Using Regression Analysis to Address Methodological and Theoretical Issues in IT Cost Benchmarking

Vlad Krotov
Murray State University
vkrotov@murraystate.edu

Abstract: The practice of IT cost benchmarking using IT managerial control ratios suffers from a number of methodological and theoretical issues. These issues arise from the following three assumptions: (1) the functional form of the relationship between a numerator and denominator used in an IT managerial control ratio is strictly proportional; (2) the nature of the underlying probability distribution of industry samples of IT managerial control ratios is normal and (3) position of an organization in relation to an industry norm can be unambiguously interpreted. If these assumptions are not met, then determining and interpreting a company's position with respect to IT costs in relation to industry averages and other companies within the industry is subject to some ambiguity. This paper uses empirical tests and theoretical arguments to show that these three assumptions may not hold true in practice. It is then argued that regression-based analysis of IT costs can be used to address these issues. Further theoretical and empirical work is needed to develop these regression models so that practitioners can have a reliable and valid method for estimating and interpreting their company's position with respect to IT costs in relation to an industry norm. At the minimum, practitioners should not rely on IT cost benchmarking for setting their IT budgets without taking into account the methodological and theoretical issues.

Keywords: IT costs, IT spending, IT budget, justification, benchmarking, methodology, theory

1. Introduction

Managing and justifying information technology (IT) costs is one of the most pressing issues facing senior technology leaders (Kappelman et al, 2014). This issue becomes especially acute during economic downturns, when senior executives look at the IT function for cost minimization opportunities. Because of that, the justification of IT costs has received substantial attention from both practitioners and researchers (Schryen, 2013). This interest has led to a myriad of IT cost evaluation and justification methodologies (Irani, 2002). One popular method of justifying IT spending is benchmarking using IT managerial control ratios (e.g. annual IT budget divided by annual revenue) (Krotov and Ives, in press). This practice involves comparing IT costs of one organization against industry averages or companies known for their "best in class" IT organizations (Doll et al, 2003). Determining the relative standing of one's IT organization in relation to the industry norm or other companies can help CIOs justify current or additional investments in IT or, if required, cost-cutting initiatives depending on whether a firm or business unit is below or above the industry mean, a sister organization, or a company deemed to be "best in class".

Validity of IT cost benchmarking rests on the following three assumptions: (1) the functional form of the relationship between a numerator and denominator used in an IT managerial control ratio is strictly proportional; (2) the underlying probability distribution of industry samples of IT managerial control ratios is normal and (3) position of an organization in relation to an industry norm can be unambiguously interpreted. If these assumptions are not met, then determining and interpreting a company's position with respect to IT costs in relation to industry averages and other companies within the industry is subject to some ambiguity. This can potentially lead to sub-optimal capital allocation decisions with respect to IT. This paper aims to educate both researchers and practitioners about these methodological and theoretical issues related to IT cost benchmarking using IT managerial control ratios and propose a regression analysis approach as a possible way of addressing these issues.

The rest of this paper is structured as follows. First, the paper educates the reader about IT managerial control ratios – parsimonious measurement instruments that are used in cross-sectional benchmarking of IT costs. To accomplish this goal, the paper provides a formal definition of an IT managerial control ratio and lists commonly used numerator and denominator variables. Second, the paper discusses how these ratios are used in IT costs benchmarking. The practice of IT cost benchmarking is first described from a practical standpoint and then formalized for subsequent analysis of this practice. Third, the paper discusses various methodological and theoretical issues related to the practice of IT cost benchmarking. These issues include: (1) the functional form of the relationship between a numerator and denominator used in an IT managerial control ratio; (2) the

Reference this paper as: Krotov V, "Using Regression Analysis to Address Methodological and Theoretical Issues in IT Cost Benchmarking" *The Electronic Journal Information Systems Evaluation Volume 19 Issue 1 2016*, (pp22-35), available online at www.ejise.com

nature of the underlying probability distribution of industry samples of IT managerial control ratios and (3) plausibility of numerous theoretical explanations as to why a particular organization has a ratio value below or above the industry norm. As it is argued in the paper, these issues can potentially create ambiguities in determining a company's position in relation to an industry norm, interpreting a company's position in relation to an industry norm and interpreting the value of an IT managerial control ration in relation to other firms in the industry. Fourth, it is proposed that regression-based benchmarking should be used to overcome these issues. It is argued that regression-based analysis of IT costs can be used to address the three problems discussed in the paper. The paper concludes with recommendations for practitioners in the light of the issues discussed in this paper and urges researchers to develop and empirically validate theoretically sound regression models that can potentially improve reliability and validity of IT cost benchmarking.

2. IT Managerial Control Ratios

2.1 Definition

IT managerial control ratios are the heart of the practice of IT cost benchmarking. The ratios are used as parsimonious quantitative measurement instruments for measuring and comparing organizational IT costs of companies of various sizes. Mathematically, an IT managerial control ratio of a firm f can be expressed as the relationship with IT cost variable (y_f) in the numerator and organizational size variable (x_f) in the denominator (Nikkinen and Sahlstrom, 2004):

$$\beta_f = \frac{y_f}{x_f}$$

Where:

$y_f =$ Annual IT spending variable

$x_f =$ Organizational size variable

$\beta_f =$ Organizational IT spending relative to size

IT managerial control ratios accomplish two related goals: (1) they serve as measures of organizational IT spending relative to size and (2) they control for systematic size differences among companies so that comparisons can be made. The two goals that IT managerial control ratios accomplish are interrelated. Control for size differences is necessary so that comparison can be made among companies of different sizes with respect to IT spending. Similarly, measuring IT spending within an organization via a ratio with organizational size variable in the denominator controls for size differences among companies.

2.2 Types of IT Managerial Control Ratios

A variety of numerators and denominators have been used in practice and research to measure IT costs relative to organizational size (see Table 1). Each of the numerator and denominator variables has certain advantages and disadvantages mostly related to the reliability of these measures (Krotov, in press).

Table 1: Numerator and Denominator Variables

Numerator Variables	**Denominator Variables**
Commonly used: • IT expense (and its categories: hardware, software, personnel, etc.) • IT budget (and its categories) • Less commonly used: Less commonly used: • Operating IT budget • IT stock (and its categories)	Commonly used: • Revenues • Operating expense • Employees Less commonly used: • Assets • Selling, general & administrative expense

While a variety of ratios have been used in practice and research to measure organizational IT spending, the ratio of annual IT budget to annual revenue (ITB/R) of an organization is one of the most commonly used IT cost benchmarks. All major IT research firms, such as *Forrester Research*, *Gartner Research*, and *Information Week*, use it (either as a mean or a medium value, see Table 2).

Table 2: Ratios in Benchmarking

Organization	Ratio(s)	Industry Standard
Forrester	ITB/R	Median
Gartner	ITB/R; ITB/OE	Mean
Information Week	ITB/R	Mean, Median

Two factors explain the popularity of this ratio. First, it has been argued that this ratio closely corresponds to executive economic thinking (Koch ,2006). It is relatively easy for a senior technology executive to provide an accurate value of their annual IT budget (as opposed to trying to estimate annual IT expense, for example). Also, it is more natural for an executive to think about costs in terms of proportion of revenues rather than as a proportion of operating expenses or on a per employee basis. All this makes data collection for this ratio easier and more accurate for IT research firms administering these surveys. Second, unlike other ratios, this ratio is not sensitive to the labor-capital mix of organizations. For example, a company substituting human labor with machines may have an unusually high ratio of IT spending per employee.

3. IT Cost Benchmarking

3.1 The Practice of IT Cost Benchmarking

IT managerial control ratios are commonly used in cross-sectional comparison of IT costs where an IT managerial control ratio of one organization compares to industry averages (Doll et al, 2003). Ratio data are collected on an annual basis by IT market research firms, such as *Forrester Research*, *Gartner Research*, *Information Week*, etc. These firms obtain the data either through self-reports from senior IT managers or by asking senior IT managers for their dollar IT expense and then using external data sources (e.g. Compustat) to obtain values for organizational size variables (e.g. revenue, operating expenses, number of employees). The data are obtained either through random sampling across industries or by obtaining data from a preexisting set of companies known for their "best in class" IT organizations (e.g. in the case of *Information Week 500* survey).

After ratio data is gathered, it is categorized into industry groups. Subsequently, "averages" are computed for each of the industry groups. These averages are viewed as "industry norms". The most commonly used averages are the median ratio **(** $\tilde{x}$ **)** and the mean ratio **(** $\bar{x}$ **).** Senior IT managers compare these averages to the values of IT managerial control ratios of their own IT organizations in order to benchmark their costs against the industry norm. Depending on an organization's position relative to the industry standard, senior IT managers may adjust their budget levels to be closer to the industry norm (Kobelsky et al, 2008), operating on the assumption that a ratio that is too high or too low may indicate that organizational IT costs are sub-optimal – something that may attract unwelcome senior management attention.

3.2 Formalization of Cross-Sectional Benchmarking

The practice of cross-sectional IT cost benchmarking using IT managerial control ratios discussed above can be formalized as follows. First, an industry standard is set by a line with a slope $\beta_s = \bar{x}$ or $\beta_s = \tilde{x}$ (see Figure 1). In other words, a standard within an industry can be represented by a straight line with a slope equal to either mean or median ratio. An implicit assumption here is that companies, regardless of their size, should have a constant ratio of IT cost to size. And this constant value is derived from the mean or median value of IT managerial control rations within an industry. Mathematically, the "standard line" is described by the following equation:

$$y = \beta_s x$$

Secondly, values of ratios of individual organizations (A and B in this case) are plotted on the *xy* coordinate space. Thus, A's and B's positions relative to the industry standard set by the line described above is given by points $A(x_a; y_a)$ and $B(x_b; y_b)$. In this particular example, A's position is interpreted as being above the industry norm and B's position is interpreted as being below the industry norm, since:

$$\frac{y_a}{x_a} > \beta_s \text{ and } \frac{y_b}{x_b} < \beta_s$$

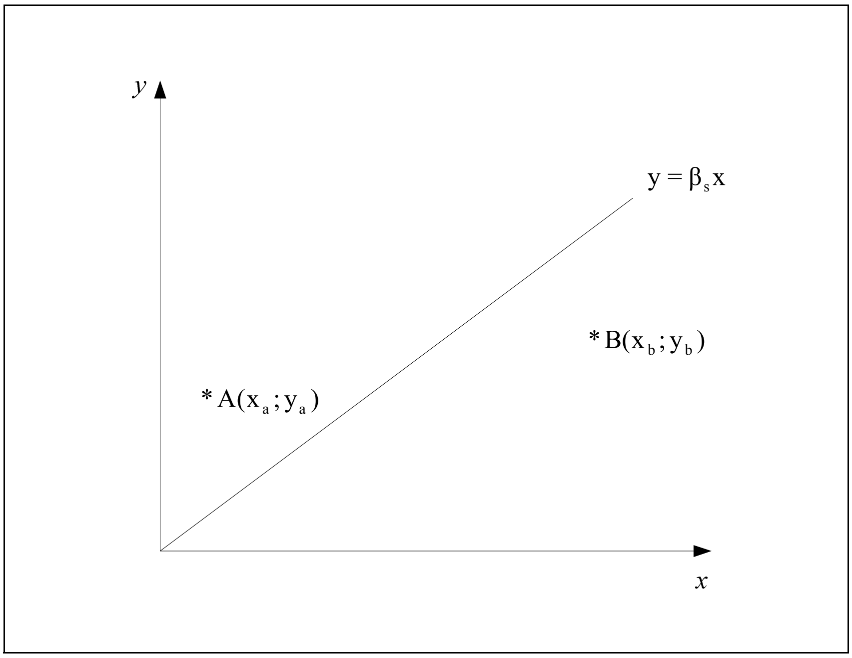

Figure 1 - Formalization of Benchmarking

4. Methodological Issues in IT Cost Benchmarking

The practice of IT cross-sectional IT cost benchmarking described and formalized in the previous section can potentially lead to wrong inferences in the light of several important methodological issues associated with the practice. First, determining a company's position (e.g. above or below) relative to the industry may be quite ambiguous in the absence of a strong theory regarding the functional form of the relationship between a numerator and denominator variable used to form a managerial control ratio. Second, interpreting a company's position in relation to other firms in the industry can be problematic if the industry sample of IT managerial control ratios is not normally distributed. Each of these issues are discussed in more detail below.

4.1 Determining Position of Individual Companies Relative to the Industry Norm

When performing a cross-sectional comparison of values of the individual IT managerial control ratios against an industry norm (e.g. mean or median) it is implicitly assumed that the relationship between a numerator and a denominator variable is strictly proportional (see Figure 1). Companies with ratios above or below this "industry line" are assumed to be deviating from the norm. The relation between numerator and denominator of IT managerial control ratios is strictly proportional when y is a constant percentage of x, with the percentage being given by β (with β representing the value of the industry norm). Graphically, this relationship can be represented by a line with a slope β (see Figure 1).

While it's plausible that the relationship between a particular numerator and denominator variable is strictly proportional (e.g. companies with various levels of revenues tend to allocate a fixed percentage of revenues to IT budget), at least some of the pairs of numerator and denominator variables may deviate from proportionality. For example, continuing the example of a ratio with an annual IT budget in the numerator (represented by y) and revenues in the (represented by x), the relationship between the two variables may deviate from strictly proportional in the following ways (Whittington, 2007):

1. The presence of a constant term β_0 in the relationship between x and y:

$$y = \beta_1 x + \beta_0$$

2. Non-linear nature of the relationship between x and y, for example (Gurbaxani et al, 1997):

$$y = ax^{\beta}$$

The resulting bias from these two possible forms of deviation from proportionality is discussed below.

4.1.1 Presence of a Constant Term

A hypothetical situation when the relationship between the numerator and denominator includes a constant term is depicted in Figure 2. In this example, the industry norm is set by a proportional line $y = \beta_n x$, where $\beta_n = \text{mean}$ or $\beta_n = \text{median}$. The true relationship is reflected by the line $y = a + \beta_t x$, which includes a constant term a. This type of functional relationship between a numerator and denominator is possible if the provision of IT services includes a fixed cost component (represented by a). One can see that in this particular example firm A is classified as being above the industry norm based on the average (e.g. mean or median) value of an IT managerial control ration within an industry. At the same time, the true relationship between the numerator and denominator suggests that the position is, in fact, below the industry norm. Similarly, firm B is classified as being below the industry norm, while the relationship suggests that the position is, in fact, above the industry norm. Obviously, if one were to measure deviations from the industry norm (e.g. in terms of the percent of current budget) one would get estimates of deviations which are contradictory to the relationship between the numerator and denominator.

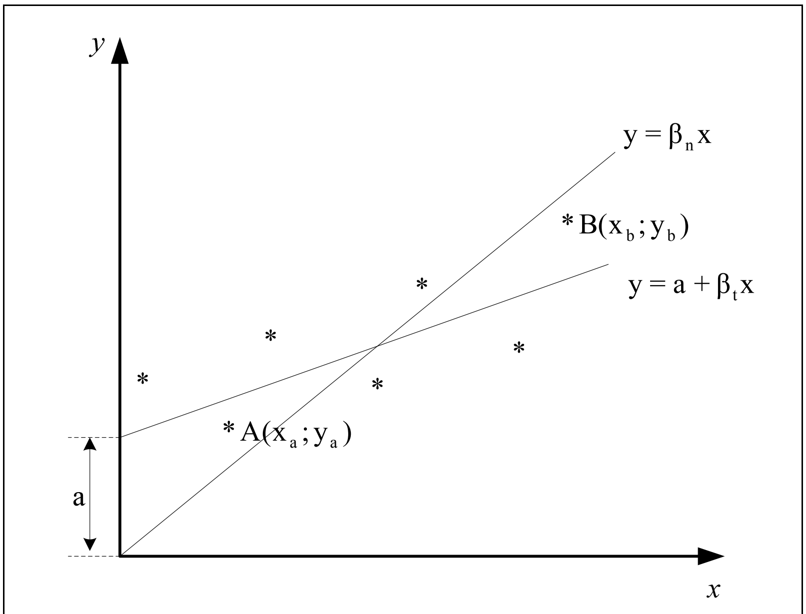

Figure 2: Presence of a Constant Term

4.1.2 Non-Linear Relationship

Another possible form of deviation from proportionality is a curvilinear relationship between the numerator and denominator. For example, in the case of presence of economies of scale in the production of IT services, the relationship between the numerator and denominator of an IT managerial control ratio may be characterized by the following relationship (Gurbaxani et al, 1997):

$$y = ax^{\beta}, \beta < 1$$

In this case, the industry standard will be set by a line implying a proportional relationship ($y = \beta_n x$) while the true relationship is curvilinear ($y = ax^{\beta}, \beta < 1$). Because of that organization $A(x_a; y_a)$ will be incorrectly classified as having a ratio above the industry norm, while the relationship between the numerator and denominator suggests that organization is below the norm set by the true relationship between a numerator and a denominator. Similarly, organization $B(x_b; y_b)$ will be classified as being below the industry norm, while, in reality, it is above the true industry norm (see Figure3).

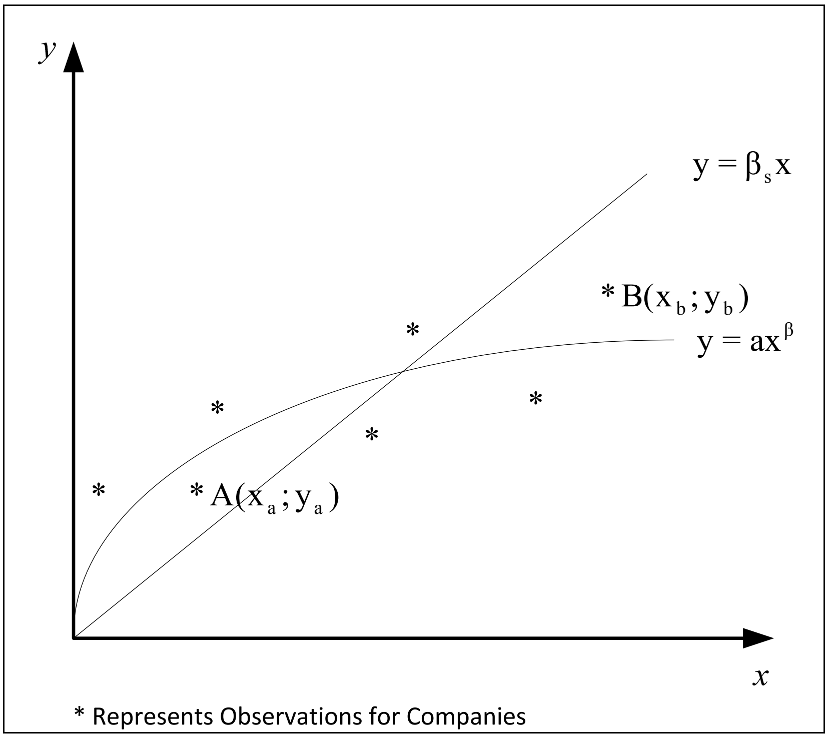

Figure 3: Curvilinear Relationship

4.2 Interpreting Position of a Company in Relation to Other Firms

Another issue in cross-sectional benchmarking of IT costs using IT managerial control ratios arise when the underlying probability distribution of an industry sample of IT managerial control ratios is not distributed normally (Barnes, 1982). The underlying probability distribution may deviate from normal either via lack of symmetry (skewness) or "peakedness" (kurtosis).

Symmetric probability distribution is important for benchmarking for several reasons. If a probability distribution is not symmetric, then the mean cannot be viewed as an adequate representation of a norm within an industry sample (Lev and Sunder, 1979). This is because the mean value of a skewed distribution does not divide an industry sample into two equivalent groups. For example, in the case of a positively skewed probability distribution, there will be more cases below the mean than above the mean.

Some practitioners suggest that in the case of a skewed probability distribution, the median ratio represents a better standard against which ratios can be compared (Bartels, 2007). There are several problems with using the median. First of all, in the case of skewed distribution the median ratio may not represent modal class - that is a class of the most typical values within an industry sample. Secondly, equal deviations from the median in different directions are likely to be associated with different probabilities. Thus, while the median ratio may be a somewhat better representation of an industry norm, the probability distribution still has to be taken into account for correct interpretation of a company's position relative to the norm set by the median.

At least in the context of setting an industry norm, "peaked" yet symmetric probability distribution of IT managerial control ratios is considered to be a lesser problem compared to that of an asymmetric probability distribution. For a peaked yet symmetric distribution the industry norm can be unambiguously defined with an average value. The problem with a peaked distribution is that the probability associated with a particular ratio cannot be approximated with a normal probability distribution. Consequently, for a distribution with a high positive kurtosis (that is a distribution with a high level of "peakedness") minor deviations from the norm in terms of the ratio value will be associated with a major deviation in frequency (probability) associated with a particular ratio. Thus, being slightly below or above the mean will put a company into a category of extremely high or extremely low IT spenders - contrary to what a user of benchmarking services may conclude (i.e. small deviations from the norm may be interpreted by the user as an evidence that his or her organizational IT spending is roughly in line with the industry norm).

5. Empirical Test of the Methodological Issues

5.1 Ratio Investigated

The methodological issues discussed in the previous section are investigated using a sample of the ratio of IT budget to revenue (ITB/R). While the choice of the ratio is dictated by data availability, it can be argued that the ITB/R ratio is the most commonly used ratio in IT spending benchmarking (see Table 2). Moreover, this ratio is commonly used in information systems research (Bharadwaj et al, 1999; Kobelsky et al. 2008; Mitra and Chaya, 1996). Thus, the properties of the ratio uncovered through this study will be of relevance to both practitioners (senior IT managers and IT industry analysts) as well as IS researchers. Second, other denominators commonly used in IT spending benchmarking (such as total operating expense and total number of employees) show a high degree of correlation in the dataset used in the study. IT managerial control ratios formed with identical numerators and highly correlated denominators are likely to exhibit similar theoretical and empirical properties (Lev and Sunder 1979). Therefore, properties uncovered through the analysis of the ITB/R ratio are likely to be generalizible to other ratios used in IT spending benchmarking, such as IT budget as a percentage of operating expense (ITB/OE) or IT budget per employee (ITB/EMP).

5.2 Data Source

Data on ITB/R ratio was obtained from an annual survey of senior IT executives conducted by an IT research firm. The survey contains responses from a random sample of approximately 3000 organizations from Europe, Asia and North America. The goal of the original survey was to obtain data on organizational IT spending (measured by the ITB/R ratio) as well as data on other aspects of an IT function, such as vendor preferences, top issues facing CIOs, plans for the upcoming year, etc. Ratio data was obtained through self-reports from senior IT managers.

5.2.1 Dataset for Testing Normality

Since differences in business environments among countries may influence ratio distribution (Nikkinen and Sahlstrom, 2004), only organizations from the United States were included in the dataset. This resulted in a sub-sample of N=795 organizations (the largest sub-sample in the dataset). Moreover, since grouping ratio data into industry groups reduces data variability, the dataset of American companies was clustered into 17 industry groups based on an industry classification code supplied with the dataset. While academic studies often recommend 4-digit SIC code for classifying companies' intro industries, IT spending benchmarking services typically use a classification scheme parallel yet less stringent than the 4-digit SIC classification. The type of the industry classification scheme used in this data set is common to all major IT spending benchmarking services (Bartels, 2007) and has been used in IS research (Lee and Bose, 2002) .

5.2.2 Dataset for Testing Proportionality

For modeling the relationship between IT budget and organizational revenues, the sample of 795 US companies was matched with data from Compustat North America. Wharton Research Data Services (WRDS) database was used to access data from Compustat. Data were matched using each company's name and 4-digit SIC code. A match was found for 96 organizations.

Three variables were obtained from the two datasets: ITBUDGET, REVENUE, and INDUSTRY. Definitions of the variables, the procedure used to derive them, as well as source of the variables are provided in Table 3.

Table 3: Description of Variables Used in Proportionality Analysis

Variable	Description	Source
ITBUDGET (USD)	Total annual IT budget. Includes expensed new investments and depreciated expenses. The dollar value of IT budget was obtained by multiplying 2006 revenues by the ITB/R ratios reported in the IT spending survey supplied by the IT research company. 2006 revenues were used in calculation because the IT budget survey was administered at the end of 2006, so respondents could only use 2006 revenues as a reference point. A similar procedure for matching survey and financial data and deriving dollar values of IT budget categories was employed by (Brynjolfsson & Hitt 1996).	Survey; Compustat North America (WRDS)
REVENUE (USD)	2006 Revenues, as reported in Compustat North America (WRDS)	Compustat North America (WRDS)
INDUSTRY (Dummy Variable: 0 – Service Sector; 1 – Production Sector)	Similarly to (Gurbaxani et al, 1997), this study controls for industry effect by splitting organizations into production or service sectors. The categorization was done using industry classification provided in the dataset.	Survey

Descriptive statistics for the variables described in Table 3 are provided in Table 4.

Table 4: Descriptive Statistics for Variables Used in Proportionality Analysis (USD, millions)

Variable	**N**	**Min**	**Max**	**Mean**	**Med.**	**S. D.**
ITBUDGET	96	0.0078	3152.75	180.55	27.88	479.90
REVENUE	96	0.78	91658	5450.04	784.013	14448.51

In order to model a non-linear relationship between ITBUDGET and REVENUE, the two variables were transformed using natural logarithm. Descriptive statistics for the log-transformed variables can be found in Table 5.

Table 5: Descriptive Statistics for Variables Used in Proportionality

Variable	**N**	**Min**	**Max**	**Mean**	**Med.**	**S. D.**
LNITBUDGET	96	-4.85	8.06	3.24	3.32	2.12
LNREVENUE	96	-0.24	11.42	6.76	6.66	2.02

5.3 Testing Normality of Probability Distribution

Normality was tested using the Kolmogorov-Smirnov test and Shapiro-Wilks test. The reason the two tests are used together is that Kolmogorov-Smirnov test is believed to perform better for large samples, while Shapiro-Wilks test is believed to perform better for smaller samples. The test statistics as well as significance levels were computed using SPSS software.

5.4 Exploring the Functional From of the Relationship

Proportionality of the relationship between the numerator (ITBUDGET) and denominator (REVENUE) were tested using the procedure recommended by Whittington (1980) and the modeling approach to economic analysis of IT budgets suggested by Gurbaxani et al (1997). First, the following model (further referred to as Model 1 or "base-line model") was estimated using the OLS procedure:

$$\text{ITBUDGET} = \beta_{11}\,\text{REVENUE} + \beta_{21}\,\text{INDUSTRY} + \varepsilon_1$$

Model 1 assumes a simple proportional relationship between IT BUDGET and REVENUE. Thus, Model 1 constitutes a base-line model against which alternative specifications can be compared. Two criteria were used to assess the adequacy of the model. The first criterion is the significance of the term attached to REVENUE (β_1). The second criterion is the overall model fit, as measured by adjusted R-squared.

Secondly, the following model is estimated (Model 2):

$$\text{ITBUDGET} = \beta_{02} + \beta_{12}\,\text{REVENUE} + \beta_{22}\,\text{INDUSTRY} + \varepsilon_2$$

Model 2 contains a constant term, β_{02} , in the relationship between ITBUDGET and REVNUE. Two criteria were used to assess the extent to which Model 2 provides a better specification compared to Model 1. The null hypothesis of Model 2 including a constant is given by:

$$H_0: \ \beta_{02} = 0$$

Thus, the statistical significance of the intercept term would suggest the presence of an intercept in the relationship between ITBUDGET and REVENUE (Whittington, 1980). In addition to observing significance of the constant term β_{02} , the extent to which Model 2 provides a better fit compared to Model 1 was tested. Since Model 1 and Model 2 are nested models, the following formula can be used to formally test improvement in fit (Lattin et al, 2003):

$$F = \frac{(R_f^2 - R_r^2)/(df_f - df_r)}{(1 - R_f^2)/df_f},$$

Where:

R_f^2 = R-squared for the full model (Model 2)

R_r^2 = R-squared for the restricted model (Model 1)

df_f = Degrees of freedom associated with the full model

df_r = Degrees of freedom associated with the restricted model

After estimating Model 1 and Model 2, a non-linear model (Model 3) suggested by Gurbaxani et al. (1997) was estimated:

$$\text{ITBUDGET} = \beta_{03}\,\text{REVENUE}^{\beta_{13}}\,\varepsilon_3$$

Where:

β_{03} = Constant

β_{13} = Elasticity coefficient

ε_3 = Error term

Since Model 3 is non-linear, it cannot be estimated using a linear equation. However, the model can be transformed into a linear form by taking a logarithm of both sides and adding a dummy variable (INDUSTRY) to control for industry effects:

$$\text{LNITBUDGET} = \beta_{03} + \beta_{13}\,\text{LNREVENUE} + \beta_{23}\,\text{INDUSTRY} + \varepsilon_3$$

The null-hypothesis of linearity under this specification is given by:

$$H_0: \ \beta_{13} = 1$$

The null-hypothesis of linearity is tested using Student's t-test:

$$t = \frac{\beta_0 - 1}{SE_{\beta_0}}$$

Since Model 3 uses log-transformed variables, R-squared could not be used for comparing fit with Model 1 (Dougherty, 2007). Instead, standardized residuals were used to compare fit of the two models. In order to make the comparison, standardized residuals from Model 1 and Model 3 were saved as variables and two normality tests (Kolmogorov-Smirnov test and Shapiro-Wilks test) were applied to standardized residuals to compare fit between Model 1 and Model 3.

5.5 Discussion of Results

5.5.1 Test of the Distribution Assumption

The null-hypothesis of normality was rejected by Kolmogorov-Smirnov test for all industries at α = 0.001 significance level; the null-hypothesis of normality was rejected by Shapiro-Wilk's test for 16 industries at α = 0.001 significance level and at α = 0.01 for one industry (see Table 6). The results provide strong evidence that industry samples of IT managerial control ratios may not be normally distributed, resulting the previously discussed issues in interpreting a company's position relative to the industry norm (e.g. relative to the mean value of IT budget to revenue in an industry sample)

Table 6: Kolmogorov-Smirnov and Shapiro-Wilk Tests

Industry	**N**	**Kolmogorov-Smirnov**		**Shapiro-Wilk**	
		Stat.	**df**	**Stat.**	**df**
1. Production	37	.300***	37	.583***	37
2. Consumer Products	39	.226***	39	.770***	39
3. Chemicals and Petroleum	32	.248***	32	.721***	32
4. High-Tech Products	32	.235***	32	.712***	32
5. Industrial Products	41	.252***	41	.762***	41
6. Retail	59	.248***	59	.743***	59
7. Wholesale	43	.267***	43	.630***	43
8. Transportation and Logistics	37	.302***	37	.635***	37
9. Professional Services	75	.248***	75	.850***	75
10. Construction and Engineering	45	.259***	45	.597***	45
11. Media, Entertainment, and Leisure	50	.227***	50	.813***	50
12. Utilities	26	.260***	26	.697***	26
13. Telecom Carriers	19	.326***	19	.826**	19
14. Financial Services	66	.243***	66	.852***	66
15. Insurance	42	.258***	42	.787***	42
16. Public Services	81	.219***	81	.795***	81
17. Government	71	.228***	71	.821***	71

*** Significant at $\alpha = 0.001$; ** Significant at $\alpha = 0.01$

5.5.2 Test of the Specification Assumption

Estimation of Model 1 produces a high value of adjusted R^2=0.754, which suggests good fit with the proportional relationship implied by Model 1 (see Table 7). The model is significant with F = 148.493 and p < 0.001. The coefficient β_1 associated with REVENUE is estimated at 0.028 (which suggest that companies in the service sub-sample spend 2.8% of their revenues on IT) is highly significant. The coefficient β_2 associated with INDUSTRY is statistically indistinguishable from zero.

Table 71: OLS Estimation of Model 1

ITBUDGET= β_1 REVENUE + β_2 INDUSTRY + ε			
Parameter	**Coeff.**	**S.E.**	**t-value**
β_1 (REVENUE)	.028	.002	16.749***
β_2 (INDUSTRY)	-10.010	40.336	-.248
SSE = 250.108 Adj. R^2 = 0.754 F = 148.493*** p < 0.001 N = 96			

*** Significant at $\alpha = 0.001$

Constant β_o is not statistically significant (see Table 8). This suggests the absence of a constant term in the relationship between ITBUDGET and REVENUE. Moreover, adding a constant term to the equation estimated in Model 1 does not result in improvement in fit: adjusted R-squared drops from R^2 =0.754 (Model 1) to R^2 =0.725 (Model 2).

Table 8: OLS Estimation of Model 2

ITBUDGET= $\beta_o + \beta_1$ REVENUE + β_2 INDUSTRY + ε			
Parameter	**Coeff.**	**S.E.**	**t-value**
β_o (CONSTANT)	44.895	34.816	1.290
β_1 (REVENUE)	.028	.002	15.887***
β_2 (INDUSTRY)	-51.405	51.440	-0.99
SSE = 249.231 Adj. R^2 = 0.725 F = 126.375*** N = 96			

*** Significant at α = 0.001

Estimation of Model 3 (see Table 9) results in a small improvement in R^2 relative to Model 1. Adjusted R^2 improves from R^2 = 0.754 (Model 1) to R^2 = 0.757 (Model 3). All parameters of the model and the model itself are statistically significant. Test of linear restriction of β_1 =1 (organizational revenues being proportional to size) produces t-statistics of -1.648 (93 df). The associated p-value is 0.051 (one-tailed). Thus, the null hypothesis of proportionality can be rejected if α=0.10.

Table 9: OLS Estimation of Model 3

LNITBUDGET= $\beta_o + \beta_1$ LNREVENUE + β_2 INDUSTRY + ε			
Parameter	**Coeff.**	**S.E.**	**t-value**
β_o (CONSTANT)	-2.750	0.395	-6.958***
β_1 (LNREVENUE)	0.911	0.054	16.996***
β_2 (INDUSTRY)	-0.417	0.219	-1.905†
SSE = 0.766 Adj. R^2 = 0.757 F = 148.893*** Sig. < 0.001 N = 96 Test of linear restriction of β_1 =1 (organizational revenues being proportional to size) produces t-statistics of -1.648 (95 df). The associated p-value is 0.051 (one-tailed). Thus, the null hypothesis of proportionality can only be rejected at α=0.10			

*** Significant at α = 0.001

† Significant at α = 0.10

Since a formal significance test for improvement in fit cannot be used (due to Model 3 parameters involving log-transformed variables), Kolmogorov-Smirnov test was used to test normality of residuals (see Table 10). The test reveals that Model 3 may provide a better fit with the data, since, unlike Model 2, residuals of Model 3 are normally distributed.

Table 10: Normality of Residuals

Variable	**N**	**K-S Stat**
RES_MODEL1	96	3.408***
RES_MODEL2	96	3.219***
RES_MODEL3	96	.758

***Significant at α = 0.001

Collectively, the results indicate that the relationship between the numerator and denominator in the sample of IT managerial control ratios (ITB/R) used in this study is approximated best using a non-linear (exponential) functional form. This means that, other things being equal, larger companies are likely to have smaller values of IT managerial control ratios. Thus, size may not be adequately controlled for when these ratios are used. Moreover, this functional form of the relationship is different from the one assumed by the benchmarking practice. As it was discussed previously, the practice of benchmarking assumes a strictly proportional relationship between a numerator and denominator. Moreover, it is also assumed that a mean or median ratio adequately approximates this proportional relationship. Thus, ambiguities in deciding whether a particular company is below or above the industry norm may arise (as it was illustrated in the previous sections).

6. Theoretical Issues in IT Cost Benchmarking

Even if a company's true position (e.g. above or below) an industry norm is unambiguously established and probability associated with this position is determined using a normal curve, interpretation of the value of an organization's IT managerial control ratio is subject to some theoretical ambiguity. The problem is that there can be a number of often orthogonal theoretical explanations as to why a particular organization is below or above the industry norm with respect to a particular ratio (see Table 11). Therefore, the validity of a decision to increase or decrease IT spending depends on correctness of understanding as to why a particular organization is below or above the norm with respect to IT spending.

Table 11: Alternative Theoretical Explanations of a Position Relative to an Industry Norm

Below the Norm	**Above the Norm**
• Underspending on IT	• Overspending on IT
• High efficiency in utilizing IT assets	• High reliance on IT in operations or strategy
• Low reliance on IT in operations or strategy	• IT infrastructure build-up or upgrade
• Mature IT infrastructure	

There are several reasons as to why a particular organization may have an IT spending ratio below the industry norm. The first possible interpretation could be that an organization is not spending enough on IT (underspending). This can potentially result in technological obsolescence of the organizations current IT infrastructure and lead to competitive disadvantage. In this case, the organization needs to increase its IT spending. On the other hand, it could be that an organization is below the industry norm due to superior efficiency in utilizing IT resources. This can be a result of the company being in possession of superior managerial IT skills (Mata et al, 1995). Thus, below average IT spending may be accompanied by an adequate quality of IT services (DeLone and McLean 2003). In this case, there may be no need to increase the budget. Alternatively, the position below the industry can be explained by the fact that an organization does not make a substantial use of IT in its strategy and operations. For example, the organization may be using highly personalized customer service approach as a part of its differentiation strategy (as opposed to relying on E-Commerce). In this case, increasing IT budget may not be justified. Finally, an organization may have a mature IT infrastructure. Naturally, this organization is likely to spend less on IT compared to organizations that are building up or upgrading their IT infrastructure. In this case, no adjustment to IT spending may be necessary.

Similarly, there can be several alternative explanations as to why an organization's IT spending ratio is above the industry norm. An organization can be above the industry norm with respect to IT spending if the organization's IT resources are not utilized efficiently. This can potentially lead to a case of the so-called "productivity paradox" (Roach, 1991), where increased IT spending does not result in any improvement in organizational performance. In fact, this may lead to deterioration in organizational performance as unnecessary IT spending may cut into the organization's bottom line. In this case, it may be necessary to initiate cost-cutting initiatives in relation to IT. Alternatively, an above average IT spending may be a result of an organization's heavy reliance on IT in its operations and strategy. For example, an organization may be using IT as a part of its IT-enabled strategic initiative, aimed at improving its competitive position (Piccoli and Ives 2005). Finally, an organization may be at the stage of IT infrastructure build-up or upgrade. In this case, it is natural that the organization's IT spending is above the industry norm.

7. Regression Analysis as an Approach to Benchmarking

In the light of the methodological and theoretical issues discussed in the article may consider adopting a regression-based approach to IT cost benchmarking (Whittington, 2007). In its simplest form, the relationship between a numerator (y) and a denominator (y) of an IT managerial control ration can be estimated via the following regression model:

$$y = \hat{\beta}\, x + \hat{a}$$

After the regression line is estimated, deviation from the relationship can be measured by residuals, where the residual of an organization is given by:

$$e_i = \hat{y}_i - y_i$$

Where:

e_i = Residual of an i-th firm

$\hat{y}_i$ = Estimated value of variable of interest (e.g. IT budget)

y_i = Observed value of a variable of interest (e.g. IT budget)

One advantage of the regression approach is that regression can be used to estimate the true structural relationship between x and y, since a regression model can allow for an intercept term and a non-linear relationship between numerator and denominator. Secondly, a regression model can include a number of control variables to rule out alternative theoretical interpretations as to why particular company may be above or below an industry norm. Finally, one of the properties of a correctly specified regression model is that the error term is normally distributed with a mean equal to zero (Barnes, 1982). Thus, normal probability distribution can be used to estimate probabilities associated with individual values of residuals. This will alleviate the ambiguity in interpreting a company's position relative to other companies within the industry.

8. Implications

Several recommendations can be made for practitioners using IT managerial control ratios for benchmarking their IT costs in relation to other companies within the industry based on the theoretical and methodological issues discussed in this article. First, practitioners should use scatterplots of ratio data to infer the nature of the relationship between a numerator and a denominator used to form a managerial control ratio. If a non-linear relationship is detected, managers should understand that a mean or median ratio may be a misleading baseline for determining a company's position relative to the industry norm. Second, a histogram of ratio values within an industry should be built to get a sense of the underlying probability distribution of these ratios. If the data is skewed, then median should be used as a measure of the norm. Overall, visualizing a probability distribution via a histogram should help practitioners to interpret their company's position with respect to IT spending relative to other organizations within the industry. Finally, given the fact that there can be many alternative explanations as to why a particular company has a high or low value of an IT managerial ratio, benchmarking results should not be viewed as a final verdict, but rather as a starting point for a further, qualitative discussion regarding the level of IT spending and the contribution of IT to the company's overall performance.

Researchers can help practitioners improve robustness of IT cost benchmarking by developing theoretical models that would explain IT budget levels with the help of several independent variables. These theoretical models, once confirmed empirically, can be used for creating regression models that take into account not only organizational size as the main determinant of IT budgets, but also other important factors (e.g. maturity of IT infrastructure, the dependence of a firm on IT in implementing its strategies, efficiency in utilizing IT resources, etc.) Because of the properties of a correctly specified regression model, these regression-based benchmarking models may help improve the reliability and validity of IT cost benchmarking.

9. Conclusion

Cross-sectional IT spending benchmarking can be a quick and simple approach for evaluating organizational IT spending. However, as it was argued in this paper, this practice of evaluating organizational IT spending is plagued by numerous theoretical and methodological issues. This paper does not suggest that, in the light of these problems, companies should not rely on IT managerial control ratios for benchmarking their IT costs. In some situations, managers may not have enough time or other resources for a more rigorous analysis of their IT spending data using more formal and robust evaluation techniques, such as ROI, NPV, balance scorecards, or real options analysis. In that case, IT managerial control ratios may serve as a quick check of the appropriateness of the level of IT spending within an organization. At the minimum, determining a company's position relative to the industry norm can serve as a useful "conversation starter" between a senior technology leader and top management regarding adequacy of financial resources allocated to the IT function and how effectively these resources are utilized. Yet, important capital allocation decisions should not be made without an understanding of the theoretical and methodological issues in relation to IT cost benchmarking using IT managerial control ratios discussed in this article. If a more robust approach to determining and interpreting a company's position is needed, then practitioners should consider adopting the regression-based approach to benchmarking IT costs described in this article. However, additional theoretical and empirical work is needed to develop and validate these regression models. If developed, these models can potentially help practitioners improve reliability and validity of IT cost benchmarking.

References

Barnes, P. (1982) 'Methodological implications of non-normally distributed financial ratios', *Journal of Business Finance & Accounting*, vol. 9, no. 1, pp. 51-62.

Bartels, A. (2007) 'US IT spending benchmarks for 2007', *Forrester Research*.

Bharadwaj, A.S. Bharadwaj, S.G. & Konsynski, B.R. (1999) 'Information technology effects on firm performance as measured by Tobin's q', *Management Science*, vol. 45, no. 7, pp. 1008-24.

Brynjolfsson, E. & Hitt, L. (1996) 'Paradox lost? Firm-level evidence on the returns to information systems spending', *Management Science*, vol. 42, no. 4, pp. 541-58.

DeLone, W.H. & McLean, E.R. (2003), 'The DeLone and McLean model of information systems success: A ten-year update', *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9-30.

Doll, W.J., Deng, X. & Joseph, A.S. (2003), 'A process for post-implementation IT benchmarking', *Information & Management*, vol. 41, no. 2, pp. 199-212.

Dougherty, C. (2007) *Introduction to Econometrics*, London: Oxford University Press.

Gurbaxani, V., Kraemer, K. & Vitalari, N. (1997) 'Note: An economic analysis of IS budgets', *Management Science*, vol. 43, no. 12, pp. 745-1755.

Irani, Z. (2002) 'Information systems evaluation: navigating through the problem domain', *Information & Management*, vol. 40, pp. 11-24.

Kappelman, L., McLean, E., Johnson, V. & Gerhart, N. (2014) 'The 2014 SIM IT key issues and trends study', *MIS Quarterly Executive*, vol. 13, no. 4, pp. 237-63.

Kobelsky, K.W., Richardson, V.J., Smith, R.E. & Zmud, R.W. (2008), 'Determinants and consequences of firm information technology budgets', *The Accounting Review*, vol. 83, no. 4, pp. 957-95.

Koch, C. (2006), 'The metrics trap...And how to avoid it', *CIO Magazine*, vol. 2015, no. December 7.

Krotov, V. (in press) 'Reliability and validity issues in analysis of IT spending using IT managerial control ratios', *Benchmarking: An International Journal*.

Krotov, V. & Ives, B. (in press) 'IT cost benchmarking: Drawing the right conclusions', *Business Horizons*.

Lattin, J., Carrol, D. & Green, P. (2003) *Analyzing Multivariate Data*, Thompson Learning.

Lee, J. & Bose, U. (2002) 'Operational linkage between diverse dimensions of information technology investments and multifaceted aspects of a firm's economic performance', *Journal of Information Technology (Routledge, Ltd.)*, vol. 17, no. 3, pp. 119-31.

Lev, B. & Sunder, S. (1979) 'Methodological issues in the use of financial ratios', *Journal of Accounting and Economics*, no. 1, pp. 187-210.

Mata, F.J., Fuerst, W.L. & Barney, J.B. (1995) 'Information technology and sustained competitive advantage: A resource-based analysis', *MIS Q.*, vol. 19, no. 4, pp. 487-505.

Mitra, S. & Chaya, A.K. (1996) 'Analyzing cost-effectiveness of organizations: The impact of information technology spending', *Journal of Management Information Systems*, vol. 13, no. 2, pp. 29-57.

Nikkinen, J. & Sahlstrom, P. (2004) 'Distributional properties and transformation of ginancial tatios: The impact of the accounting environment', *Advances in International Accounting*, vol. 17, pp. 85–101.

Piccoli, G. & Ives, B. (2005) 'Review: IT-dependent strategic initiatives and sustained competitive advantage: a review and synthesis of the literature', *MIS Q.*, vol. 29, no. 4, pp. 747-76.

Roach, S.S. (1991) 'Services under diege - The restructuring imperative. ', *Harvard Business Review*, vol. 69, no. 5, pp. p82-91.

Schryen, G . (2013) 'Revisiting IS business value research: What we already know, what we still need to know, and how we can get there', *European Journal of Information Systems*, vol. 22, no. 2, pp. 139-69.

Whittington, G. (1980), 'Some basic properties of accounting ratios', *Journal of Business Finance & Accounting*, vol. 7, no. 2, pp. 219-33.