistics for the variables for each model, the coefficients of determination (R^2), F-values, and the L_{\max} values for all six models. Given that the λ s for cost for Models II-VI are natural log or near natural log transformations (i.e., λ s ranging from .01 to .09), it is appropriate to compare these models based on R^2 values, F-statistics, t-statistics, and L_{\max} values. However, given that the transformation of the cost (dependent) variable for Model I (the linear model) differs significantly from natural log, caution should be used when comparing Model I to the other five models based on these statistical criteria.

Model VI has the highest R^2 value, F-value, and L_{\max} value, while the linear model (Model I) has the lowest R^2 value, F-value, and L_{\max} value. The other four models fall somewhere in between the above two extremes. The F-value and t-statistics for the individual variables for Models II-V are all significant at the 0.01 level. A comparison of all six models in terms of F-value, R^2 , and L_{\max} indicates that Model VI is marginally better than Models II-V and clearly superior to Model I.

Table 5. T-Statistics, R^2 , F-Values, and L_{\max} for Models I, II, III, IV, V, and VI

Variable	Models (T-Statistics)					
	I	II	III	IV	V	VI
C80386	15.024	14.593	14.944	14.879	14.706	14.819
C8088	-1.272	-5.020	-4.982	-4.569	-4.752	-4.972
SERPORTS	5.035	7.318	6.229	6.955	7.472	7.689
C80386SX	7.424	11.516	11.784	11.395	11.840	12.058
YEAR88	5.296	10.657	10.265	10.306	9.583	9.307
RAMLOW	8.448	5.282	6.516	5.757	6.264	6.534
C80286	4.185	6.215	6.929	6.143	6.940	7.047
EXPSLOTS	-7.081	-3.451	-6.414	-4.061	-6.123	-6.022
YEAR89	-0.588	4.467	3.567	4.088	3.179	2.697
R^2	0.6493	0.7103	0.7193	0.7111	0.7250	0.7267
F	142.11	189.89	198.45	190.62	204.15	205.89
L_{\max}	-5,128.8	-4,821.9	-4,811.2	-4,823.2	-4,803.5	-4,801.3

Three other measures frequently used to aid in the model selection and verification process are the mean square error (MSE), the mean absolute deviation (MAD), and the mean absolute percentage error (MAPE) (Lin, 1989). The MSE measure penalizes a model more for large errors than for small ones. MAD does not have this problem. MAPE relates the magnitude of the forecasting error to the actual value.

Table 6 presents the performance of all six models on these three measures. The linear model performs the worst on the MAD and MAPE measures although it performs the best in terms of the MSE measure. Model II, the double natural log model, performs the worst on the MSE measure. Model V, with $\lambda_1^* = 0.01$ for the dependent variable and $\lambda_2^* = 0.55$ for the three non-dummy independent variables, performs the best on the MAD and MAPE measures. At the same time, it performs the second best based on the MSE measure. Model VI, the most generalized Box-Cox transformed model, with different power transformation values for the dependent variable and each of the three non-dummy independent variables, performs better than all the other models, except Model V, on the MAD and MAPE measures and it falls in the middle in terms of the MSE measure. The standard deviation for cost is \$2,381 which is approximately 79 percent of the mean machine cost of \$3,010 (see Table 3). MAPEs around 32 to 34 percent and MADs around \$910 to \$940 represent significant improvements in contrast to the large standard deviation. As shown in Table 6, in general, the power transformed models (i.e., models II-VI) perform better than the linear model based on the MAD and MAPE measures. The implications of the empirical results will be discussed in the next section.

Table 6. MAPE, MAD, and MSE for Models I-VI

	MAPE (%)	MAD (\$)	MSE (\$ ²)
<u>Linear and Log Models:</u>			
I. Linear	38.73	939.61	1,997,593
II. Double Log	33.61	938.33	2,287,823
<u>Box-Cox Models:</u>			
III. $\lambda^* = .04$	32.89	915.14	2,115,207
IV. $\lambda^* = .09$	33.72	936.36	2,214,260
V. $\lambda_1^* = .01, \lambda_2^* = .55$	32.63	908.22	2,087,092
VI. $\lambda_1^* = .01, \lambda_2^* = .39, \lambda_3^* = .97, \lambda_4^* = .70$	32.65	913.35	2,117,136

Note: MAPE = mean absolute percentage error
MAD = mean absolute deviation
MSE = mean squared error

ANALYSIS AND IMPLICATIONS

As noted previously, a well formulated, statistically significant computer attribute cost impact model could be used as a screening mechanism for identifying desktop alternatives that would be worthy of in-depth consideration by purchasers. For example, decision makers could eliminate from consideration machines identified as overpriced by the model, while giving a more detailed examination to machines identified as underpriced by the model.

The models identify nine variables that have a highly significant impact on machine cost for the desktop market. The statistics (t-values, F-values, coefficients of determination, and the L_{max} values) of the data transformed models (i.e., Models II-VI) outperform those of the linear model. Given that the double natural log model (i.e., Model II) has been identified by prior researchers (Dave and Fitzpatrick, 1991; Lynch et al., 1990) as applicable to the desktop or microcomputer market and that Models II through VI are clustered reasonably close together in terms of their statistical results (e.g., R^2 , L_{max} , F-values, and t-statistics), the double natural log model will be used for purposes of discussion of results. Further, when results are similar, simpler model formulations (i.e., Model II) are preferred to more complex model formulations (i.e., Models III-VI).

Table 7 presents the effect by variable for the double natural log model (i.e., Model II). The variables, in order from largest to smallest effect, are RAMLOW (approximately 1.253 to 1.345 for values near the mean for RAMLOW), C80386 (.996), C80386SX (.842), YEAR 88 (.440), C80286 (.335), C8088 (-.311), YEAR89 (.201), and EXPSLOTS (-.041 for a value near the mean for EXPSLOTS), and SERPORTS (.000 for a value near the mean for SERPORTS). In terms of effect on machine cost, four out of the first six variables are related to computer chips. This implies that computer chips play a very important role in determining the cost of a given desktop. From a rank order perspective, the cost effects are in exact order from the newest to the oldest chip. The C8088 was developed first followed by the C80286, C80386SX, and C80386, respectively. In fact, it is very logical to expect that the more powerful the chip, the more expensive the desktop computer. More specifically, all the chip variables have positive effects on cost except for the C8088.

Table 7. Analysis of Effect by Variable for Model II

Variable Identifier	Coeff.	Variable Values	LN (Variable)	Effect
C80386	0.996	0		0.000
		1		0.966
C8088	-0.311	0		0.000
		1		-0.311
SERPORTS	0.025	0.000001	-13.816	-0.345
		1	0.000	0.000
		2	0.693	0.017
		3	1.099	0.027
C80386SX	0.842	0		0.000
		1		0.842
YEAR88	0.440	0		0.000
		1		0.440
C80286	0.335	0		0.000
		1		0.335
RAMLOW (Kbytes)	0.194	512	6.238	1.210
		640	6.461	1.253
		1024	6.931	1.345
		2048	7.625	1.479
		4096	8.318	1.614
		8192	9.011	1.748
YEAR89	0.201	0		0.000
		1		0.201
EXPSLOTS	-0.023	0.000001	-13.816	0.318
		1	0.000	0.000
		2	0.693	-0.016
		3	1.099	-0.025
		6	1.792	-0.041
		8	2.079	-0.048

NOTE: For natural log transformations zero values were approximated with a value close to zero (i.e., 0.000001) as suggested by Kang (1989), Kang and Pick (1989), and Lynch et al. (1990).

It is important to note that the negative effect for C8088 implies that the machine cost is less if the C8088 chip is present. This makes intuitive sense, since the C8088 represents a very inferior CPU in comparison to the other chips. This is further verified by its near obsolescence by 1989 (i.e., it is present in only 1.7% of the 1989 observations). More generally, we would expect the cost effects of any chip to initially be significantly positive (i.e., when

it represents state-of-the-art technology — e.g., the C80386 and C80386SX). Then, over time, its cost effect would decline (e.g., the C80286) and eventually become negative (e.g., the C8088). A negative cost effect would indicate that a chip was at the end of its product life cycle — i.e., no longer a viable option for vendors configuring new machines for the market place. Thus, if future analyses were conducted on new data in conjunction with the data used here, the C80386, C80386SX, and C80286 chips would all decline in terms of cost effect and eventually become negative and obsolete like the C8088 is presently.

The RAMLOW (minimum random access memory in kilobytes that comes with the computer) is the most important variable in the model. The most popular configurations of RAMLOW are 512, 640, 1012, 2048, 4096, and 8192 kilobytes. It is interesting to note that RAMLOW is increasing from 1987 to 1989 (see Table 2). As noted earlier, more powerful and sophisticated software requires an increase in RAM capabilities in conjunction with an increase in CPU capabilities. This makes economic sense. Typically, one would expect simultaneous increases in inputs to yield greater marginal utility in terms of output, as opposed to drastically increasing only one input.

The dummy variable YEAR88 has the second largest effect on cost of the non-chip variables. This implies that, in general, a machine is more expensive in 1988 than in 1987. As noted earlier, this effect may result from the need of vendors to increase the capabilities of their older models (i.e., change the machines' configurations) to be more competitive with machines based on the state-of-the-art chip technology. Additionally, machines based on the state-of-the-art chip technology usually come equipped with more advanced technology (e.g., super VGA monitor, larger capacity hard disk, etc.) which could tend to further increase machine COST. This rationale probably accounts for the significant increase in machine COST from 1987 to 1988.

On the other hand, YEAR89 has a significantly smaller positive effect on machine COST than YEAR88. This implies that given the same machine, it is cheaper in the year 1989 as compared to the year 1988. The explanation for this is probably the maturing of the desktop market (particularly with respect to C80386 chip technology) and the increasing use of price cuts to drive sales. Competition based on machine price was even more prevalent in 1991 than in the data analyzed here (Fitzgerald, 1991; Quinlan, 1991).

The negative impact on machine cost for EXPSLOTS is consistent with the finding by Lynch et al. (1990). They argued that the negative cost effect of expansion slots was due to non-standard expansion slot configurations. In 1987 (their data year), the standard configuration was eight expansion slots. Thus, desktops with a nonstandard configuration of expansion slots (e.g., less than eight) would cost significantly more than those with the standard configuration of eight expansion slots. However, the mean number of expansion slots is decreasing from 1987 to 1989 (see Table 3). This implies that the standard number of expansion slots is decreasing. For example, IBM's Personal System Model line comes with three expansion slots. This decrease could be attributed to the evolution of LANs. With LANs, many machines can be linked together and they can share the same resources. Thus, the need to enhance an individual machine by inserting expansion boards into expansion slots is not as great as in the past. Two primary implications may be drawn: (i) desktops with standard configuration are cheaper than those with nonstandard configurations; and (ii) machines based on older standard configurations (e.g., 8 expansion slots) cost less than machines with state-of-the-art expansion slot configurations.

The final variable in the model, SERPORTS (number of serial ports that come with the machine), has the lowest impact on machine COST. The mean number of serial ports has increased from 1987 to 1989. As noted earlier, the increasing use of mice and modems may explain the increase in the mean of SERPORTS.

In summary, machine memory (i.e., RAMLOW), the chip variables (i.e., C80386, C80386SX, C80286, and C8088), the passage of time (i.e., YEAR88 and YEAR89), and variables related to connectivity and expandability (i.e., SERPORTS and EXPSLOTS) were shown to have significant effects on machine cost. The findings are consistent with past literature which shows that variables that differentiate machine types (e.g., the chip variables) have a large impact on machine cost or price (Dave and Fitzpatrick, 1991; Lynch, et al., 1990). In essence, chip technology is rapidly becoming the primary determinant of machine cost. This is particularly true if the chip technology in question represents the state-of-the-art chip technology for the data under analysis. The cost effect of EXPSLOTS was found to be negative which confirms the finding of Lynch et al. (1990). The discussion related to the YEAR88 and the YEAR89 variables showed that the increase in machine COST due to the passage of time is slowing. For example, compared to YEAR88, the impact of the YEAR89 variable on the cost of a desktop is significantly less. It is logical to deduce from the analysis that technological change is relatively rapid in the desktop industry, especially in terms of the chip segment. In fact, chipmakers believe that they can continue to double chip power every two years, as they have done throughout the 1980s. If the chipmakers are right, desktops of the late 1990s will have power surpassing today's supercomputers (Depke and Brandt, 1991).

CONCLUSIONS

The results of this research indicate that the minimum amount of RAM memory and the microchip provided with the machine have the largest impact on machine cost for desktop computers. The finding related to primary memory size (RAM) is consistent with the findings of prior studies (e.g., Cale et al., 1979; Sircar and Dave, 1986; Dave & Fitzpatrick, 1991), while the finding related to chip type supports the results of previous research (e.g., Dave & Fitzpatrick, 1991; Lynch et al., 1990). Other important cost impact variables identified by this study include the year of observation (i.e., measurement), and variables associated with connectivity and expandability. An interesting result is that direct access storage devices were not found to be significant. This contrast with past research related to larger machines that have typically identified secondary storage memory as an important determinant of machine cost.

A fundamental contribution of this paper is the application of the generalized Box-Cox methodology to differentiate between different functional forms. The generalized Box-Cox transformation process is a rigorous methodological approach which is especially useful for comparing and contrasting alternative model formulations. The results of this analysis support the appropriateness of the double natural log formulations used by past researchers (e.g., Dave & Fitzpatrick, 1991; Lynch et al., 1990). The results suggest that the Box-Cox methodology is a viable approach to a more realistic and less subjective study of economic phenomena in the information systems field. It is our belief that utilizing more rigorous statistical methodologies (e.g., the Box-Cox methodology) will advance the field of information systems research. A limitation of the use of this methodology is that it is extremely

computation intensive. However, the availability of appropriate computer programs (e.g., LIMDEP) has greatly reduced the burden of computation and computational cost.

A key limitation of this study is that (due to the problem of data availability) some variables not captured in the data set (such as marketing promotions, economies of scale due to unit sales volume, and the cost of labor) may significantly affect the cost of desktops. In particular, a more detailed model of the competitive market place, which captures how equilibrium prices for desktops change as a function of competitor actions and market responses, could help managers discern the optimal timing of their desktop purchases particularly with respect to chip product life cycle effects on desktop pricing decisions by vendors.

In summary, the double natural log model (i.e., Model II) was identified as an appropriate model formulation for the desktop hardware market and accounted for approximately 71 percent of a desktop's cost. Further, state-of-the-art computer chips were shown to play a central role in determining desktop cost. It is expected that chip technology will continue to change rapidly. A logical implication for decision makers is that an effective information systems planning function is necessary to ensure that the firm can anticipate (rather than respond) to technological changes. For instance, due to the increasing power of desktops, decision makers need to more closely scrutinize the possibility of the downsizing of the hardware infrastructure (i.e., transferring work from the larger machines to desktops) and the impact of downsizing on the organization.

ACKNOWLEDGEMENTS

The authors wish to thank Michael Cheng for his assistance with the use of the LIMDEP computer program, and Wonsu Hu and Eugene Johnson for providing research assistance.

REFERENCES

- Are personal computers a commodity market? (1986, July 24). *Purchasing* 101, 2, 78-95.
- Box, G. E. P. & Cox, D. R. (1964, April). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26(2), 211-243.
- Box, G. E. P. & Tidwell, P. W. (1962, November). Transformation of the independent variables. *Technometrics*, 4, 531-550.
- Bozdogan, H. & Ramirez, D. E. (1986). Testing for model fit: Assessing and Box-Cox transformations of multivariate data to 'near' normality. *Computational Statistics Quarterly*, 3, 127-150.
- Cale, E.G., Gremillion, L. L. & McKenney, J. L. (1979, April). Price/performance patterns of U.S. computer systems. *Communications of the ACM*, 22(4), 225-232.
- Dave, D. S. & Fitzpatrick, K. E. (1991, March). Price/performance of desktop computers in the U. S. computer industry. *Information & Management*, 20(3), 161-182.
- Depke, D. A. & Brandt, R. (1991, Aug. 12). PCs: What the future holds. *Business Week*, 58-64.

- Ehrlich, I. (1977). Capital punishment and deterrence: Some further thoughts and additional evidence. *Journal of Political Economy*, 85(4), 741-748.
- Ein-Dor, P. (1985, February). Grosch's law re-revisited: CPU power and the cost of computation. *Communications of the ACM*, 28(2), 142-151.
- Ein-Dor, P. & Feldmesser, J. (1987, April). Attributes of performance of central processing units: A relative performance prediction model. *Communications of the ACM*, 30(4), 308-317.
- Ein-Dor, P. & Feldmesser, J. (1989, February). Authors' response to comments on "Attributes of the performance of central processing units: A relative performance prediction model." *Communications of the ACM*, 32(2), 259-261.
- Fitzgerald, M. (1991). Compaq becomes latest PC giant to falter. *Computerworld*, 25(30), 4.
- Greene, W. H. (1989). *LIMDEP*, Version 5.1. New York: Econometric Software, Inc.
- Grosch, H. A. (1953, April). High-speed arithmetic: The digital computer as a research tool. *Journal of Optical Society of America*, 43(4), 306-310.
- Hillkirk, J. (1989, April 26). Workstations, PCs gain popularity. *USA Today*, B1-B2.
- Kang, Y. M. (1989, May). Computer hardware performance: Production and cost function analyses. *Communications of the ACM*, 32(5), 586-593.
- Kang, Y. M., Miller, R. B. & Pick, R. A. (1986, August). Comment on "Grosch's law re-revisited: CPU power and the cost of computation." *Communications of the ACM*, 29(8), 779-781.
- Kang, Y. M. & Pick, R.A. (1989, February). Comments on "Attributes of the performance of central processing units: A relative performance prediction model." *Communications of the ACM*, 32(2), 256-259.
- Laudon, K. C. & Laudon, J. P. (1988). *Management information systems: A contemporary perspective*. New York: Macmillan.
- Levenbach, H. & Cleary, J. P. (1984). *The modern forecaster: The forecasting process through data analysis*. Lifetime Learning Publishers.
- Lewis, B. C. & Crews, A. E. (1985, March). The evolution of benchmarking as a computer performance evaluation technique. *Management Information Systems Quarterly*, 9(1), 7-15.
- Lin, W. T. (1989, June). Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches. *International Journal of Forecasting*, 5(2), 195-208.
- Lin, W. T., Rao, H. R. & Lynch, B. D. (1992, May). Response to comments on the economic analysis of microcomputer hardware. *Communications of the ACM*, 35(5), 18-21.
- Long, L. (1989). *Management information systems*. Englewood Cliffs, NJ: Prentice-Hall.
- Lynch, B. D., Rao, H. R. & Lin, W. T. (1990). Economic analysis of microcomputer hardware. *Communications of the ACM*, 33(10), 119-129.

- Mendelson, H. (1987, December). Economies of scale in computing: Grosch's law revisited. *Communications of the ACM*, 30(12), 1066-1072.
- Meth, C. (1987, February). Buyers' guide to personal computers. *Administrative Management*, 48(2), 27-31.
- Pindyck, R. S. & Rubinfeld, D. L. (1981). *Econometric models & economic forecasts*, second edition. New York: McGraw-Hill.
- Quinlan, T. (1991). Slow sales fuel PC industry price war. *Infoworld*, 13(30), 1-6.
- Reichert, A. K., Moore, J. S. & Byler, E. (1988, Winter). Financial analysis among large U. S. corporations: Recent trends and the impact of the personal computer. *Journal of Business Financial Accounting (UK)*, 15(4), 469-485.
- Remmlinger, E. & Waldmann, R. G. (1988, October). Avoid mistakes when implementing microcomputer systems. *Healthcare Financial Management*, 42(10), 90-94.
- Ryan, A. J. (1988, October 3). PCs, workstations, and small systems. *Computerworld*, 22(40), 65-92.
- Sircar, S. & Dave, D. (1986, March). The relationship between benchmark tests and microcomputer price. *Communications of the ACM*, 29(3), 212-217.
- Spitzer, J. J. (1982, December). A primer on Box-Cox estimation. *Review of Economics and Statistics*, 64, 307-313.
- Spitzer, J. J. (1978, September). A Monte Carlo investigation of the Box-Cox transformation in small samples. *Journal of the American Statistical Association*, 73(363), 488-495.
- Xenakis, J. (1987, May 25). Spotlight; Line shifts in the PC field. *Computerworld*, S1-S24.
- Xenakis, J. (1989, September 25). PCs, workstations, and small systems. *Computerworld*, 23(39), 69-106.