Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2004 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2004

Applying Data Mining Techniques to Understand the Impact of Information Technology on Organizational Productivity

Kweku-Muata Osei-Bryson
Virginia Commonwealth University

Myung Ko
The University of Texas at San Antonio

Applying Data Mining Techniques to Understand the Impact of Information Technology on Organizational Productivity

Kweku-Muata Osei-Bryson
Department of Information Systems and
The Information Systems Research Institute
Virginia Commonwealth University
Richmond, VA 23284, U.S.A.
E-Mail: Kweku.Muata@isy.vcu.edu

Myung Ko
Department of Information Systems
College of Business
The University of Texas at San Antonio
San Antonio, TX 78249
E-Mail:Mko@utsa.edu

ABSTRACT

We explore conditions under which investments in Information Technology (IT) have a statistically significant impact on productivity. Rather than using a traditional analytical technique such as regression, we use two data mining techniques (i.e. regression trees and regression splines) for doing data analysis and integrated their results. Our analysis indicates the following: 1) IT investments have a positive impact on organizational productivity only when IT investments meet certain conditions; and 2) the IT impact is not uniform but varies depending on the amounts invested in other related areas, such as *Non-IT Labor*, *Non-IT Capital*, and *IT Stock*. Thus, our study leads to some suggestions to top managers that organizations should assess the current state of investments in *Non-IT Labor*, *Non-IT Capital*, and *IT Stock* before making any further commitments to invest in IT as this current state partially determines the potential impact of additional investments in IT on organizational productivity.

Keywords

Regression Tree; Regression Splines; IT impact; Organizational productivity

INTRODUCTION

Understanding the impact of information technology (IT) on organizations has been a constant concern for both researchers and practitioners for more than two decades. While some recent studies have claimed that IT productivity paradox no longer exists (e.g. Brynjolfsson and Hitt, 1996; Shao and Lin, 2001), other studies suggest that this issue has not been completely resolved. For example, the empirical study of Strassman (1997) suggests that the investments in IT have no significant association with organizational performance. Lee and Menon (2000) found that while IT capital has a positive impact on productivity, IT labor did not. Morgan Stanley reported that U.S. companies wasted $130 billion on technology during the first two years of this millennium (Ward 2002). It should also be noted that most of these studies have used a single technique to examine the issue of the IT productivity paradox in terms of its existence or non-existence, rather than conditions under which the IT productivity paradox would or would not exist.

This study is also focused on understanding the impact of IT investments on organizational productivity, but we use multiple data mining analytical techniques (i.e. regression trees and regression splines) to identify conditions under which the impact of IT on organizational productivity would or would not exist. It has been well known in data mining research (e.g. Bauer and Kohavi, 1999) that for some datasets, a combination of individually trained predictive model can give better performance than any of the individual models. Given this fact, we believe that the use of multiple techniques can provide the opportunity for deeper exploration of our research issue, if the researcher factors in the capabilities and limitations of each technique by integrating the responses obtained from each technique as we used in our study.

LITERATURE REVIEW

Table 1 provides a brief summary of the previous studies. While some earlier studies have found no impact or mixed results (Weill, 1992; Loveman, 1994), more recent studies have found a positive impact on productivity (Hitt and Brynjolfsson, 1996; Dewan and Min, 1997; Menon, Lee, and Eldenburg, 2000; Shao and Lin, 2001).

Study	Research Method	Year Studied	Findings
Weill (1992)	Regression	1982-1987	Transactional IT: ↑ Strategic or informational IT: ↔
Loveman (1994)	Regression	1978-1984	Productivity: ↔
Hitt & Brynjolfsson (1996)	Ordinary least squares (OLS), the iterated seemingly unrelated regression (ISUR)	1988-1992	Productivity and consumer value: ↑ Business profitability: ↔
Dewan & Min (1997)	Non-linear least squares and OLS regressions	1988-1992	IT capital is a substitute for both capital and labor
Menon, Lee, Eldenburg (2000)	Stochastic frontier	1976-1994	Between non-IT labor, IT labor, IT capital, medical IT capital and productivity: ↑
Shao & Lin (2001)	Stochastic production frontier / data envelopment analysis (DEA)	1988-1992	IT has a positive effect on technical efficiency and thus, it lead to the productivity growth.

Table 1: Summary of the Previous IT and Organizational Productivity Studies

Legend:
↑: positive relationship
↔: No effect

DATA AND VARIABLES

We used a dataset that has also been used in an IT and productivity study (Menon et al., 2000). The dataset includes all hospitals in the state of Washington for the period from 1976 to 1994 excluding specialized hospitals. A total of 1130 observations are included in our analyses. Each observation represents charges and costs incurred by each hospital per year.

IT Stock, which represents IT investments, is constructed by combining *IT Capital, Medical IT Capital,* and a capitalized value of *IT Labor* expenses. *IT Capital* includes capital expenses incurred mainly for administrative purposes in the departmental accounts and *Medical IT Capital* includes capital expenses incurred for the equipment used for diagnosing and therapeutics in the departmental accounts. Table 2 provides a description of the variables and related departmental accounts. *IT labor* includes salaries and employee benefits incurred in departments where their capital expenses were classified as *IT Capital*. The previous study treated *IT Labor* as a type of expenses that produce a capital asset, which lasts 3 years on the average (Hitt and Brynjofsson, 1996) and thus, it is also included as IT investments. *Non-IT Capital* includes capital expenses incurred for the equipment used only for therapeutics purposes and also includes any capital expenses in remaining departmental accounts. The output variable, *Adjusted Patient Days,* represents a hospital performance measure. In the previous studies*, Adjusted Patient Days* has been used as a useful proxy for hospital performance (MacLean and Mix, 1991; Menon et al., 2000). For detailed description of variables, refer to the study by Menon et al. (2000).

Variable	Description (or Departmental Account)
Adjusted Patient days (*Q*)	Sum of *Inpatient Days* and *Outpatient Days*. Deflated by the output price (see below).
IT Stock *(T)*	Calculated as *IT Capital* plus *Medical IT Capital* plus three times *IT Labor*
IT Capital	Data Processing, Communications, Admitting, Patient Accounts, Central Services, Purchasing, Accounting, Medical Records, Personnel, Medical Library, Medical Staff, and Utilization Management. Deflated by Price Deflator for Fixed Investment for IT from WEFA-1994

Variable	Description (or Departmental Account)
Medical IT Capital	MRI, CT Scanning Services, Surgical Services, Recovery Room, Anesthesiology, IV Therapy Services, Electrodiagnosis, Radiology-Diagnostic, Radiology-Therapeutic, Emergency Room, Nuclear Medicine, Electromyography, Lithotripsy, Organ Acquisitions, Outpatient Chemical Deposit. Deflated by Price Deflator for Fixed Investment for IT from WEFA-1994
Non-IT Capital *(K)*	Intensive/Coronary Care, Semi-Intensive Care, Acute Care, Physical Rehabilitation, Psychiatric, Nursery, Laboratory, Pharmacy, Home Care Services and any remaining accounts. Deflated by Price Deflator for Fixed Investment for Non-IT from WEFA –1994.
IT Labor	Salaries and employee benefits charged to IT Capital accounts. Deflated by Labor Price (see below).
Non-IT Labor *(L)*	Salaries, employee benefits, and physicians' salaries charged to accounts other than IT Capital accounts. Deflated by Labor Price (see below).
Labor Price	Employment Price Index for health care services from Bureau of Labor Statistics (BLS) (1995)
Output Price	Consumer Price Index for health care services from WEFA (1994)

Table 2: Variable Definitions (Source: Menon et al. (2000) & Menon's SAS Program)

THE PRODUCTION FUNCTION

Like other studies, we also use production theory as the theoretical base for this study. We assume that a hospital's *Adjusted Patient Days (Q)* depends on the use of various inputs, *Non-IT Capital (K)*, *IT Stock (T)*, and *Non-IT Labor (L)*, and so our production function has the following form:

$$Q = f(K, L, T) \tag{1}$$

Because we are concerned about the conditions under which the impact of on productivity could occur, our analyses involve the use of the Translog production function, which is more flexible functional form than the Cobb-Douglas function (Evans et al., 2000), allowing for the exploration of interactions between the input variables. The relevant Translog production function can be expressed as

$$\log_e Q = \beta_0 + \beta_K \log_e K + \beta_L \log_e L + \beta_T \log_e T + \tfrac{1}{2}\beta_{KK}(\log_e K)^2 + \tfrac{1}{2}\beta_{LL}(\log_e L)^2 + \tfrac{1}{2}\beta_{TT}(\log_e T)^2 + \beta_{KL}\log_e K \log_e L + \beta_{LT}\log_e L \log_e T + \beta_{TK}\log_e T \log_e K \tag{2}$$

OVERVIEW ON ANALYTICAL TECHNIQUES

In this section, we provide overviews on regression trees and regression splines since many readers may not be familiar with these techniques.

Regression Trees

A decision tree (DT) is a tree-shaped knowledge structure, consisting of nodes, branches, and leaves. For a given decision problem, each non-leaf node is associated with one of the decision variables, each branch from a non-leaf node is associated with a subset of the values of the corresponding decision variable, and each leaf node is associated with a value of the target (or dependent) variable.

There are two main types of DTs are 1) classification trees and 2) regression trees (RT). For a classification tree, the target variable takes its values from a discrete domain, and for each leaf the DT associates a probability for each class. For the RT, the target variable takes its values from a continuous domain, and for each leaf, the DT associates the mean value of the target variable.

The generation of a DT involves partitioning the model dataset into at least two parts: the training and the validation (test) datasets. Once an RT is generated from the training dataset, it is evaluated against the validation (or test) dataset and a subtree that has the lowest error rate against the validation dataset is generated.
While the most commonly used performance measure for an RT is based on its *predictive accuracy* (e.g. R-squared, average squared error), among the other important performance measures are *simplicity* and *stability*. *Simplicity* is referred to as the interpretability of the RT, is often based on the number of leaves in the RT. *Stability* of the RT refers to obtaining similar results for the training and validation datasets. Although there is no standard quantitative measure for stability, one way to

assess the stability of the RT can be achieved by comparing the predicted mean value of the target variable based on the training dataset and the corresponding value for the validation dataset for each rule of the RT.

Although RTs are similar to regressions since both techniques are used for the prediction, the RT model is a step function, whereas the regression model is a continuous function (Clark and Pregibon, 1992). Compared to regression models, RTs provide a model with better interpretability because the model represents interpretable English rules or logic statements. There have been instances where a decision tree has shown clues to data sets while a traditional linear regression analysis could not clearly indicate them (Breiman, et al., 1984). However, perturbations in data could cause instability of RTs and thus, it can cause the predictive capabilities of a tree (Nerini et al., 2000; Hastie, Tibshirani, and Freeman, 2001). To minimize instability, we can generate multiple trees and choose the best model that fits one's objective.

Although the RT technique has not been used in IT and productivity research, it has been successfully applied in various fields including software engineering (e.g. Gokhale and Lyu, 1997), epidemiology (Ciamplie et al., 1995), and production management (e.g. Markham et al., 1998).

Regression Splines

A Regression Splines (**RS**) approach models the mean outcome as piecewise polynomial function $f(x)$ which can be obtained by dividing the range of each predictor variable into one or more intervals and representing f by a separate polynomial in each interval (Hastie et al., 2001). A regression spline function can be expressed as a linear combination of piecewise polynomial *basis functions* ***(BF)*** that are joined together smoothly at the knots, where a *knot* specifies the end of one region of data and the beginning of another (Steinberg, Colla, and Martin, 1999). The coefficient of each basis function is estimated by minimizing the sum of square errors, which is similar to the estimation process of regression, but involving local data for the given region.

Multivariate Adaptive Regression Splines (MARS) approach was motivated by adaptive regression spline (Hastie and Tibshirani, 1990) and the recursive partitioning regression (RPR) approach (Breiman et al., 1984). MARS automatically selects locations and degree of knots. It builds a model using a forward stepwise regression selection and a backward stepwise deletion strategy. In the first phase, MARS builds an overfitted model by adding basis functions. In the second phase, basis functions that have the least contribution to the model are deleted and the model is optimized (Steinberg et al, 1999).

MARS uses the basis functions in pairs of the form $(x - t)_+$ and $(t - x)_+$ where t is the knot. The "+" represents positive part, thus, $(x - t)_+$ means $x - t$ if $x > t$ or 0 if otherwise and $(t - x)_+$ means $t - x$ if $x < t$ or 0 if otherwise (Hastie and Tibshirani, 1990; Hastie et al., 2001). MARS provides ANOVA decomposition, which identifies the relative contributions of each of the predictor variables and the interactions between variables (Friedman, 1991). Although MARS has not been used in IT productivity research, this technique has been successfully applied in various fields including geography (e.g. Abraham & Steinberg, 2001), genetics (York & Eaves, 2001), and finance (e.g. Abraham, 2002).

EMPIRICAL ANALYSIS

Results of Regression Tree Based Analysis

We used the data mining software, SAS Enterprise Miner (EM), version 4.1. Following the traditional data mining approach for supervised learning, we partitioned the dataset into *Training* (R) and *Validation* (A) datasets and generated a regression tree. The *predictive accuracy* obtained from the RT in terms of *R-squared* is 0.869 for the *Training* dataset and 0.858 for the *Validation* dataset. Table 3 includes a rule set obtained from our RT based analysis. As shown in Table 3, the rule set from the RT generated the fourteen rules. Each row represents a rule and the *Condition Component* columns represent the range of values for the relevant input variables for the relevant rule. The *Target* columns represent the predicted mean values obtained from the *Training* and the *Validation* datasets for the target variables, where the standard deviation (SD) is enclosed in parentheses in the *Training* column. For example, the first rule can be expressed as "*If $\log_e$ (Non-IT Labor) is less than 14.5717, the predicted mean $\log_e$ (Adjusted Patient Days) is 8.9747 with a standard deviation of 0.3338*." The *IT Impact* column indicates whether the *IT Stock* variable was included in the relevant rule and specifies whether IT makes a contribution to the target value. Also, predicted mean values of the target variable from the *Training* dataset and the

Validation dataset in Table 3 are very close to each other. Thus, the RT demonstrates the stability although instability is one of limitations of the regression tree analysis.

Rule	Condition Component			Target: Mean $\log_e V$		IT Impact
	Non-IT Labor **($\log_e L$)**	***Non-IT Capital*** **($\log_e K$)**	***IT Stock*** **($\log_e T$)**	**Training (SD)**	**Validation**	
1	[0, 14.5717]	Not selected	Not selected	8.9747 (0.3338)	8.9436	No
2	[14.5717, 14.8966]	Not selected	Not selected	9.1983 (0.4421)	9.2660	No
3	[14.8966, 15.1752]	Not selected	Not selected	9.6050 (0.4720)	9.8419	No
4	[15.1752, 15.7554]	[12.8680, ∞]	Not selected	9.6536 (0.3026)	9.7182	No
5	[15.1752, 15.7554]	[0, 12.8680]	Not selected	10.1528 (0.3606)	10.1772	No
6	[15.7554, 16.3942]	[13.1910, ∞]	Not selected	10.2371 (0.2712)	10.2582	No
7	**[15.7554, 16.3942]**	**[0, 13.1910]**	**[0, 15.2176]**	**10.4320 (0.4778)**	**10.5603**	**Yes**
8	**[15.7554, 16.3942]**	**[0, 13.1910]**	**[15.2176, ∞]**	**10.9523 (0.2111)**	**10.8925**	**Yes**
9	[16.3942, 16.8873]	[13.6985, ∞]	Not selected	10.8374 (0.2164)	10.8514	No
10	[16.3942, 16.8873]	[0, 13.6985]	Not selected	11.2650 (0.2160)	11.2827	No
11	[16.8873, 17.2691]	[14.3513, ∞]	Not selected	11.0363 (0.1762)	11.1423	No
12	[16.8873, 17.6502]	[0, 14.3513]	Not selected	11.5347 (0.2714)	11.5509	No
13	[17.2691, 17.6502]	[14.3513, ∞]	Not selected	11.3476 (0.1306)	11.4420	No
14	[17.6502, ∞]	Not selected	Not selected	11.9474 (0.3217)	11.6782	No

Table 3: The Ruleset of RT – Sorted by $\log_e(L)$ and mean $\log_e(V)$ for Training Dataset

With regards to the impact of investments in the IT stock on productivity, these RT-based results suggest that:

1) *IT Stock* has a positive impact on target variable only when *Non-IT Labor* expenses are within the middle range associated with rules 7 and 8 (i.e. $\log_e L \in$ [15.7554, 16.8773]). However, when the *Non-IT Labor* is out of this range (i.e. $\log_e L \notin$ [15.7554, 16.8773]), investments in IT have no impact on the target variable.
1) Even when *IT Stock* has a positive impact on target variable, its impact is not uniform since the impact of IT on the target variable is conditioned by the amounts invested in *IT Stock, Non-IT Labor*, and *Non-IT Capital* (see rules 7 and 8).
1) The mean value for the target variable is lower in the range where *IT Stock* has a positive significant impact on the target variable (see rules 7 and 8) than it is out of the range where both investments in *Non-IT Labor* are the highest and there is no significant IT investments impact (see rules 9, 10, 11. 12, 13, and 14). This suggests that further investments in IT might not necessarily increase organizational productivity, once organization reaches its maximum level of IT investments.

In order to validate our findings, we have also generated three additional RTs that varied *Splitting Criterion, the Minimum Number of Observations per Leaf*, and the *Observations Required for a Split Search.* Then we compared the rulesets that were generated from these three RTs. Their results are consistent with findings from our initial RT.

Results of Regression Splines Based Analysis

We used the Multivariate Adaptive Regression Splines (MARS) software, version 2.0 by Salford Systems. The *R-squared* for the RS model was 0.90 thus indicating that it has relatively high predictive power. Table 4 displays the results of the RS with two-way interaction, which contains a constant (basis function 0), 11 basis functions (regions), their coefficients, a variable that is directly related to each basis function and a variable that is interacting with another variable if any, and a knot location for each basis function.

The knots for predictor variables are as follows:

- *Non-IT Labor*: L_{cv1} and L_{cv2}, where $\log_e(L_{cv1}) = 16.703$ and $\log_e(L_{cv2}) = 15.350$
- *IT Stock*: T_{cv1}, T_{cv2}, and T_{cv3}, where $\log_e(T_{cv1}) = 14.086$, $\log_e(T_{cv2}) = 15.967$, and $\log_e(T_{cv3}) = 14.738$

- *Non-IT Capital*: K_{cv1} and K_{cv2}, where $\log_e(K_{cv1}) = 12.220$ and $\log_e(K_{cv2}) = 13.714$.

	Basis Function (BF)	Coefficient	Variable	Parent	Knot (log Value)
0		11.923			
1	BF1 = max (0, $\log_e$L – 16.703);	1.114	$\log_e$L		16.703
2	BF2 = max (0, 16.703 – $\log_e$L);	-1.243	$\log_e$L		16.703
3	BF3 = max (0, $\log_e$K – 12.220);	-0.589	$\log_e$K		12.220
4	BF4 = max (0, 12.220 – $\log_e$K)				
5	BF5 = max (0, $\log_e$L – 15.350)*BF4;	-1.067	$\log_e$L	$\log_e$K	15.350
6	BF6 = max (0, 15.350 – $\log_e$L)*BF4;	0.370	$\log_e$L	$\log_e$K	15.350
7	BF7 = max (0, $\log_e$T –14.086) *BF4;	0.611	$\log_e$T	$\log_e$K	14.086
10	BF10 = max (0, 15.967 - $\log_e$T);	-0.207	$\log_e$T		15.967
11	BF11 = max (0, $\log_e$T – 14.738) * BF3;	0.050	$\log_e$T	$\log_e$K	14.738
12	BF12 = max (0, 14.738 - $\log_e$T) * BF3;	0.856	$\log_e$T	$\log_e$K	14.738
13	BF13 = max (0, $\log_e$K – 13.714) * BF2;	0.597	$\log_e$K	$\log_e$L	13.714

Table 4: Final Model

Based on the model shown in Table 4, our Translog function can be expressed as follows:

$$\log_e Q = 11.923 + 1.114*BF1 – 1.243*BF2 - 0.589*BF3 – 1.067*BF5 + 0.370*BF6 + 0.611*BF7 - 0.207*BF10 + 0.050*BF11 + 0.856*BF12 + 0.597*BF13$$

If the sign of the coefficient of a basis function is the same as the sign of the variable in that function (e.g. *BF1, BF10*), then the contribution of given variable in terms of that basis function is positive, while if the corresponding signs are different, then the contribution of given variable in terms of that basis function is negative (e.g. *BF3, BF12*).

With regards to the impact of IT investments on productivity, these RS-based results suggest that:

1) *IT Stock* has an impact on productivity since *IT Stock* involved in four (4) of the basis functions (e.g. *BF7, BF10, BF11,* and *BF12*).
1) The overall impact on productivity is conditioned both by the amount invested in *IT Stock* (see *BF10*) and the investments in *Non-IT Capital* (see *BF7, BF11*, and *BF12*).
1) The impact of investments in *IT Stock* is not conditioned by investments of *Non-IT Labor* as none of the basis functions that involve *IT Stock* has *Non-IT Labor* as a parent.
1) The overall impact on productivity is not uniform because coefficients for each of the basis functions that involve *IT Stock* are different.
1) Under certain conditions, investments in *IT Stock* have a positive impact on productivity. For example if the investments in *Non-IT Capital* is less than K_{cv1}, then the impact of *IT Stock* on productivity is positive since both of basis functions *BF7* and *BF10* have the same sign of the variable and the sign of the coefficient. However, if the investments in *Non-IT Capital* are greater than K_{cv1}, and the investments in *IT Stock* is less than T_{cv3} (see *BF12*), investments in *IT Stock c*ould have a negative impact on productivity. In this case, although the contribution of the *IT Stock* from *BF10* is positive, the one from *BF12* is negative (since the sign of $\log_e$T in *BF12* is negative while the sign of the coefficient of *BF12* is positive), and so the overall impact could be negative. Thus, for those situations where even after additional investments in the *IT Stock* make the total *IT Stock* investments still below $T_{cv3,}$ there are no resulting increases on productivity.

OVERALL RESULTS FROM BOTH OF DATA MINING TECHNIQUES

Capabilities of Techniques

Regression Trees and Regression Splines techniques have different capabilities and limitations, which restrict the nature of responses that they can provide to the research issue in study. Thus, when each technique is used, it is important that capabilities and the limitations of the technique be factored in any interpretation of the results provided by the technique. Table 5 describes the capabilities of each technique used in our analyses.

Capability	RT	RS: Two-Way Interactions
Importance	Yes	Yes
Coefficient	No	Yes
Partitioning	Yes	Yes
Hierarchy	Yes	Yes

Capability:
Importance: Can identify order of importance of the variables in the predictive model
Coefficient: Estimates value of the coefficient for each variable
Partitioning: Can provide a model with conditional response by partitioning a variable's values
Hierarchy: Can automatically identify the hierarchical nature of interaction between variables

Table 5: Capabilities of Analytical Techniques

Integration of Responses from Each Technique

In this subsection, we summarize the findings by integrating the responses obtained from each technique. Table 6 provides the summary of the results of each analysis. There are two research questions, which we are trying to answer. First, do IT investments have a positive impact on organizational productivity? Second, is the impact of IT on organizational productivity uniform?

Research Questions	RT	RS	Overall Summary
A positive IT impact on productivity	YES (under some conditions); NO (under other conditions);	YES (under some conditions) NO (under other conditions)	**YES,** only when IT investments meet some conditions
The impact of IT is uniform	NO. It is conditioned by the amounts invested in other areas.	NO. It is conditioned by the amounts invested in other areas.	**NO,** the impact of IT is not uniform but it is conditioned by the amounts invested in other related areas.

Table 6: The Results of Each Analysis

The results from the both data mining techniques indicate that IT investments have a positive impact on organizational productivity only when they meet some conditions, where each condition is described in the results of each analysis. Both techniques consistently indicate that the impact of IT is not uniform but is conditioned by the amounts invested in *Non-IT Capital* and/or *IT Stock* and/or *Non-IT Labor.* This is due to the differences of each technique's capabilities and limitations, which restrict the nature of responses to the research in study. We believe that identifying these differences and integrating responses from each technique advance our understanding of the IT impact on organizational productivity.

CONCLUSION

While most previous studies have attempted to assess the impact of IT investments using a single technique, we used two data mining techniques (i.e. regression trees and regression splines) to understand the impact of IT investments on organizational productivity. This approach provided us with the opportunity for both forming a consensus result and developing a better understanding of the impact of IT investments on productivity. By integrating the responses obtained from both data mining techniques, we found that IT investments have a positive impact on productivity only when they meet some conditions. In addition, the impact of IT is not uniform but is conditioned by the amounts invested in *Non-IT Labor*, *Non-IT Capital*, and *IT Stock*.

Our study is different from the previous studies that examined the impact of the IT productivity in terms of its existence or non-existence. Rather, this study explores the conditions under which the impact of IT on productivity would or would not exist. Findings from our study lead to some suggestions. When an organization considers making additional investments in IT, top managers should assess the level of the organization's current state with regards to its investments in *Non-IT Labor*, *Non-IT Capital*, and *IT Stock* before making any further commitments to invest in IT as this current state partially determines the potential impact of additional investments in *IT* on organizational productivity. Results of our study can also help making other decisions. Either increasing or reducing the overall investments in other areas can lead to the increase in organizational productivity. By identifying these relationships, organizations can use their resources more efficiently.

ACKNOWLEDGMENTS

We are grateful to Professors Byungtae Lee and Nirup Menon for sharing their dataset and Professor Menon's SAS program with us, thus facilitating this study.

REFERENCES

1. Abraham, A. and D. Steinberg. (2001). Is Neural Network a Reliable Forecaster on Earth? A MARS Query!, *International Work – Conference on Artificial and Natural Neural Networks*, Springer-Verlag Germany, Jose Mira and Alberto Prieto (Eds.), 679 –686, June 2001.
2. Abraham, A. (2002). Analysis of Hybrid Soft and Hard Computing Techniques for Forex Monitoring Systems, *World Congress on Computational Intelligence*, May 2002, 1616-1622.
3. Bauer, E., and R. Kohavi. (1999) An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants *Machine Learning*, 36, 105-139.
4. Breiman, L., J. H. Friedman, R. Olshen, and C. Stone. (1984). Classification and Regression Trees. Belmond, California, Wadsworth International Group.
5. Brynjolfsson, E. and L. Hitt. (1996). Paradox Lost? Firm-level Evidence on the Returns to Information Systems Spending, *Management Science,* 42,4, 541-558.
6. Ciampi, A., Negassa, A., and Z. Lou. (1995) Tree-Structured Prediction for Censored Survival Data and the Cox Model., *Journal of Clin Epidemiol*, 48 ,5, 675-689.
7. Clark, L. and D. Pregibon (1992) Tree-Based Models. Statistical Models in S. J. M. Chambers and T. J. Hastie. Pacific Grove, California, Wadsworth & Brooks / Cole Advanced Books & Software, 377-419.
8. Dewan, S. and C. Min (1997). The Substitution of Information Technology for Other Factors of Production: A Firm Level Analysis, *Management Science* , 43,12, 1660-1675.
9. Evans, D., Green, C., and Murinde, V. (2000) The Importance of Human Capital and Financial Development in Economic Growth: New Evidence Using the Translog Production Function, Finance and Development Research Program Working Paper Series, Paper No. 22, November: 1-31.
10. Friedman, J. (1991) Multivariate Adaptive Regression Splines, *The Annals of Statistics*, Vol. 19, 1-141.
11. Gokhale, S. and M. Lyu (1997) Regression tree modeling for the prediction of software quality in *Proceedings of Third ISSAT International Conference on Reliability & Quality in Design*, Hoang Pham, Ed., Anaheim, CA, 31- 36. http://citeseer.nj.nec.com/gokhale97regression.html
12. Hastie, T. and R. Tibshirani. (1990) Generalized Additive Model, London, Chapman and Hall.
13. Hastie, T., Tishirani, and J. Friedman (2001) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
14. Hitt, L. M. and E. Brynjolfsson (1996) Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value, *MIS Quarterly*, 20,2,121-142.
15. Ko, M. and K.-M. Osei-Bryson (2004) The Productivity Impact of Information Technology in the Healthcare Industry: An Empirical Study Using a Regression Spline-Based Approach, *Information and Software Technology*, 46,65-73.

16. Kuhnert, P., Kim-Anh D., and R. McClure (2000) Combining Non-Parametric Models with Logistic Regression: An Application to Motor Vehicle Injury Data, *Computational Statistics & Data Analysis,_* 34, 371-386.
17. Lee, B. and N. Menon (2000) Information Technology Value Through Different Normative Lenses. In *Journal of Management Information Sstems* 16,4, 99-119.
18. Lichtenberg, F. (1995) The Output Contributions of Computer Equipment and Personnel: A Firm-Level Analysis, *Economics of Information and New Technology,* 3,4, 201-217.
19. Loveman, G. (1994)"An Assessment of the Productivity Impact of Information Technologies, *Information Technology and the Corporation of the 1990s: Research Studies*. T. J. Allen and M. S. Scott Morton, Oxford University Press**:** 84-110.
20. MacLean, M. and P. Mix. (1991) Measuring Hospital Productivity and Output: The Omission of Outpatient Services, *Health Reports*, 3, 229-244.
21. Markham, I., R. Mathieu, and B. Wray (1998) A Rule Induction Approach for Determining the Number of Kanbans in a Just-In-Time Production System, 34**,** 4, 717-727.
22. Menon, N., Lee, B., and L. Eldenburg (2000) Productivity of Information Systems in the Healthcare Industry, *Information Systems Research*, 11, 83-92.
23. Nerini, D., Durbec, J., and C. Mante. (2000) Analysis of Oxygen Rate Time Series in a Strongly Polluted Lagoon Using a Regression Tree Method, *Ecological Modelling***,** 133, 95-105.
24. Shao, B. and W. Lin (2001) Measuring the Value of Information Technology in Technical Efficiency with Stochastic Production Frontiers, *Information and Software Technology,* 43, 447-456.
25. Steinberg, D. Colla, P., and K. Martin (1999) MARS User Guide, San Diego, CA, Salford Systems.
26. Weill, P. (1992) The Relationship Between Investment in Information Technology and Firm Performance: A Study of the Valve Manufacturing Sector*, Information Systems Research,* 3, 4, 307-333.
27. Strassman, P. (1997) Will Big Spending on Computers Guarantee Profitability? *Datamation,* February 1997.
28. Ward, S. (2002) Companies Squander Billions on Tech, *USA Today, May 20,* 2002.
29. York, T. and L. Eaves (2001) Common Disease Analysis Using Multivariate Adaptive Regression Splines (MARS): Genetic Analysis Workshop 12 Simulated Sequence Data, *Genetic Epidemiology,* 21, 649-654.