2002

# A Regression Tree Based Exploration of the Impact of Information Technology Investments on Firm Level Productivity

Myung Ko
*Virginia Commonwealth University*, msko@vcu.edu

Kweku-Muata Bryson
*Virginia Commonwealth University*, Kweku.Muata@isy.vcu.edu

# A REGRESSION TREE BASED EXPLORATION OF THE IMPACT OF INFORMATION TECHNOLOGY INVESTMENTS ON FIRM LEVEL PRODUCTIVITY

**[1]Myung Ko and [2] Kweku-Muata Bryson**
Department of Information Systems and The Information Systems Research Institute
Virginia Commonwealth University
Richmond, VA 23284, U.S.A.
EMail: [1]Msko@vcu.edu, [2]Kweku.Muata@isy.vcu.edu

## ABSTRACT

*The issue as to whether investments in information technology (IT) contribute significantly to organizational productivity has been of major concern for many years, and various studies have lead to seemingly contradictory results. In this paper, we analyze the relationship between IT investments and firm level productivity using regression trees (RT). Use of this data mining technique represents a novel approach to identifying elements of this relationship as most previous studies have primarily used econometrics-based techniques. While the use of traditional techniques has provided valuable results, our RT-based analysis revealed additional findings that were not identified in the previous studies. For example unlike the econometric-based studies that identify a uniform impact of IT investments on productivity, our RT-based analysis suggests that IT has an impact on productivity only when Non-IT Labor expenses are within the interior interval. Also, even within this range, the impact of IT is not uniform.*

**Keywords:** Information Technology Investments, Productivity, Productivity Paradox, Regression Trees, Data Mining

## 1. INTRODUCTION

Investments in information technology (IT) have grown continuously over the past thirty years (Dewan and Min, 1997; Fernberg, 1995; King, 1998; Shu, 2001). While organizations have used IT as a means to improve company's profitability, substantiating business value of the IT investments is not an easy task and has been a major concern among IT managers and information systems (IS) researchers. Many researchers have attempted to examine the contribution of IT to output but have failed to show any evidence of IT impact on productivity in spite of the increased IT investments. This so called the "IT productivity paradox" that has been debating issues among IS researchers since mid-1980s (Brynjolfsson, 1993; Brynjolfsson and Hitt, 1996; Hitt and Brynjolfsson, 1996; Jurison, 1997; King, 1998; Rai et al., 1997; and Sircar et al., 2000). Possible explanations for the productivity paradox include 1) mismeasurement of outputs and inputs, 2) time lags due to learning and adjustment, 3) redistribution of profits, and 4) mismanagement of IT (Brynjolfsson, 1993; Loveman, 1994); 5) Inappropriateness of traditional productivity measures. Some claimed that inconsistent findings from IT productivity research were due to interchanging terms between productivity and financial performance and also lack of adequate data (Sircar et al., 1998 & 2000). However, recent studies have claimed that IT productivity paradox no longer exists (e.g. Garretson, 1999; McGee, 2000;

Brynjolfsson and Hitt, 1996; Dewan and Min, 1997; Shao and Lin, 2000, 2001). Brynjolfsson and Hitt (1996) assert that the productivity paradox had disappeared by the early 1990s.

The goal of this paper is to explore the relationship between IT investments and organizational productivity using a Regression Tree (RT) technique from the field of Data Mining (DM). DM is a popular application in analyzing and discovering hidden information from the datasets. While the majority of the previous studies have applied the Econometric analysis or Data Envelopment Analysis (DEA), we applied the RT based analysis on the dataset of Hitt and Brynjolfsson (1996), which was repeatedly used in previous research (Brynjolfsson and Hitt, 1996; Dewan and Min, 1997; Shao and Lin, 2000 & 2001). Employing the same dataset from the previous studies could help us to compare findings from our study with previous studies without any bias since any contradicting findings that may be caused by the different datasets could be reduced or eliminated. While the use of traditional techniques has provided valuable results, our RT-based analysis revealed additional findings that were not identified in the previous studies. For example unlike the econometric-based studies that identify a uniform impact of IT investments on productivity, our RT-based analysis determines that Non-IT Labor is the most significant variable for predicting productivity but partitions Non-IT Labor expenses into three intervals such that IT Stock is only relevant as a predictor variable when the Non-IT Labor expenses fall within the interior interval. This suggests that IT has an impact on productivity only when Non-IT Labor expenses are within this interior range. The RT-based analysis also suggests that even within this range, the impact of IT investments is not uniform. In addition, the value of productivity is lower when the Non-IT Labor amount is within this range than it is above the range. Thus, IT investments do not play a role in determining the highest value of productivity and our findings provide an evidence of the 'IT productivity paradox.'

The remainder of the paper is organized as follows. The next section discusses the overview on previous research of the IT impact on productivity at the firm level. Section 3 describes the dataset including variables. Section 4 describes the production function, on which our research is based. Section 5 introduces the methodology used in our study, a Regression Tree technique. Section 6 discusses the empirical results including comparison of our results with previous studies. Section 7 concludes the paper including suggestions for future research.

## 2. PREVIOUS RESEARCH ON IT AND PRODUCTIVITY

Various researchers have examined the IT impact on productivity at the firm level and their findings have been conflicting. Some studies have found no impact between IT investments and productivity. Weil (1992) did not find any significant relationship between total IT investments and firm's performance. Loveman (1994) also examined the IT productivity at the firm level. His study provided no evidence of IT impact on output or labor productivity despite of disaggregating IT according to IT intensity, industry, and market share. On the other hand, recent studies have found a positive relationship. Lichtenberg (1995) reported that computer capital and IS labor have contributed to firm's output substantially. Hitt and Brynjolfsson (1996) found that IT spending has a positive impact on productivity and provided the significant value for consumers. However, their analysis did not show any evidence of improvement in business profitability. Dewan and Min (1997) assessed IT substitutability for other inputs and found that IT capital is a substitute for capital and labor. Shao & Lin (2001) and Shao (2000) examined the impact of IT investments on technical efficiency in the firm's production process using the dataset of Hitt and Brynjolfsson (1996). While Shao and Lin (2000) used the Cobb-Douglas and translog stochastic production frontiers in their study, Shao (2000) used both stochastic production frontiers and a data envelopment analysis (DEA). Both studies indicated IT has a positive effect on technical efficiency in the production process whether IT investments are treated as a firm-specific factor or a production factor.

**Table 1: Research Summary of IT Impact on Firm Level Productivity**

| Study | Research Method | Dataset / period | Measures | Results of IT impact |
|---|---|---|---|---|
| Weil (1992) | Econometrics | 33 manufacturing firms during 1982-1987 | IT investment type, (transactional, strategic, informational), financial measures (sales growth, return on assets (ROA), measures of labor productivity | Transactional IT - positive relationship with performance<br>Strategic or informational IT - no relationship |
| Loveman (1994) | Econometrics | 60 manufacturing business units during 1978 - 1984 | Material expenditure, non-IT purchased services expenditure, total labor compensation, non-IT capital, IT capital | No evidence of productivity gains from IT investments |
| Lichtenberg, Frank (1995) | Econometrics | Firms reported by Computerworld and InformationWeek during the period 1988-1991 | IS budget, Computer capital, non-computer capital, non-computer labor, IS labor, | Output contributions of IS capital and IS labor are substantial. Computer capital and labor jointly contribute about 21 percent of output |
| Hitt & Brynjolfsson (1996) | Econometrics | 370 firms during 1988-1992 | Variables in the dimensions of Productivity, profitability, and consumer value | Increased productivity and consumer value but no impact on business profitability |
| Dewan & Min (1997) | Econometrics | 370 firms during<br>1988-1992 | IT substitutability for other inputs | IT capital is a substitute for capital and labor. Evidence of excess returns on IT capital relative to labor |
| Shao & Lin (2001) / Shao (2000) | Econometrics / data envelopment analysis (DEA) | 370 firms during<br>1988-1992 | Capital, labor, IT investments | IT has a positive effect on technical efficiency and thus, it lead to the productivity growth. |

## 3. DATA AND VARIABLES.

We employed the same dataset used by Hitt and Brynjolfsson (1996). There were several previous studies, which have also used this dataset (Brynjolfsson and Hitt, 1996; Shao and Lin, 2000; Shao, 2000; Shao and Lin, 2001). Since more detailed descriptions of the data are available in previous studies, brief descriptions of the data are included in this paper.

IT related data were obtained from the International Data Group (IDG) and other data were obtained from Standard & Poor's Compustat. Since annual survey by IDG includes mainly large U.S. firms that are publicly traded, IT related data were matched to financial related data in Compustat and then the data were converted into constant 1990 dollars. The data are collected for the period from 1988 to 1992 and represented an unbalanced panel of 370 firms with 1252 observations. Compared to previous studies, the number of observations in our study is higher because Regression Trees treats the missing value as an acceptable value unless more than 50 % of information is missing.

Variables in our study are included in Table 2. The variable, IT STOCK (T), represents IT investments, which is computer capital and a capitalized value of IS labor expenses. Computer Capital represents the total market value of central processors and PCs. IS labor considered as capital expenditure that produce an asset which last three years on the average and thus is included as part of IT STOCK (See description of IT Stock in Table 2).

**Table 2: Variable definitions (Source: Hitt and Brynjolfsson, 1996, p. 128)**

| **Variable** | **Description** | **Source** |
|---|---|---|
| OUTPUT | Gross Sales deflated by Output Price (see below) | Compustat |
| VALUE ADDED (V) | Output minus non-labor expense. Non-labor expense is calculated as total firm expenses (excluding interest, taxes, and depreciation) divided by Output Price less Labor (see below) | Compustat |
| IT STOCK (T) | Calculated as Computer Capital plus three times IS Labor | Calculation |
| COMPUTER CAPITAL | Market value of central processors plus value of PCs and terminals obtained from IDG survey. Deflated by Computer Price (see below). Average value of PC determined as weighted average of PC price from Berndt and Griliches (1990) and value of PC from IBM. Resulting estimate is $2,840 in 1990 dollars. | IDG Survey |
| NON-COMPUTER CAPITAL (K) | Deflated book value of Capital less Computer Capital as calculated above (for deflator see below). | Compustat |
| LABOR (L) | Available labor expenses or estimated labor expenses based on sector average labor costs times number of employees minus IS Labor. Deflated by Labor Price (see below). | Compustat |
| IS LABOR | Labor portion of IS budget. Deflated by Labor Price (see below). | IDG Survey |
| INDUSTRY | Primary Industry at the 2-digit SIC level. | Compustat |
| COMPUTER PRICE | Gordon’s deflator for computer systems– extrapolated to current period at same rate of price decline (-19.7% per year) | Gordon, 1993 |
| OUTPUT PRICE | Output deflator based on 2-digit industry from BEA estimates of industry price deflators. If not available, sector level deflator for intermediate materials, supplies, and components. | Bureau of Economic Analysis, 1993 |
| LABOR PRICE | Price index for total compensation | Council of Economic Advisors, 1992 |

## 4. THE PRODUCTION FUNCTION.

We employed the production function approach, which was used by many previous researchers (Loveman, 1994; Hitt & Brynjolfsson, 1996; Dewan & Min, 1997; Shao 2000). The production function assumes that a firm uses various inputs to produce outputs and can be expressed as the following form:

$$V = f(T, K, L) \qquad (1)$$

Where the OUTPUT (V) is the firm VALUE ADDED and the three variables, IT STOCK (T), NON-COMPUTER CAPITAL (K), and the LABOR (L) are input variables. Because IT Stock (T) includes IS Labor and computer capital, we used NON-COMPUTER CAPITAL for CAPITAL (K), and NON-IT LABOR amount for LABOR (L). This is not to include amounts twice in the model.

Thus, production of V depends on the use of IT STOCK (T), NON-COMPUTER CAPITAL (K), and the NON-IT LABOR (L). The simplest production function is known as the Cobb-Douglas function and it is one of the most commonly used production functions due to empirical validity (e.g. Hitt & Brynjolfsson, 1996). The Cobb-Douglas function can be expressed as the following form:

$$V = T^{\beta 1} K^{\beta 2} L^{\beta 3} \quad (2)$$

Where parameters $\beta_1$, $\beta_2$, and $\beta_3$ are estimated constants. We can convert equation (2) to an equation that can be expressed in terms of linear regression by taking natural logarithms and adding an error term $\varepsilon$:

$$\text{Log } V = \beta_0 + \beta_1 \log T + \beta_2 \log K + \beta_3 \log L + \varepsilon \qquad (3)$$

Our study is based on the Cobb-Douglas production using Regression Tree (RT) based approach. While the linear regression model estimates a continuous linear equation and predicts $\beta_1$ value for the IT impact, RT is a stepwise linear equation that predicts the target value and provides set of decision rules. Previous researchers who have examined the IT impact applied either regression or the DEA approach. Thus, our approach provides a different perspective from previous studies.

## 5. OVERVIEW ON REGRESSION TREES.

Data mining techniques allow organizations to explore and discover meaningful, previously hidden information from huge organizational databases. An important knowledge structure in data mining activities is the decision tree (DT). A DT is a tree structure representation of the given decision problem such that 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 decision trees (DTs): 1) classification trees (CT); 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 (and in some cases a value) for each class (i.e. value of the target variable). The class that is assigned with a given leaf of the classification tree results from a form of majority voting in which the winning class is the one that provides the largest class probability even if that probability is less than fifty percent (50%). For the regression tree (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. Thus, a DT is an alternative approach to continuous linear models for regression problems and to linear logistic models for classification problems (Clark and Pregibon, 1992).

To generate a DT, the model dataset is partitioned into at least two parts: the training dataset and the validation dataset (commonly referred to as the test dataset). Then it undergoes two major phases of process: the *growth phase*, and the *pruning phase* (e.g. Kim and Koehler, 1995). The *growth phase* involves constructing a DT from the training dataset in a top-down recursive manner (Han and Kamber, 2001; Hand et al., 2001). In this phase, either each leaf node is associated with a single class or further partitioning of the given leaf would result in the number of cases in one or both child nodes being below some specified threshold. The *pruning phase* aims to generalize the DT that was generated in the *growth phase* in order to avoid overfitting. In this phase, the DT is evaluated against the test (or validation) dataset in order to generate a subtree of the DT generated in the *growth phase* that has the lowest error rate against the validation dataset. The DT that results from this two-phase process is the subtree of the pruning phase that had the smallest error (i.e. average squared error for

RT) when applied to the validation dataset. It follows that this DT is not independent of the training dataset or the validation dataset.

There are several criteria for measuring performance of RTs. Although the *predictive accuracy* (R-squared, average squared error) is the most commonly used performance measure for an RT, *simplicity* and *stability* are also important measures for an RT. *Simplicity* refers to the interpretability of the RT and is often based on the number of leaves in the RT, but the chain lengths of the corresponding rules could also be used to determine this criterion of the RT. On the other hand, a stability of an 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.

In our study, we employ RTs to explore the impact of IT investments on productivity at the firm level. Although RTs are similar to regressions since both of them are used for prediction, the main difference between the two models is that RTs use a step function and the regressions use continuous functions (Clark and Pregibon, 1992). Also, RTs have some advantages over the regression models. First, a model generated by RTs is easier to understand and interpret (Breiman et al., 1984; Torgo, 1997; Edelstein, 1996; Hand et al., 2001). Second, RTs can be used for an alternative approach for regression problems. There have been instances where a decision tree has shown clues to datasets while a traditional linear regression analysis could not clearly indicate them (Breiman et al., 1984). Third, although both approaches handle missing data, RTs handle them better. While regressions omit data that has any missing values automatically, RTs accommodate missing values by using surrogate rules for back up. For instance, users can specify that omit data only when missing values are more than 50 percent. Fourth, RTs provide computational efficiency because they take less time in computation and require less storage (Torgo, 1997). However, RTs also have some limitations. The simple models in the leaves might provide functions that approximate (Torgo, 1997). Perturbations in data could cause instability in RTs. Thus, multiple trees can be generated and selecting one that fits one's objective can be a strategy in generating the best model.

## 6. EMPIRICAL ANALYSIS.

### 6.1 Experimental Approach.

We used the SAS Enterprise Miner (EM) data mining software to generate multiple RTs in order to see if the different RTs offer a consistent message about the relationship between IT investments and firm level productivity. The process of generating each RT involved partitioning the model dataset into Training (R) and Validation (A) based on the traditional data mining approach for supervised learning. For the generation of each RT, we used a stratified sampling approach to partition the dataset. Two variables, YEARNO (the Year of the Observation) and INUM (the Industry - two digit primary SIC level or Sector of the Economy in which a firm operates) are included as stratification variables to ensure that characteristics of both training and validation data sets are close to each other. By varying the value of the Seed parameter in relevant EM *Partition Node* that partitioned the input, we were able to ensure that different Training and Validation datasets were used in the growth and pruning phases of the different RTs. In order to ensure even further variation in our experimentation we varied other parameter values in EM Tree Node, the component of EM that does decision tree induction. We varied *the Splitting Criterion,* and two parameters that are used for pre-pruning, *the Minimum Number of Observations per Leaf*, and *the Observations Required for a Split Search*. Table 3 displays parameter values used in the generation of our original RT1 and 3 additional RTs, and also some summary statistics (i.e. Average Squared Error (ASE) and R-Squared) on RTs.

**Table 3: Data Partition and Parameter Settings**

| | RT1 (Original) | RT2 | RT3 | RT4 |
|---|---|---|---|---|
| **Data Partitioning** | R: 70%<br>A: 30% | R: 70 %<br>A: 30 % | R: 60 %<br>A: 40 % | R: 60 %<br>A: 40 % |
| **Splitting criterion** | F-Test | Variation reduction | F-Test | Variance reduction |
| **Minimum Number of Observations per Leaf** | 20 | 30 | 20 | 30 |
| **Observations Required for a Split Search** | 40 | 60 | 40 | 60 |
| **Model Assessment Measure** | ASE | ASE | ASE | ASE |
| **Selecting Subtree** | Best Assessment Value | Best Assessment Value | Best Assessment Value | Best Assessment Value |
| **ASE (R, A)** | (0.108, 0.103) | (0.113, 0.126 ) | (0.109, 0.112) | (0.118, 0.149) |
| **R-Squared (R, A)** | (0.908, 0.892) | (0.895, 0.899) | (0.901, 0.902) | (0.891, 0.873) |

### 6.2 Results from the Regression Tree Based Approach.

The results from our initial regression tree, RT1 are included in Table 4, which represent a ruleset. Each row represents a rule and the *condition component* column represents the range of values for the relevant input variables for each rule. The *target* column represents the predicted mean values obtained from the *training* and the *validation* datasets for the target variable. The Standard Deviation (SD) is enclosed in parentheses in *Training* column. For example, first rule can be expressed as "if log(Non-IT Labor Expenses) is less then 5.880335 and log(Non-Computer Capital) is less then 5.186935, the predicted mean log(Value Added) is 4.8832 with a standard deviation of 0.7362. 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. As shown in Table 4, the ruleset from our RT generated the fourteen rules. Also, the predicted mean values of the target variable from the training dataset and the validation dataset in Table 4 are very close to each other. This demonstrates the stability of our RT1.

The ruleset described in Table 4 revealed following facts:

- Our RT based analysis agrees with the Econometrics–based analysis and DEA-based analysis in that IT has an impact on productivity of the firm. However, our RT based analysis offers the insight that this impact is not uniform and may not be significant for some cases.
- Non-IT Labor (log L) is the most significant variable for predicting productivity. Our RT-based analysis partitions Non-IT Labor expenses into three intervals such that IT Stock (log T) is relevant as a predictor variable only when Non-IT Labor expenses fall within the interior interval. Thus, IT has an impact on productivity only when Non-IT Labor expenses are within this interior range. The bold amounts in Table 4 represent this range. When the Non-IT Labor (log L) is out of this range, there is no IT impact on the firm's output.
- The mean value for the output is lower when the Non-IT Labor expenses are within the interior range than it is above the range. This indicates that IT is not a factor generating the highest mean

value for the output. Thus, our RT-based analysis provides an evidence for the apparent 'IT Productivity Paradox.'

**Table 4: The Ruleset of RT1 – Sorted by log(L) and mean log(V) for Training**

| Condition Component | | | Target: mean log (V) | | IT Impact |
|---|---|---|---|---|---|
| Non-IT Labor Expenses (log L) | Non-Computer Capital (log K) | IT Stock (log T) | Training / (SD) | Validation | |
| [0.000000, 5.880335] | [0.000000, 5.186935] | Not Selected | 4.8832 (0.7362) | 5.1487 | No |
| [0.000000, 5.880335] | [5.186935, + ∝] | Not Selected | 6.1914 (0.4679) | 6.3707 | No |
| [5.880335, 6.332945] | [0.000000, 8.202245] | Not Selected | 6.6119 (0.2712) | 6.5871 | No |
| [5.880335, 6.906925] | [8.202245, 9.461535] | Not Selected | 7.2442 (0.2972) | 7.2335 | No |
| [5.880335, 6.906925] | [9.461535, + ∝] | Not Selected | 7.6466 (0.1968) | 7.7375 | No |
| [6.332945, 6.709300] | [0.000000, 8.202245] | Not Selected | 6.9347 (0.2312) | 6.9455 | No |
| [6.709300, 6.906925] | [0.000000, 8.202245] | Not Selected | 7.2196 (0.2131) | 7.2451 | No |
| **[6.906925, 7.714730]** | **Not Selected** | **[0.0000, 5.38717]** | **7.7239 (0.2786)** | **7.7260** | **Yes** |
| **[6.906925, 7.714730]** | **[0.000000, 8.513730]** | **[5.38717, + ∝]** | **8.0273 (0.3571)** | **7.8585** | **Yes** |
| **[6.906925, 7.714730]** | **[8.513730, + ∝]** | **[5.38717, + ∝]** | **8.2805 (0.3189)** | **8.3354** | **Yes** |
| [7.714730, 8.176160] | [0.000000, 9.935670] | Not Selected | 8.3724 (0.1983) | 8.2903 | No |
| [7.714730, 9.039965] | [9.935670, + ∝] | Not Selected | 9.1152 (0.2584) | 9.0513 | No |
| [8.176160, 9.039965] | [0.000000, 9.935670] | Not Selected | 8.9071 (0.2682) | 8.8089 | No |
| [9.039965, + ∝] | Not Selected | Not Selected | 10.1877 (0.3862) | 10.3663 | No |

### 6.3. Discussion.

In order to check the validity of our findings based on initial regression tree, RT1, we compared the rulesets generated from three additional trees RT2, RT3, and RT4 that are described in the appendix A. Overall, results from the all three additional RTs are consistent with findings from RT1. The ruleset from each RT indicates that the IT Stock has an impact on the Value-Added (log V) amount when Non-IT Labor (log L) amounts are within the interior ranges. Also, every ruleset from each RT describes that the IT Stock does not play a role in determining the highest value of the Value-Added (V). Thus, results from the additional three RTs confirmed our findings.

We believe that our analysis using the RT technique not only confirmed findings from the previous studies that have examined the impact of IT on firm level productivity, but also revealed additional facts that have not identified from the other studies. In Table 5, we compared our study with previous studies that have used the same dataset on IT productivity research.

**Table 5: Comparison with Previous Research of IT Impact on Organizational Productivity**

| Study | Approach | Results |
|---|---|---|
| Hitt and Brynjolfsson (1996) | Regression Analysis | The effect of IT on productivity is positive. |
| Shao (2000) / Shao & Lin (2001) | Cobb-Douglas and translog Stochastic Production Frontiers / Data Envelopment Analysis (DEA) | IT investments have a favorable total effect on the firm's productive efficiency in the production process. |
| Ko and Bryson (2002) | Regression-Tree technique from the field of Data Mining | 1. IT investments have an impact on productivity of the firm but the impact is not uniform and may even be negligible for some cases.<br>2. IT has an impact only when non-IT labor amounts are within the interior ranges in which IT was selected as a predictor variable in determining productivity.<br>3. The value of productivity is lower when non-IT labor amount is within the range than it is above the range. Thus, IT does not play a role in determining the highest value of productivity.<br>4. Evidence of "IT Productivity Paradox" |

## 7. CONCLUSION.

As organizations have increased investments in IT continuously hoping to improve their organizational profitability, many researchers have tried to estimate the impact of IT on firm level productivity or profitability. In this study, we examined the impact of IT on productivity at the firm level using a RT technique. Overall, our results agree with previous studies in that IT investments have an impact on organizational productivity. However, our RT analysis revealed additional facts that have not identified from previous studies. Our study indicated that Non-IT Labor is the most important variable for predicting productivity but partitions Non-IT Labor investments into three intervals. Only when Non-IT Labor investments are within the interior interval, IT Stock is relevant as predictor variable. Thus, it suggests that IT has an impact on productivity only when Non-IT Labor expenses are within this interior ranges. Even within this range, the impact of IT investments is not uniform. In addition, the value of productivity is lower when the non-IT labor amount is within this range than it is above the range. Thus, IT does not play a role in generating the highest value of productivity and our findings provide an evidence of the IT 'Productivity Paradox."

We also generated additional RTs varying the data partitioning and parameter values and checked the validity of our findings. Results from all three additional RTs confirmed our findings. We believe our RT-based analysis provided additional insights to the IT productivity research.

For future research, determining value of enterprise resource planning (ERP) systems or electronic commerce (EC) on productivity at the firm level would be valuable for organizations since many organizations have invested huge amount of organizational resources.

## ACKNOWLEDGMENTS

We are grateful to Professors Erik Brynjolfsson and Lorin Hitt for kindly sharing their dataset with us, thus facilitating this study.

## REFERENCES

Bharadwaj, A. (2000). "A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation." MIS Quarterly **24**(1): 169-196.

Breiman, L., J. H. Friedman, et al. (1984). Classification and Regression Trees. Belmond, California, Wadsworth International Group.

Brynjolfsson, E. (1993). "The Productivity Paradox of Information Technology." Communications of the ACM **36**(12): 66-76.

Brynjolfsson, E. and L. M. Hitt (1996). "Paradox Lost? Firm-level Evidence on the Returns to Information Systems Spending." Management Science **42**(4): 541-558.

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 Boorks & Software: 377-419.

Dewan, S. and C.-K. Min (1997). "The Substitution of Information Technology for Other Factors of Production: A Firm Level Analysis." Management Science **43**(12): 1660-1675.

Dos Santos, B. and K. Peffers (1993). Firm Level Performance Effects: A Framework for Information Technology Evaluation Research. Strategic Information Technology Management: Perspective on Organizational Growth and Competitive Advantage. R. Banker, Kauffman and M. A. Mahmood. Harrisburg, Idea Group Publishing.

Edelstein, H. (1996). "Data Mining: Exploiting the Hidden Trends in Your Data." DB2 Online Magazine.

Fernberg, P. (1995). "The Changing Workplace: Finding Hope Amid Chaos." Managing Office Technology **40** (5): 14-18.

Han, J. and M. Kamber (2001). Data Mining Concepts and Techniques, Academic Press.

Hastie, Trevor. Tishirani, et al., (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hand, D. Mannila, et al., (2001). Principles of Data Mining. 2001.

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.

Jurison, J. (1997). "Reevaluating Productivity Measures." Information Systems Management: 30-34.

Kim, H. and G. J. Koehler (1995). "Theory and Practice of Decision Tree Induction." Omega **23** (6): 637-652.

King, W. R. (1998). "IT-Enhanced Productivity and Profitability." Information Systems Management: 64-66.

Kivijarvi, H. and T. Saarinen (1995). "Investment in Information Systems and the Financial Performance of the Firm." Information & Management **28**: 143-163.

Lichtenberg, Frank R. (1995). "The Output Contributions of Computer Equipment and Personnel: A Firm-Level Analysis." Economics of Information and New Technology **3** (4): 201-217.

Loveman, G. W. (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.

Mahmood, M. and G. J. Mann (1993). "Measuring the Organizational Impact of Information Technology Investment: An Exploratory Study." Journal of Management Information Systems **10**(1): 97-122.

Rai, A., R. Patnayakuni, et al. (1997). "Technology Investment and Business Performance." Communications of the ACM **40**(7): 89-97.

SAS Institute Inc., The SAS System for Windows, Version 8.2, Cary, NC, 2001.

Shao, B. (2000). Investigating the Value of Information Technology in Productive Efficiency: An Analytic and Empirical Study. The Graduate School. Buffalo, State University of New York.

Shao, B. and W. Lin (2000). "Examining the Determinants of Productive Efficiency with IT as a Production Factor." Journal of Computer Information Systems **41**(1): 25-30.

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.

Shu, W. (2001). "Will the New Economy Emerge as Information Technology Pays Off?" Journal of Association for Information Systems **2**(1): 1-23.

Sircar, S., J. L. Turnbow, et al. (1998). "The Impact of Information Technology Investments on Firm Performance: A Review of the Literature." Journal of Engineering Valuation and Cost Analysis **1**: 171-181.

Torgo, L (1997). Functional Models for Regression Tree Leaves. International Conference on Machine Learning.

Weil, 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.

## APPENDIX A

Three additional RTs have created to check the validity of our findings. The rulesets for each RT are described in Table 6, 7, and 8 as below:

**Table 6: The Ruleset of RT2 – Sorted by log(L) and mean log(V) for Training**

| Condition Component | | | Target : Mean log (V) | | IT Impact |
|---|---|---|---|---|---|
| Labor (log L) | Non-Computer Capital (log K) | IT Stock (log T) | Training (SD) | Validation | |
| [0.000000, 4.635155] | Not Selected | Not Selected | 5.21072 (0.81661) | 5.30834 | No |
| [4.635155, 5.954584] | [0.000000, 6.824960] | Not Selected | 6.10035 (0.39052) | 5.97392 | No |
| [4.635155, 5.954584] | [6.82496, + ∝] | Not Selected | 6.49744 (0.41477) | 6.61088 | No |
| [5.954584, 6.543785] | [0.000000, 8.118965] | Not Selected | 6.71384 (0.26035) | 6.66839 | No |
| [5.954584, 6.902305] | [9.462845, + ∝] | Not Selected | 7.67813 (0.19744) | 7.70282 | No |
| [5.954584, 6.902305] | [8.118965, 9.462945] | Not Selected | 7.26719 (0.25386) | 7.23279 | No |
| [6.543785, 6.902305] | [0.000000, 8.118965] | Not Selected | 7.11914 (0.22914) | 7.08455 | No |
| **[6.902305, 7.420990]** | **Not Selected** | **[0.000000, 5.390322]** | **7.64067 (0.24730)** | **7.65891** | **Yes** |
| **[6.902305, 7.714730]** | **[0.000000, 8.509035]** | **[5.390322, + ∝]** | **8.03293 (0.35623)** | **7.74537** | **Yes** |
| **[6.902305, 7.714730]** | **[8.509035, + ∝]** | **[5.390322, + ∝ ]** | **8.28358 (0.33335)** | **8.35516** | **Yes** |
| **[7.420990, 7.714730]** | **Not Selected** | **[0.000000, 5.390322]** | **7.97664 (0.23620)** | **7.98391** | **Yes** |
| [7.714730, 8.534740] | [0.000000, 9.948815] | Not Selected | 8.47906 (0.27127) | 8.41804 | No |
| [7.714730, 8.534740] | [9.948815, + ∝] | Not Selected | 9.01009 (0.16382) | 9.04068 | No |
| [8.534740, + ∝] | Not selected | Not Selected | 9.63092 (0.55903) | 9.68804 | No |

**Table 7: The Ruleset of RT3 – Sorted by log(L) and mean log(V) for Training**

| Condition Component | | | Target: Mean log (V) | | IT Impact |
|---|---|---|---|---|---|
| Labor (log L) | Non-Computer Capital (log K) | IT Stock (log T) | Training (SD) | Validation | |
| [0.000000, 5.271235] | [5.186935, 8.586980] | Not Selected | 5.92304 (0.51147) | 5.65002 | No |
| [0.000000, 6.204105] | [0.000000, 5.186935] | Not Selected | 5.10048 (0.75576) | 5.07355 | No |
| [0.000000, 6.204105] | [8.586980, + ∝] | Not Selected | 6.96130 (0.31746) | 7.04028 | No |
| [5.271235, 6.204105] | [5.186935, 8.586980] | Not Selected | 6.40629 (0.32876) | 6.39960 | No |
| [6.204105, 6.715340] | [0.000000, 9.365200] | Not Selected | 6.97588 (0.27878) | 6.89633 | No |
| [6.204105, 6.940255] | [9.365200, + ∝] | Not Selected | 7.74252 (0.16986) | 7.64515 | No |
| [6.715340, 6.940255] | [0.000000, 9.365200] | Not Selected | 7.35241 (0.29336) | 7.24015 | No |
| **[6.940255, 7.462740]** | **Not Selected** | **[0.00000, 5.378768]** | **7.67200 (0.26355)** | **7.67256** | **Yes** |
| **[6.940255, 8.009865]** | **[0.000000, 10.05267]** | **[5.378768, + ∝]** | **8.22101 (0.22127)** | **8.11403** | **Yes** |
| **[6.940255, 8.009865]** | **[10.052670, + ∝]** | **[5.378768, + ∝]** | **8.73145 (0.22127)** | **8.71910** | **Yes** |
| **[7.462740, 8.009865]** | **Not Selected** | **[0.00000, 5.378768]** | **8.06870 (0.21151)** | **8.01160** | **Yes** |
| [8.009865, 8.801680] | [0.000000, 9.866430] | Not Selected | 8.73337 (0.32081) | 8.70397 | No |
| [8.009865, 8.801680] | [9.866430, + ∝] | Not Selected | 9.19571 (0.22544) | 9.17264 | No |
| [8.801680, + ∝] | Not selected | Not Selected | 10.05120 (0.54214) | 9.81628 | No |

**Table 8: The Ruleset of RT4 – Sorted by log(L) and mean log(V) for Training**

| Condition Component | | | Target: Mean log (V) | | IT Impact |
|---|---|---|---|---|---|
| Labor (log L) | Non-Computer Capital (log K) | IT Stock (log T) | Training (SD) | Validation | |
| [0.000000, 4.777320] | Not Selected | Not Selected | 5.28037 (0.86486) | 5.38524 | No |
| [4.777320, 5.933865] | [0.000000, 6.824360] | Not Selected | 6.09830 (0.33149) | 6.06655 | No |
| [4.777320, 5.933865] | [6.824360, + ∝] | Not Selected | 6.51625 (0.39907) | 6.53601 | No |
| [5.933865, 6.543785] | [0.000000, 8.073070] | Not Selected | 6.70439 (0.26626) | 6.65415 | No |
| [5.933865, 6.906925] | [8.073070, 9.357415] | Not Selected | 7.25574 (0.25910) | 7.22999 | No |
| [5.933865, 6.906925] | [9.357415, + ∝] | Not Selected | 7.64888 (0.24165) | 7.6453 | No |
| [6.543785, 6.906925] | [0.000000, 8.073070] | Not Selected | 7.11505 (0.23351) | 7.10373 | No |
| **[6.906925, 7.460675]** | **Not Selected** | **[0.0000, 5.37897]** | **7.65327 (0.25846)** | **7.66297** | **Yes** |
| **[6.906925, 7.533715]** | **[0.000000, 9.844750]** | **[5.37897, + ∝]** | **8.01790 (0.38825)** | **7.71943** | **Yes** |
| **[6.906925, 7.970095]** | **[9.844750, + ∝]** | **[5.37897, + ∝]** | **8.62840 (0.29824)** | **8.69331** | **Yes** |
| **[7.460675, 7.970095]** | **Not Selected** | **[0.0000, 5.37897]** | **8.05257 (0.24046)** | **8.01085** | **Yes** |
| **[7.533715, 7.970095]** | **[0.000000, 9.844750]** | **[5.37897, + ∝]** | **8.30747 (0.19743)** | **8.23629** | **Yes** |
| [7.970095, 8.546030] | Not Selected | Not Selected | 8.75222 (0.30787) | 8.79299 | No |
| [8.546030, + ∝] | Not selected | Not Selected | 9.67009 (0.54542) | 9.73461 | No |